



Universiteit  
Leiden  
The Netherlands

## Statistical modelling of time-varying covariates for survival data

Spreafico, M.

### Citation

Spreafico, M. (2022, October 12). *Statistical modelling of time-varying covariates for survival data*. Retrieved from <https://hdl.handle.net/1887/3479768>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3479768>

**Note:** To cite this publication please use the final published version (if applicable).

# **Statistical modelling of time-varying covariates for survival data**

Proefschrift

ter verkrijging van  
de graad van doctor aan de Universiteit Leiden,  
op gezag van rector magnificus prof.dr.ir. H. Bijl,  
volgens besluit van het college voor promoties  
te verdedigen op woensdag 12 oktober 2022  
klokke 13.45 uur

door

**Marta Spreafico**  
geboren te Bellano, Italië  
in 1993

**Promotor:**

Prof.dr. M. Fiocco

Universiteit Leiden, Leids Universitair Medisch Centrum,  
Prinses Máxima Centrum

**Co-promotor:**

Dr. F. Ieva

Politecnico di Milano, Human Technopole

**Promotiecommissie:**

Prof.dr. P. Stevenhagen

Universiteit Leiden

Prof.dr. P. Grünwald

Universiteit Leiden, Centrum Wiskunde & Informatica

Prof.dr. A. Farcomeni

Università degli Studi di Roma Tor Vergata

Dr. E. Musta

Universiteit van Amsterdam

Prof.dr. H. Putter

Leids Universitair Medisch Centrum, Universiteit Leiden



**POLITECNICO**  
MILANO 1863



**Universiteit**  
**Leiden**  
The Netherlands

This doctoral dissertation was part of a cotutelle agreement and was presented to Politecnico di Milano (Italy) in fulfillment of the requirements for the degree of Doctor in *Mathematical Models and Methods in Engineering* (Metodi e Modelli Matematici per l'Ingegneria), and to Universiteit Leiden (The Netherlands) in fulfillment of the requirements for the degree of Doctor in *Mathematics* (Wiskunde).

*A mio nonno Ermanno<sup>†</sup>  
con infinito amore*

*“Uau”*



# Contents

<b>Introduction</b>	<b>1</b>
I.1 Basics of Survival Analysis . . . . .	3
I.2 Epidemiological and clinical framework . . . . .	6
I.2.1 Pharmacoepidemiology in Heart Failure . . . . .	6
I.2.2 Chemotherapy in Osteosarcoma . . . . .	8
I.3 Overview of the thesis . . . . .	9
<b>PART I Pharmacoepidemiology in Heart Failure</b>	<b>13</b>
<b>1 A new method for measuring adherence to polypharmacy</b>	<b>15</b>
1.1 Materials and Administrative data . . . . .	16
1.1.1 Study setting . . . . .	16
1.1.2 Data sources . . . . .	16
1.1.3 Study population . . . . .	17
1.2 Methodologies . . . . .	18
1.2.1 Target dosages according to guidelines . . . . .	18
1.2.2 Adherence measures . . . . .	18
1.2.3 Adherence to polypharmacy . . . . .	19
1.2.4 Outcome measure . . . . .	20
1.2.5 Survival Analysis: multivariable Cox regression models . . . . .	20
1.3 Results . . . . .	21
1.3.1 Cohort selection . . . . .	21
1.3.2 Standardized Daily Dose . . . . .	25
1.3.3 Patients' adherence measures . . . . .	25
1.3.4 Multivariable Cox models for survival outcome . . . . .	27
1.4 Final remarks . . . . .	29
<b>2 Joint modelling of time-varying adherence to medication and survival</b>	<b>33</b>
2.1 Statistical Methodologies . . . . .	35
2.1.1 Pharmacological time-varying covariates . . . . .	35
2.1.2 Joint model specification . . . . .	37
2.2 Materials and Administrative data . . . . .	39
2.2.1 Study setting . . . . .	39
2.2.2 Administrative data sources . . . . .	40
2.2.3 Pharmacological time-varying covariates for ACE/ARB therapy . . . . .	41

2.2.4	Outcome measure . . . . .	42
2.3	Results . . . . .	42
2.3.1	Study cohort . . . . .	42
2.3.2	Joint models for time-varying consumption and adherence to ACE/ARB therapy . . . . .	43
2.4	Final remarks . . . . .	52
<b>3</b>	<b>Functional modelling of recurrent events on time-to-event processes</b>	<b>55</b>
3.1	Materials and Administrative data . . . . .	57
3.1.1	Administrative data sources . . . . .	57
3.1.2	Study design and outcome measure . . . . .	58
3.2	Statistical Methodologies . . . . .	59
3.2.1	Marked point process formulation for recurrent events . . . . .	59
3.2.2	Functional linear Cox regression model with multiple functional compensators . . . . .	62
3.3	Data application . . . . .	64
3.3.1	Step 1: Data preprocessing & clinical history . . . . .	64
3.3.2	Step 2: Modelling compensators of marked point processes . . . . .	65
3.3.3	Step 3: Summarize compensators through Functional Principal Component Analysis . . . . .	69
3.3.4	Step 4: Predictive functional Cox model for overall survival . . . . .	71
3.4	Final remarks . . . . .	72
A.	Appendix to Chapter 3 . . . . .	75
<b>PART II</b>	<b>Chemotherapy in Osteosarcoma</b>	<b>77</b>
<b>4</b>	<b>Modelling time-varying covariates effect on survival via Functional Data Analysis</b>	<b>79</b>
4.1	Statistical Methodologies . . . . .	81
4.1.1	Time-varying covariates and survival frameworks . . . . .	81
4.1.2	Time-Varying covariate Cox Model . . . . .	83
4.1.3	Functional covariate Cox Model . . . . .	83
4.2	MRC BO06 randomized clinical trial data . . . . .	86
4.2.1	Trial protocol . . . . .	86
4.2.2	Sample cohort selection and baseline characteristics . . . . .	87
4.3	Results . . . . .	88
4.3.1	Time-varying characteristics . . . . .	88
4.3.2	Time-Varying covariate Cox Model . . . . .	91
4.3.3	Functional covariate Cox Model . . . . .	92
4.4	Final remarks . . . . .	98
<b>5</b>	<b>A novel longitudinal method for quantifying multiple overall toxicity</b>	<b>103</b>
5.1	BO06 data . . . . .	104
5.1.1	Treatment-related factors . . . . .	105

5.2	Statistical Methodologies . . . . .	107
5.2.1	Longitudinal Multiple Overall Toxicity (MOTox) scores and outcomes	107
5.2.2	Statistical analysis . . . . .	108
5.3	Results . . . . .	109
5.3.1	Non-haematological longitudinal Overall Toxicity scores . . . . .	109
5.3.2	Multivariable logistic regression models for high overall toxicity over cycles . . . . .	111
5.4	Final remarks . . . . .	112
B.	Appendix to Chapter 5 . . . . .	117
<b>6</b>	<b>Modelling longitudinal profiles of latent probability and relative risk via latent Markov models and compositional data</b>	<b>119</b>
6.1	Statistical Methods . . . . .	121
6.1.1	Motivations for latent Markov models for longitudinal toxicity data	121
6.1.2	Latent Markov model with covariates . . . . .	122
6.1.3	Model selection . . . . .	124
6.1.4	Longitudinal profiles: latent probability and relative risk . . . . .	125
6.2	Data application . . . . .	127
6.2.1	Longitudinal toxicity data: item-response categories . . . . .	127
6.2.2	Latent Markov model for longitudinal toxicity data . . . . .	128
6.2.3	Longitudinal profiles of Latent Overall Toxicity . . . . .	133
6.3	Final remarks . . . . .	135
C.	Appendix to Chapter 6 . . . . .	138
<b>7</b>	<b>Investigating the causal effects of joint-exposure on survival outcome in presence of time-varying confounders</b>	<b>139</b>
7.1	Data description . . . . .	141
7.1.1	Control arms protocol and Cohort selection . . . . .	142
7.1.2	Complexity of chemotherapy data . . . . .	143
7.1.3	Chemotherapy exposure characteristics . . . . .	146
7.2	Causal inference structure and methods . . . . .	150
7.2.1	Event-Free Survival Outcome . . . . .	151
7.2.2	Causal inference assumptions for marginal structural models . . . . .	151
7.2.3	Causal structure of chemotherapy data . . . . .	153
7.2.4	Joint-exposure and marginal structural Cox model for DAG-1 . . . . .	156
7.2.5	Joint-exposure and marginal structural Cox model for DAG-2 . . . . .	158
7.3	Results . . . . .	161
7.3.1	Joint-exposure descriptive and association with EFS . . . . .	161
7.3.2	IPTW diagnostics . . . . .	162
7.3.3	Causal inference through marginal structural Cox models . . . . .	165
7.4	Final remarks . . . . .	167
D.	Appendix to Chapter 7 . . . . .	170
	<b>Conclusions</b>	<b>178</b>



*Contents*

<b>Bibliography</b>	<b>192</b>
<b>Summary</b>	<b>193</b>
<b>Samenvatting</b>	<b>195</b>
<b>Sommario</b>	<b>197</b>
<b>List of Publications</b>	<b>199</b>
<b>Acknowledgments</b>	<b>201</b>
<b>Curriculum Vitae</b>	<b>203</b>

# INTRODUCTION

---

Survival analysis [97, 106, 202] is an important field of statistics dealing with time-to-event (or life-time) data. In life-time data the outcome variable of interest, named *survival time*, is the amount of time elapsed from a so-called origin/initiating event until an event of interest. Many examples exist in several research fields: time from diagnosis of disease until death in medicine, time to failure of a machine part in engineering, or duration of unemployment in socio-economic sciences. A key characteristic of survival data is that they are generally *partially observed*, coming as a mixture of complete and incomplete observations. Some individuals might not have experienced the event of interest at the end of the study period or have dropped out before the event has occurred. It is only known that the event had not occurred before the last observation time but the exact event time is unknown. This type of mechanism is called *right-censoring*. Whether studying a specific event or a sequence of events, special statistical methodologies are required to handle this particular type of partially observed data, as *censoring* complicates all technical issues involved in life-time analyses.

The work carried out in this thesis is motivated by specific medical questions. For example in chronic diseases or cancers, survival models can be used to investigate if a patient's age, gender, medical treatment, or other covariates are associated to the risk of death. Typically, the term covariate refers to *time-fixed* predictive or explanatory variable, whose value does not change over time (e.g., demographic or baseline information). However, medical follow-ups are characterized by *time-varying covariates* (i.e., attributes that may have different values at different time points), such as drugs intake, treatment doses, biomarkers or toxicity, or by *repeated events* occurring during the study, like office visits, subsequent drug consumption or hospital admissions. These processes that change or re-occur over time are of great interest because the way their dynamic patterns evolve may affect patient's health status and disease progression.

Due to the complexity of these phenomena, a *piecewise-constant* or *fixed-baseline* approach is usually preferred in the clinical literature, discarding their dynamic and/or temporal components. In this way, the information that these processes can provide if the association between time-varying and time-event data is properly captured is completely lost. Complex mathematical methods are therefore required to model disease evolution and characterise its relationship to the dynamic nature of time-dependent features.

The current thesis arises in cross-sectional fields of biostatistics and healthcare research and focuses on developing mathematical and statistical methods to represent time-varying processes from complex raw data, and model them within the context of time-to-event analysis. The main purpose is to enrich the knowledge available for modelling survival

with relevant features related to the dynamic characteristics of interest. These aspects are rarely addressed in the literature and may provide new insights for medical research, representing a challenge of both clinical relevance and statistical interest. This work focuses on time-varying processes and, more specifically, examines (i) *dynamic representation* and (ii) *modelling* in a time-to-event setting.

In terms of *representation*, the main issue consists on identifying appropriate dynamic characterizations of the processes under study. Here, several levels of complexities must be considered. On one hand, producing models that are suitable for dealing with complex data is not straightforward. A huge amount of data-integration and preprocessing work is needed to make data suitable for the statistical analysis preserving clinical interpretability. On the other hand, when defining the mathematical formulation, the nature of the processes and aspects such as temporal dynamics, categorical levels or recurring events must be taken into account. Different methodologies are therefore proposed throughout this thesis, including:

- complex data integration to define novel pooled or longitudinal representations related to time-dependent covariates (Chapters 1, 2, 5);
- recurrent events modelling and point processes theory to retrieve the trajectories of compensators related to appropriate stochastic processes for recurrent events (Chapter 3);
- functional data analysis techniques to reconstruct features able to incorporate trends and variations in the evolution of the processes as continuous smoothed functions of time (Chapters 3, 4);
- latent Markov models and compositional data analysis to model latent disease evolution on the basis of interval-based categorical observations (Chapter 6);
- direct acyclic graphs to identify all possible confounders and their relation with the time-dependent exposure under study and then engage a causal inference paradigm (Chapter 7).

At this stage, the main challenge is to best represent the time-varying characteristics of interest by managing the complex trade-off between clinical interpretability and mathematical formulation.

In terms of time-to-event *modelling*, innovative statistical models for identifying and quantifying the association between time-varying processes and patient survival are proposed. In medical statistics the Cox proportional hazards model [46] has been widely used for survival data due to its flexibility. It has also been extended to account for time-dependent covariates [202, 97] using piecewise-constant values among different time measurements. Discarding the continuous nature of the process underlying the data, this approach leads to biased results and fails to account for possible measurement errors [16]. Therefore, the main purpose during the survival *modelling* phase is to develop advanced versions of Cox-type methods capable of handling dynamic covariates, while preserving clinical

interpretability. Depending on the context of the study, different approaches are then proposed, such as:

- a joint modelling to simultaneously analyse longitudinal and time-to-event data through appropriate mixed-effect and Cox-type models (Chapter 2);
- dimensionality reduction techniques for functional data in order to incorporate dynamic predictors into advanced version of traditional Cox regression models (Chapters 3, 4);
- Cox-type marginal structural models to assess the causal effects of a joint-exposure on survival outcomes, in presence of time-varying confounders (Chapters 7).

The main purpose is to add to the current literature relevant survival models which are able to incorporate dynamic information usually discarded by standard approaches.

This work has an impact on the community of researchers in mathematics and statistics, but it provides also useful tools to support doctors and clinicians in their daily work. All research topics are motivated by specific clinical questions related to two different medical domains, corresponding to the two main parts of the thesis: cardiological (Part I) and oncological (Part II) patients.

The identification of dynamic representations able to reflect variability and differences among patients may improve patients' profiling and tailor their therapies. This can lead to new knowledge for both general guidelines and personalised treatments, and make the pathway of patients through the healthcare system more efficient and effective. Therefore, the development of novel methodologies capable of extracting additional information to enrich survival models 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.

The remainder of this Chapter is composed as follows. Section I.1 introduces basic concepts of survival analysis. Section I.2 presents the motivating epidemiological and clinical frameworks. Section I.3 gives an overview of the remaining Chapters of this thesis.

## I.1. Basics of Survival Analysis

This section aims at providing notation for the rest of the thesis. A first step in understanding survival analysis [97, 106, 202] is in understanding the *partially-observed* time-to-event data it has to deal with.

Let  $T^*$  be the non-negative random variable denoting the true event time, i.e., the amount of time elapsed from the origin event until an event of interest. If a patient dropped out of the study early or the study ended before the event of interest occurred, or another event occurred, the event time may not be observed and right-censoring occurs. Let  $C$  be a random variable that denotes the time for the censoring mechanism (i.e., the last time a subject was observed in the study). The time-to-event information observed for

an individual is given by the pair  $(T, D)$ , where  $T = \min(T^*, C)$  is the *survival time* and  $D = \mathbb{1}(T \leq C)$  is the *event indicator*,  $D = 1$  if the true event time was observed or  $D = 0$  if censored. The event time  $T^*$  and the right-censoring time  $C$  are usually assumed to be independent.

To study the distribution of the survival time  $T$  different quantities are of interest. The *survival function* (or survival curve)  $S(t) = \Pr(T > t)$  at time  $t$  is equal to the probability of being event-free at time  $t$ . It is a non-increasing function with  $S(0) = 1$  because everybody is event-free at the time origin, and as  $t$  gets large as  $S(t)$  tends to 0 because everybody eventually experiences the event of interest. The *hazard function*, i.e., the instantaneous risk of failure at time  $t$ , conditional on survival to that time, is defined as:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr(t \leq T < t + \Delta t | T \geq t)}{\Delta t}.$$

The hazard function can be used to express:

- the *survival function*:

$$S(t) = \exp \left\{ - \int_0^t h(u) du \right\},$$

which can be estimated by using a non-parametric method called the product limit estimator better known as the *Kaplan-Meier* estimator [99]

$$\hat{S}(t) = \prod_{j: t_j^* \leq t} \left( 1 - \frac{d_j}{n_j} \right);$$

- the *cumulative hazard function* representing the total accumulated risk of experiencing the event of interest that has been gained by progressing to time  $t$ :

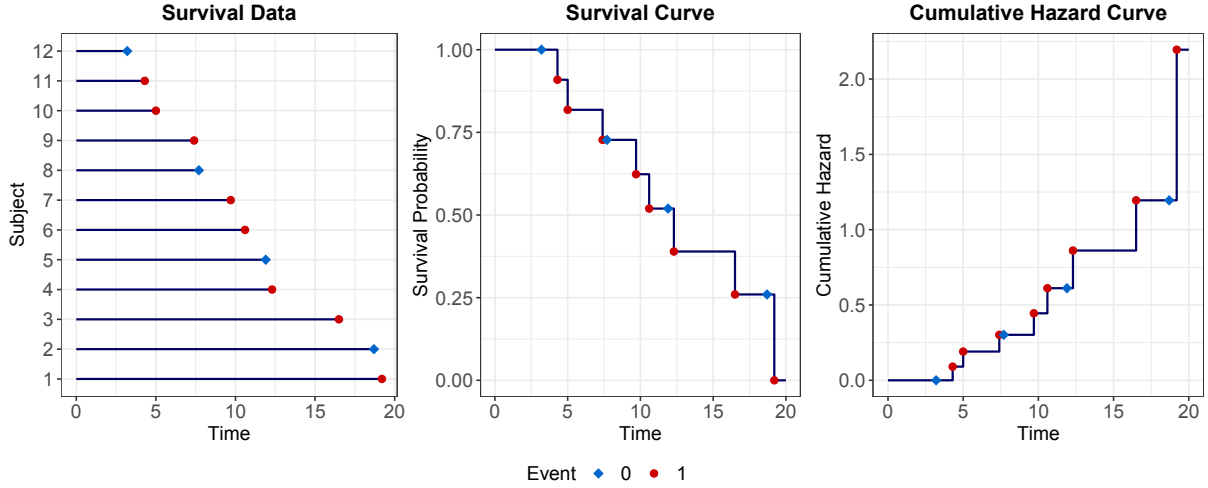
$$H(t) = \int_0^t h(u) du,$$

which can be estimated by using the *Nelson-Aalen* estimator [145, 146, 1]

$$\hat{H}(t) = \sum_{j: t_j^* \leq t} \frac{d_j}{n_j}.$$

In both estimators  $\hat{S}(t)$  and  $\hat{H}(t)$ ,  $0 < t_1^* < t_2^* < \dots < t_j^* < \infty$  denote the observed ordered true event times with  $J$  equal to the total number of events,  $d_j$  and  $n_j$  denote the number of events and the number of individual still at risk at time  $t_j^*$ , respectively. Figure I.1 shows an example of time-to-event data (left panel) and the corresponding estimated survival and cumulative hazard curves (central and right panels, respectively). Subjects 2, 5, 8 and 12 are right-censored. Both curve have steps at event times (red points) and remain unchanged at censoring times (light-blue diamonds). The censoring times however affect the size of the jumps the curves make.

The main goal of survival studies is to estimate the hazard function and to assess how the covariates affect it. The most widely used model to study the effect of a covariate



**Figure I.1.** Left panel: time-to-event data for 12 subjects (*light-blue diamonds*: censored subjects; *red points*: event subjects). Central panel: Kaplan-Meier survival estimate. Right panel: Nelson-Aalen cumulative hazard estimate.

vector  $\omega$  on the survival is the Cox proportional-hazard regression [46]. It is based on the proportional hazards assumption stating the effects of the covariates are multiplicatively related to the hazard, defined as:

$$h(t|\omega) = h_0(t) \exp\{\theta^T \omega\},$$

where  $h_0(t)$  is the unspecified non-negative baseline hazard function and  $\theta$  is the vectors of regression coefficients. Inference for coefficients  $\theta$  is based on maximizing the *partial likelihood* [46]:

$$L(\theta) = \prod_{j=1}^J \frac{\exp(\theta^T \omega_{(j)})}{\sum_{i \in R(t_j^*)} \exp(\theta^T \omega_i)}$$

where  $0 < t_1^* < t_2^* < \dots < t_j^* < \infty$  are the observed ordered event times,  $\omega_{(j)}$  denotes the covariates of the individual who experiences the event at time  $t_j^*$ ,  $\omega_i$  is the covariate vector of individual  $i$  and  $R(t_j^*)$  denotes the set of individuals still at risk at time  $t_j^*$ . The baseline hazard  $h_0(t)$  can then be estimated by *Breslow estimator* [32].

For each covariate  $l$ , the quantity  $\exp(\hat{\theta}_l) = HR_l$  is called Hazard Ratio (HR). For a categorical explanatory variable, the HR represents the ratio between the predicted hazard for a member of one group and for a member of the reference group, by holding everything else constant. For a continuous explanatory variable, the same interpretation applies to a 1-unit difference. In particular, a  $HR_l < 1$  indicates a reduction in the hazard function (i.e., an increase in the survival), meaning that the  $l$ -th covariate is a *protective* factor for the time to the event of interest; whereas the opposite ( $HR_l > 1$ ) indicates that the  $l$ -th covariate is a *risk* factor.

The Cox model has also been extended to account for a covariate vector  $\omega(t)$  which can change values during follow-up [202, 97]. Since time-dependent observations are only available at the times of measurements, the Time-Varying covariate Cox Model (TVCM) uses the last-observation-carried-forward (LOCF) approach [206]: between two subsequent

observations, the value of the time-varying covariate is kept constant at the last observed value. Under TVCM, the hazard function is

$$h(t|\boldsymbol{\omega}(t)) = h_0(t) \exp \{ \boldsymbol{\theta}^T \boldsymbol{\omega}(t) \}.$$

The partial likelihood is defined similarly to the model with only time-fixed covariates [202, 97], considering piecewise-constant values for the time-dependent covariates.

## I.2. Epidemiological and clinical framework

As mentioned before, the research topics of this thesis are related to two different applications: (i) the study of pharmacotherapy in patients with heart failure and (ii) the investigation of chemotherapy treatment in patients with osteosarcoma. The motivating clinical research issues for both cases are now introduced.

### I.2.1. Pharmacoepidemiology in Heart Failure

Pharmacoepidemiology is the study of the use and the effects of drugs in large numbers of people through the application of epidemiological methods [194, 195]. A modern definition of pharmacoepidemiology [220] is

*“the study of the use and effects/side-effects of drugs in large numbers of people with the purpose of supporting the rational and cost-effective use of drugs in the population thereby improving health outcomes”.*

The investigation about quantification, understanding and evaluation of the processes of prescribing, dispensing and consumption of medicines and their effect on patients’ clinical courses refers to a branch of pharmacoepidemiology known as Drug Utilization Research (DUR) [53]. As defined by the World Health Organization in 1977 [220], DUR consists in the

*“marketing, distribution, prescription, and use of a drug in the society, with special emphasis on the resulting medical, social and economic consequences”.*

The ultimate goal of DUR is hence to identify and communicate the proper use of drugs to patients, combining researches which belong to the medical, economical and social fields.

In DUR and pharmacotherapy, the achievement of a certain level of medication intake or adherence is an important component of patient’s care. According to the taxonomy introduced in [212], *adherence* to medication is defined as the process by which patients take their medication as prescribed. In the past decade there has been substantial growth in clinical research focused on adherence to medication, partly owing to the increasing awareness of the problem and partly due to the pervasiveness of non-adherence behaviours among patients [212]. Individual patient’s adherence is usually reported as a percentage of the actual medication taken over a defined period of time [27]. Poor adherence to

## 1.2. Epidemiological and clinical framework

medication regimens accounts for substantial worsening of disease, death and increased health care costs [147]. In particular, in long-term therapies poor adherence severely compromises the effectiveness of treatment, representing a critical issue both for quality of life and health economics [219].

Long-term therapies are typical in chronic diseases, such as Heart Failure (HF). HF is a major and growing public health issue, characterized by high costs, steep morbidity and mortality rates [129]. HF is widespread all over the world, especially among people over 65 years, with a prevalence of 1–2% in Western countries and an incidence from 5 to 10 per 1000 persons per year [150]. In particular, in Italy about 80,000 new cases per year are recorded [131] and it is the second cause of hospitalization, after vaginal delivery. This complex clinical syndrome is characterized by structural or functional cardiac disorders that impair the ability of one or both ventricles to fill with or eject blood [87]. It may be provoked by several different cause, such as myocardial ischaemia, high blood pressure, cardiomyopathies, valvular heart disease, pulmonary hypertension or congenital heart disease [150]. Due to HF, organs and tissues receive insufficient quantities of oxygen and nutrients for their metabolic needs, and there is an accumulation of excess fluid in the lungs and tissues [143]. This condition can worsen to the point of acute pulmonary oedema and death. According to data from different studies conducted in America and Europe, 30-day, 1-year, and 5-year mortality are around 10% to 20%, 30%, and 65% respectively [150].

Patients hospitalized for HF are at high risk for all-cause re-hospitalization, with a 1-month readmission rate of 25%. Self-care in HF comprises treatment adherence and health maintenance behaviours [75]. HF patients should learn to take medications as prescribed, stay physically active, restrict sodium intake, get vaccinations and understand how to monitor for signs of worsening HF [75]. Therapeutic and pharmacological interventions in HF patients aim at reducing symptoms, morbidity and mortality. Depending on the different symptoms, the following pharmacological treatments have been established as disease-modifying drugs of routine use in HF treatment: Angiotensin-Converting Enzyme (ACE) inhibitors, Angiotensin II Receptor Blockers (ARB – as an alternative for people who cannot tolerate ACE), Beta Blockers (BB), Anti-Aldosterone agents (AA) and diuretics [139, 222, 221, 154, 75]. Different studies showed that a proper and monitored drug intake in HF patients could improve their clinical status, functional capacity and quality of life, prevent hospital admission and reduce mortality [154]. However, it is well known that adherence in HF is low and not satisfactory [179], even few months after the first hospital discharge for HF. Poor adherence to medications leads to increased HF exacerbations, reduced physical function, higher risk for hospital readmission and death [154], representing a significant problem in HF management in both healthy and economic terms.

In the last decades, secondary or administrative databases have increasingly been used in the pharmacotherapy field [14], becoming one of the most employed sources to evaluate adherence to medication [102]. Although administrative data are mainly collected for managerial and economic purposes [89], their use for clinical and epidemiological purposes has become an accepted practice [90] as they are particularly suitable for investigating different areas, such as profile of drug uses [45]. This requires intensive computational



INTRODUCTION

effort to link different data sources (e.g., drug purchases, death registry, hospitalisation records) in order to create usable databases. According to the state of the art, adherence to medication is usually modelled by a numerical variable representing the percentage of the actual drug taken over a pre-defined period of time [27]. This approach does not consider *changes in drugs consumption over time* or the occurrence of *re-hospitalizations*. Moreover, the most used adherence measures [14, 102] refer to monotherapy medication, although chronic HF patients usually undergo a *polypharmacy treatment* (i.e., the simultaneous use of different medications). Given the clinical relevance of these aspects and the impact they may have on patients' survival, the development of new methodologies to overcome these problems is a challenge for both clinical research and statistical modelling. This topic is discussed in Part I of this thesis.

## 1.2.2. Chemotherapy in Osteosarcoma

Osteosarcoma is a malignant bone tumour mainly affecting children and young adults. Although osteosarcoma is the most common primary malignant bone cancer, it is a rare disease and has an annual incidence of 3-4 patients per million [185]. Osteosarcoma can occur in any bone but it is often localized in the extremities: the most common primary sites are the distal femur, the proximal tibia, and the proximal humerus, with > 50% originating around the knee [166, 15]. Local pain, followed by localized swelling and limitation of joint movement, are the typical signs and symptoms of osteosarcoma [166].

Osteosarcoma treatment typically involves surgery and chemotherapy. The goal of surgery is the complete tumour removal. Different surgical techniques are available and the surgeon must choose the most appropriate for each individual, taking into account several factors such as the size of the tumour and its location, as well as the influence the surgery will have on the patient's daily life. In case of unresectable tumours or of microscopic residual tumour foci following attempted resection, recent research also suggests radiotherapy in addition to standard therapy [166]. In modern treatment schedules, chemotherapy to kill cancer cells is usually a combination of doxoubicine (DOX) and cisplatin (CDDP), with or without high-dose methotrexate and/or ifosfamide and/or etoposide. DOX and CDDP are considered the most active agents against osteosarcoma, but the ideal combination remains to be defined [166, 15]. Since the extent of histological response to pre-operative chemotherapy (i.e., the improvement in the appearance of microscopic tissue specimens) represents the strongest prognostic factor of survival known so far in osteosarcoma [31], most current protocols include a period of pre-operative (neoadjuvant) chemotherapy. Post-operative (adjuvant) chemotherapy is then used to kill any cancer cells that might remain after surgery.

Multidisciplinary management including neoadjuvant and adjuvant chemotherapy with aggressive surgical resection [166] or intensified chemotherapy has improved clinical outcomes although the overall 5-year survival rate has remained unchanged in the last 40

years at 60-70% [15]. The impact of chemotherapy dose modification on patients' survival is still unclear [111]. In cancer trials the relationship between chemotherapy dose and clinical efficacy outcomes is difficult to analyse due to the presence of negative feedback between exposure to cytotoxic drugs and other aspects, such as latent accumulation of chemotherapy-induced toxicity. Patients may develop different types of toxic side effects, ranging in severity from minor, asymptomatic changes to life-threatening injuries or death [204]. Depending on patients' treatment history or development of toxicity, biomarkers values change and chemotherapy treatment is modified by delaying a course or reducing the dose intensity. Being at the same time risk factors for mortality and predictors of future exposure levels, toxicities are time-dependent confounders for the effect of chemotherapy on patient's survival [112].

Due to the complexity of longitudinal chemotherapy data, the ways chemotherapy doses and toxicities are accounted for into predictive models in literature and current practice for cancer research is far from being informative as it may be. Chemotherapy is usually modelled by different allocated regimens, i.e., by Intention-To-Treat (ITT) analysis [70]. This means that protocol deviations or changes in drug intake over time are not considered in the analysis [110]. Toxicities are usually incorporated as summary indexes (e.g., maximum toxicity over time, maximum grade among events, or weighted sums of individual toxic effects) discarding substantial amount of information (e.g., isolated vs repeated events, single vs multiple episodes, longer-lasting lower-grade toxicities, toxic events timing). As neglecting the time component may give an inaccurate depiction of toxicity and chemotherapy regimen intensity, characterisation of both aspects is of interest to patients and clinicians engaged in shared decision making about a treatment strategy. The development of models and methods able to deal with all these peculiar aspects is hence of statistical interest and of clinical relevance. This topic is discussed in Part II of the this thesis.

### 1.3. Overview of the thesis

The current thesis aims at developing mathematical and statistical methods to properly *represent* time-dependent processes and *modelling* them within the context of time-to-event analysis by means of appropriate Cox-type survival models.

Part I "*Pharmacoepidemiology in Heart Failure*" focuses on methods for representing drug consumption, adherence to medications or re-hospitalization events exploiting administrative databases, and modelling their effect on long-term survival in HF patients. Administrative data from *Friuli Venezia Giulia* and *Lombardia* [164] regions in Italy are analysed. In particular, records from Hospital Discharge Charts (i.e., admission to hospital), Public Drug Distribution Systems (i.e., drugs purchases) and Registries of Deaths are considered.

In Chapter 1 we investigate patients' adherence to disease-modifying therapies and the prognostic impact on survival, exploiting administrative databases of the *Friuli Venezia*

*Giulia* Italian region. A novel method to represent adherence to polypharmacy, i.e., the Patient Adherence Index (PAI), is proposed as the ratio between the number of drugs to which a patient is adherent and the number of purchased drugs. Taking advantage of the developed index, the effect of adherence to polypharmacy on patients survival is then investigated through Cox regression model, adjusting for patient-specific characteristics. Although PAI is still a time-fixed covariate, this study requires complex data integration procedures among different data sources, representing a first step forward in the pharmacoepidemiology context for HF patients as few data on polypharmacy adherence exist in a real-world setting. The content of this chapter is based upon:

- **M. Spreafico**, F. Gasperoni, G. Barbati, F. Ieva, A. Scagnetto, L. Zanier, A. Iorio, G. Sinagra and A. Di Lenarda. *Adherence to Disease-Modifying Therapy in Patients Hospitalized for HF: Findings from a Community-Based Study*. American Journal of Cardiovascular Drugs, 20:179–190, 2020 [187].

In Chapter 2 we propose an innovative method to represent adherence to medication as time-varying covariate and to investigate its dynamic effect on patients' survival using a joint modelling framework. Two different longitudinal representations are introduced: a continuous time-dependent variable, which indicated the cumulative months covered by drug consumption up to time  $t$ , and a dichotomous time-dependent variable, which indicates if the patient is adherent to the therapy at time  $t$ . The development of (generalized) mixed effect models for these longitudinal processes joint with Cox-type regression model for time-to-death allows to capture the interaction among processes over time, representing a more realistic and informative approach with respect to the commonly used time-fixed measures. Administrative databases of the *Lombardia* Italian region provided by *Regione Lombardia - Healthcare Division* [164] are analysed. The content of this chapter is based upon the following publication:

- **M. Spreafico** and F. Ieva. *Dynamic monitoring of the effects of adherence to medication on survival in heart failure patients: A joint modeling approach exploiting time-varying covariates*. Biometrical Journal, 63(2):305–322, 2021 [188].

Chapter 3 concerns the development of a new methodology to extract and summarize information from trajectories of compensators of suitable marked point processes for recurrent events intended as functional data. The developed methodology involves database integration, Functional Data Analysis (FDA) and marked point process modelling of critical events of interest, i.e., drugs purchases and re-hospitalizations. First, the functional trajectories (i.e., the compensators of such processes, which may represent the rate at which events happen) are retrieved by means of FDA theory. This new information is then included into a predictive Cox-type model, exploiting Functional Principal Component Analysis (FPCA) techniques. The introduction of this novel way to account for dynamic processes allows for modelling self-exciting behaviours, for which the occurrence of events in the past increases the probability of a new event, including a large piece of information about patient's clinical history contained in the administrative data. The developed approach is able to take into account the fact that HF patients usually experience several re-hospitalizations and consume different types of drugs at the same time, representing a novelty for clinical and pharmacological research in the direction of properly

treating multimorbidity and polypharmacy. Administrative databases of the *Lombardia* Italian region [164] are analysed. This chapter is based on the following publication:

- **M. Spreafico** and F. Ieva. *Functional modeling of recurrent events on time-to-event processes*. Biometrical Journal, 63(5):948–967, 2021 [189].

Part II “*Chemotherapy in Osteosarcoma*” focuses on methods to represent and model chemotherapy treatment and related effects in cancer patients, such as dose modifications, biomarkers changes, toxicities evolution over time and their associations with survival. Clinical data from randomized trials funded by the Medical Research Council (MRC) (<https://www.ukri.org/councils/mrc/>) and the European Organisation for Research and Treatment of Cancer (EORTC) (<https://www.eortc.org>) for patients with high-grade osteosarcoma are analysed.

In Chapter 4 we propose a Functional covariate Cox Model (FunCM) to study the association between time-varying processes and time-to-death outcome. FunCM first exploits FDA techniques to represent time-varying processes in terms of functional data. Then, information related to the evolution of the functions over time is incorporated into functional regression models for survival data through FPCA. FunCM is compared to a standard TVCM, commonly used despite its limiting assumptions that covariate values are piecewise-constant in time and measured without errors. Data from MRC BO06/EORTC 80931 randomised controlled trial [119] are analysed. Time-varying covariates related to alkaline phosphatase levels, white blood cell counts and chemotherapy dose during treatment are investigated. The proposed method allows to detect differences between patients with different biomarkers and treatment evolutions, and to include this information in the survival model. The content of this chapter is based on the following work:

- **M. Spreafico**, F. Ieva and M. Fiocco. *Modelling time-varying covariates effect on survival via Functional Data Analysis: application to the MRC BO06 trial in osteosarcoma*. Statistical Methods & Applications, 2022 [192]. <https://doi.org/10.1007/s10260-022-00647-0>

Chapters 5 and 6 focus on the methodological aspects concerning a proper representation of the overall toxicity burden over time, still lacking in the medical literature due to the complex nature of both chemotherapy protocol and data. In both cases, data from the MRC BO06/EORTC 80931 randomized clinical trial [119] are analysed. In Chapter 5 we exploit complex database processing and aggregation methods to introduce two innovative longitudinal representations of Multiple Overall Toxicity (MOTox), a continuous mean-max score and a dichotomous one. These new representations are then used to investigate the evolution of high-MOTox over treatment through the implementation of cycle-specific multivariable logistic regression models adjusted for previous toxicity levels and patient’s characteristics. Although this approach represents a flexible method for quantifying the individual evolution of overall toxicity in cancer patients compared to traditional indexes, it discards the categorical nature of the observed toxic grades. For this reason, in Chapter 6 we propose a new taxonomy based on latent Markov models with covariates and compositional data theory to (i) represent the overall toxicity as the latent process that

affects the distribution of the observed response variables (i.e., the interval-based categorical toxic levels), (ii) identify different states of Latent Overall Toxicity (LOTox) burden, and (iii) model the personalized *longitudinal LOTox 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. Together, absolute probabilities and relative risks provide a full picture of the individual LOTox dynamics during treatment, which may be considered as a proxy for patient’s quality of life and used to describe patient’s response to therapy over cycles in terms of toxic side effects. Chapter 5 is based upon the following publication:

- **M. Spreafico**, F. Ieva, F. Arlati, F. Capello, F. Fatone, F. Fedeli, G. Genalti, J. Anninga, H. Gelderblom and M. Fiocco. *Novel longitudinal Multiple Overall Toxicity (MOTox) score to quantify adverse events experienced by patients during chemotherapy treatment: a retrospective analysis of the MRC BO06 trial in osteosarcoma*. *BMJ Open*, 11(12):e053456, 2021 [190].

Chapter 6 is extracted and extended from the following work:

- **M. Spreafico**, F. Ieva and M. Fiocco. *Longitudinal Latent Overall Toxicity (LO-Tox) profiles in osteosarcoma: a new taxonomy based on latent Markov models*. arXiv:2107.12863, 2021 [191]. *[Submitted]*

In Chapter 7 we introduce marginal structural models in combination with Inverse-Probability-of-Treatment Weighted estimators to model the causal effects of chemotherapy intensity exposure on Event-Free Survival (EFS) in presence of time-dependent confounders. Statistical and clinical expertises are merged to propose a suitable characterisation of the causal structure of the chemotherapy data through the introduction of appropriate direct acyclic graphs that identify all possible (time-dependent) confounders (i.e., toxicities and other individual characteristics) and their relationships with exposure and EFS outcome. Data from the control arms of European Osteosarcoma Intergroup studies MRC BO03/EORTC 80861 [120] and MRC BO06/EORTC 80931 [119] for patients with osteosarcoma are analysed. Since drug administration is longitudinal while only the most severe side-effects are recorded, the analysis of such mixed longitudinal/non-longitudinal data requires both an original analytical strategy and an unconventional model formulation. The main contribution of this chapter is the presentation of an all-round analysis of complex chemotherapy data, with tutorial-like explanations of the difficulties encountered and the problem-solving strategies deployed. The content of this chapter is based on the following work:

- **M. Spreafico**, C. Spitoni, C. Lancia, F. Ieva and M. Fiocco. *Causal effects of chemotherapy regimen intensity on survival outcome in osteosarcoma patients through Marginal Structural Cox Models*, 2022. *[Submitted]*

In the final Conclusions we summarise the contributions and achievements of this work from both a statistical and clinical point of view, identifying the added values of the entire thesis from a global perspective.

Codes to perform the analysis is written in R software environment [161] and are available on my personal Github repository (<https://github.com/mspreafico>) or as supplementary material of the relative published papers.

## **PART I**

---

# **Pharmacoepidemiology in Heart Failure**



## CHAPTER 1

---

# A new method for measuring adherence to polypharmacy

This chapter has been published in *American Journal of Cardiovascular Drugs*, 20:179–190, 2020 as M. Spreafico, F. Gasperoni, G. Barbati, *et al.* “Adherence to Disease-Modifying Therapy in Patients Hospitalized for HF: Findings from a Community-Based Study” [187].

Heart failure (HF) is a major and growing public health issue, characterized by high costs, steep morbidity and mortality [129]. Despite the advances in the understanding the pathophysiology of chronic HF and the improvement of therapy, HF mortality and morbidity rates remain high [141, 98]. HF guidelines [139, 221, 50] have consistently focused on the benefits of neurohormonal therapy in HF patients with reduced ejection fraction to delay progression and improve survival. These recommendations also underlined up-titration of neurohormonal doses toward target, when possible, by the time of hospitalization discharge. However, medication non-adherence is a common issue, and it is associated with adverse health conditions and increased economic burden to the healthcare system especially in case of chronic diseases such as HF [158].

Recent observations suggest that up to 50% of early post discharge mortality may be associated with guidelines non-adherence [61]. However, previous epidemiological studies of adherence to polypharmacy have analyzed HF patients from surveys of highly selected populations [225]. Further, these studies were based on physician’s prescriptions [108, 107] regardless of patients’ adherence in the follow up [107]. Several concerns remain on adherence of unselected patients of real world setting to evidence-based HF treatment. To overcome the aforementioned gaps, it is possible to estimate patients’ adherence from drugs purchases, and this is particularly feasible in a public health system using healthcare administrative archives. Worth of note, methods for estimating adherence to single drug classes from drugs purchases are well established [102, 14], whereas there are few studies on patient’s adherence, especially in the setting of polypharmacy [17]. Specifically, in cardiological literature, focusing both on polypharmacy and on adherence is still an open research field [58, 63].

The present chapter aim is to investigate HF patients’ adherence to disease-modifying therapies during the first year after HF hospitalization and to estimate its prognostic impact on survival. First, we describe how evidence-based therapies are applied in a



## 1. A new method for measuring adherence to polypharmacy

real world setting, including the evaluation of target dosages based on drugs purchases. Secondly, we represent polypharmacy adherence during the first year as combinations of prescriptions and adherence to the pharmacological classes of interest. In particular, we introduce a novel method for measuring adherence to polypharmacy by computing the ratio between two quantities: the “*Polypharmacy Adherence*” (PA) and the “*Purchase Indicator*” (PI), so producing the *Patient Adherence Indicator* (PAI). Finally, we evaluate the effect of PAI on survival using different Cox models [46], adjusting for demographic characteristics, comorbidities, re-hospitalisations events and patterns of care in the year following the index HF hospitalization.

## 1.1. Materials and Administrative data

### 1.1.1. Study setting

Between January 2009 and December 2015, patients hospitalized in the *Friuli Venezia Giulia* Italian Region (FVG, a north-eastern region of Italy, with a population of about 1.2 million inhabitants) with a principal diagnostic code of HF and at least one pharmacological purchase of disease-modifying drugs for HF were recruited. Patients who were not inhabitants of the FVG region or were younger than 18 years at the time of hospitalization were excluded. Enrolment occurred from the data of discharge of HF hospitalization.

### 1.1.2. Data sources

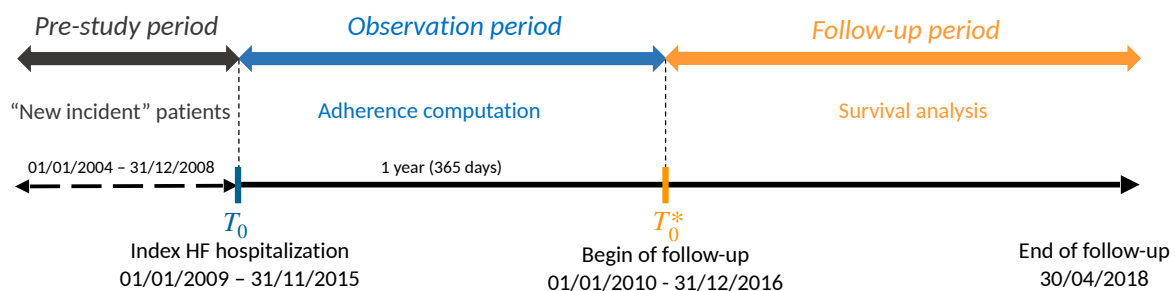
The data of healthcare administrative archives were used for identification of HF patients. The FVG regional Data Warehouse includes various sources of data, such as the Registry of Births and Deaths, Hospital Discharge, the District Healthcare Services (intermediate and home care), Public Laboratories and Public Drug Distribution System, that are object of internal routinely quality checks. Of note, the availability of laboratory analyses performed in public hospitals is a peculiar characteristic of this Region. Each record in the dataset was related to an event, which could be a HF hospitalization or hospitalization for other causes, an activation of Intermediate Care Unit (ICU) service or an Integrated Home Care (IHC). For all these events (admission to hospital or ICU/IHC), we collected dates of admission and discharge. Moreover, for each HF hospitalization, we identified with a binary flag if the patient was discharged from a Cardiological Ward (CW), if a cardiological visit and an echocardiogram were performed. In the Public Drug Distribution System, each record represented a pharmacological purchase characterized by the date of acquisition, ATC (Anatomical Therapeutic Chemical classification system) [220, 214] and AIC codes (authorization code related to Italian market) [3] and the total number of purchased boxes.

### 1.1.3. Study population

HF primary diagnosis included ICD-9CM codes for HF (428, 398.91, 402.01, 402.11, 402.91, 404.01, 404.03, 404.11, 404.13, 404.91 and 404.93) selected according to the National Outcome Evaluation Program. We focused on those patients with a first discharge (associated with a principal diagnostic code of HF) between January 2009 to December 2015 and we excluded those patients who died during the first HF hospitalization. We defined a 5 *years pre-study period* from 2004 to 2008 (Figure 1.1), in order to observe chronic comorbidities and hospitalizations for HF. This allowed us to limit underestimation of chronic comorbidities and to identify new incident HF patients (those with no hospitalizations for HF during the pre-study time-window). The *study-period* was divided into the *observation period* (365 days from the index discharge date) and the *follow-up period*: only patients alive at the end of the observation period were followed up to observe survival outcomes (Figure 1.1). Finally, only patients with at least one pharmacological purchase related to the disease-modifying drugs were included [139, 50]. Specifically, we considered the following drugs: Angiotensin-Converting Enzyme inhibitors (ACE), Angiotensin Receptor Blockers (ARB) – these two considered as a unique class (ACE/ARB), Beta-Blocking (BB) and Anti-Aldosterone agents (AA).

Patients were classified as *Worsening Heart Failure* (WHF) or *De Novo* on the basis of the presence of at least one HF hospitalization in the 5 years preceding the index HF hospitalization (Figure 1.1). Demographic, comorbidities, procedures and laboratory tests performed during hospitalization were considered. Among procedures, we considered only the major ones as Coronary Angiography, Percutaneous Transluminal Coronary Angioplasty (PTCA) (w/out implantation of stent in coronary artery), Coronary Artery Bypass Graft surgery (CABG), implantation of pacemaker, Cardioverter defibrillator or Cardiac Resynchronization Therapy (CRT), Transcatheter Aortic Valve Implantation (TAVI) or percutaneous mitral valve repair with MitraClip device. Finally, the Charlson Comorbidity Index [160] was computed using hospital diagnoses based on ICD-9CM that occurred within 5 years before the hospitalization and integrated with laboratory data and diagnosis recorded at the hospitalization, as previously reported [59].

In order to protect privacy, information retrieved from the different databases were linked



**Figure 1.1.** Study design for a HF patient of the study cohort. The *pre-study period* is used to define “new incident” HF patients. The *observation period* is used for adherence computation. The *follow-up period* is used for survival analysis. The administrative censoring date is April 30<sup>th</sup>, 2018.

## 1. A new method for measuring adherence to polypharmacy

via a single anonymous identification code by institutional technical staff. The reverse process is not possible since the generation code table is not available to the authors. Data analyses were performed by authorized staff only on remotely controlled computer. Any possibility to copy or export datasets was disabled. According to the rules from the Italian Medicines Agency [5], retrospective studies using administrative databases do not require Ethics Committee protocol approval.

## 1.2. Methodologies

### 1.2.1. Target dosages according to guidelines

In order to evaluate if the purchased drug quantity was in line with the expected target dosage, we considered an *observation period* of 365 days starting from the index date and we computed the total purchased milligrams of the main active principles for each pharmacological class of interest. Dividing these quantities by 365, we obtained the mean purchased daily doses (DD) of each active principle. Then, we divided them by the respective target dosages (see Table 1.1) as recommended in the ESC (*European Society of Cardiology*) Guidelines [139, 50] or, for those drugs not included in the guidelines, as prescribed routinely in clinical practice and verified in the Italian Drug Agency's (in Italian: AIFA – *Agenzia Italiana del Farmaco*) website [4]. Thus, we obtained the *standardized Daily Doses* (sDD) that patients assumed during the observation period:

$$\text{sDD} = \frac{\text{mean purchased daily dose (DD) during observation period}}{\text{target dose recommended in ESC or AIFA guidelines}}. \quad (1.1)$$

If the sDD was 100% (i.e., sDD = 1) the mean purchased DD was equal to the perfect target, whereas if it was < or > 100%, it was less or higher than the perfect target, respectively.

### 1.2.2. Adherence measures

Medication adherence is generally defined as the process by which patients take their medications as prescribed and three different constructs could be analyzed, i.e., *initiation* of therapy, *implementation* of the dosing regimen and *persistence* with treatment [212]. In the present chapter, we focused on *implementation*, according with the review paper [212], basing our analysis on purchased drugs instead of prescribed drugs. According to [102, 14] we calculated two measure of adherence, i.e., the Proportion of Days Covered (PDC), defined as:

$$\text{PDC} = \frac{\text{number of distinct coverage days}}{\text{number of days in the observation period}} \quad (1.2)$$

and the Medical Possession Ratio (MPR):

$$\text{MPR} = \frac{\text{number of days supply during observation period}}{\text{number of days in the observation period}}. \quad (1.3)$$

**Table 1.1.** Target dosages of each active principle recommended in the ESC (*European Society of Cardiology*) [139, 50] or AIFA (*Agenzia Italiana del Farmaco*) guidelines [4].

Pharmacological class	Active principle	Daily target dose [mg]	Guideline
<b>Anti-Aldosterone agents</b>	Canrenone	50	AIFA
	Potassium Canrenoate	50	AIFA
	Spironolactone	25	ESC
<b>Angiotensin-Converting Enzyme inhibitors</b>	Enalapril	20	ESC
	Lisinopril	20	ESC
	Ramipril	10	ESC
<b>Angiotensin Receptor Blockers</b>	Candesartan	32	ESC
	Losartan	150	ESC
	Olmesartan	40	AIFA
	Telmisartan	80	AIFA
	Valsartan	320	ESC
<b>Beta-Blocking agents</b>	Bisoprolol	10	ESC
	Carvedilol	50	ESC
	Metoprolol	200	ESC
	Nebivolol	10	ESC

The distinction between PDC and MPR consists in the numerator that is different in case of overlapping of two subsequent purchases. In particular, through PDC we considered the period covered by the first purchase entirely and the second purchase only in those days that were not covered by the first one. Conversely, through MPR we shifted the second purchase at the day after the end of the first one, preserving the duration of all purchases.

These measures were dichotomized to identify as adherent those patients with a PDC (or MPR) at least 80% [14]. For adherence computation of each of the disease-modifying pharmacological class (ACE/ARB, BB, AA) an observation period of 365 days from the index date was considered [102]. If during the observation period a patient was re-hospitalized or spent some time in ICU, we assumed that he/she was under treatment, i.e., he/she was taking all the purchased types of drug during those periods.

### 1.2.3. Adherence to polypharmacy

In order to evaluate polypharmacy, we introduced a new index, the *Patient Adherence Indicator* (PAI), based on the ratio between the *Polypharmacy Adherence* (PA) and the *Purchase Indicator* (PI). These measures are computed using observed combinations of the three pharmacological classes of interest: BB, AA and ACE or ARB. PI is defined as the number of purchased types of drug at least once and it could be 1, 2 or 3 based on patient's different purchases:

$$\text{PI} = \text{number of purchased types of drug at least once.} \quad (1.4)$$

## 1. A new method for measuring adherence to polypharmacy

PA is the number of pharmacological classes to which the patient is adherent at the defined threshold of 80% (0, 1, 2 or 3):

$$PA = (\text{adherent to ACE or to ARB}) + \text{adherent to BB} + \text{adherent to AA}. \quad (1.5)$$

Finally, PAI is the number of pharmacological classes to which the patient is adherent divided by the number of purchased types of drug:

$$PAI = \frac{PA}{PI}. \quad (1.6)$$

PAI considers adherence to polypharmacy and it could be 0, 1/3, 1/2, 2/3 or 1 (3/3). Based on the overall PAI percentage, patients were divided into two groups: those with *poor* adherence percentage to polypharmacy ( $PAI < 50\%$ , i.e.,  $< 1/2$ ) and *good* adherence percentage ( $PAI \geq 50\%$ , i.e.,  $\geq 1/2$ ).

### 1.2.4. Outcome measure

Study outcome of interest was patient's death for any cause. Deaths were collected from the Registry of Birth and Deaths included in the regional Data Warehouse. For the survival analysis, each patient was followed from one year after the index date (i.e., one year after the discharge from the index HF hospitalization –  $T_0^*$  in Figure 1.1) until the end of the study or the date of death (see *follow-up period* in Figure 1.1). The administrative censoring date was April 30<sup>th</sup>, 2018.

### 1.2.5. Survival Analysis: multivariable Cox regression models

In order to assess the role of polypharmacy adherence with respect to the overall survival time of a patient, we estimated four different Cox regression models [46], one for each of the following polypharmacy indices: PAI and PAI group, computed with both PDC and MPR adherence measures. Each model was adjusted for nine covariates: WHF condition (**whf**) and discharge from CW at the index hospitalization (**cw**), cardiological visit in 24 months before the last hospitalization of the observation period (**cardio**), number of re-hospitalizations (**rehosp**), number of ICU services (**icu**) and IHC activation during the observation period (**ihc**), Charlson index at the last hospitalization of the observation period (**charlson**), age (**age**) and gender (**gender**) at the beginning of the follow-up. The choice of these covariates was driven by clinical relevance and availability from administrative data. The hazard functions for each patient  $i$  were hence given by:

$$h_i(t|\boldsymbol{\omega}_i) = h_0(t) \exp \{ \boldsymbol{\theta}^T \boldsymbol{\omega}_i \} \quad (1.7)$$

where the covariate vector for each patient was

$$\boldsymbol{\omega}_i = (\text{whf}_i, \text{age}_i, \text{gender}_i, \text{charlson}_i, \text{cw}_i, \text{cardio}_i, \text{rehosp}_i, \text{icu}_i, \text{ihc}_i, \omega_{10,i})$$

with polypharmacy index  $\omega_{10,i}$  equal to

$$\text{PAI\_PDC}_i \text{ or } \text{PAI\_MPR}_i \text{ or } \text{PAIgroup\_PDC}_i \text{ or } \text{PAIgroup\_MPR}_i.$$

All the analyses were carried out using the free software R [161], in particular survival package [201, 202]. Covariates with p-values  $< 0.05$  were considered statistically significant.

## 1.3. Results

Patient characteristics are presented as numbers and percentages for categorical variables. For continuous variables we reported means with standard deviations or medians with interquartile ranges (IQRs), as appropriate depending on the distribution shape.

### 1.3.1. Cohort selection

A total cohort of 20,622 patients were identified with principal diagnostic code of HF. Of these, we excluded 13 paediatric patients. A substantial portion of patients (6,505, 32%) was not considered because they died during the first year after the index hospitalization. Moreover, 1,020 patients (5%) were removed since they did not present any purchase of ACE, ARB, BB or AA during the observation period. Further, since their health residence district was not in FVG region, other 146 (0.7%) patients were excluded. Thus, a total of 12,938 (63%) patients met study selection criteria (Figure 1.2).

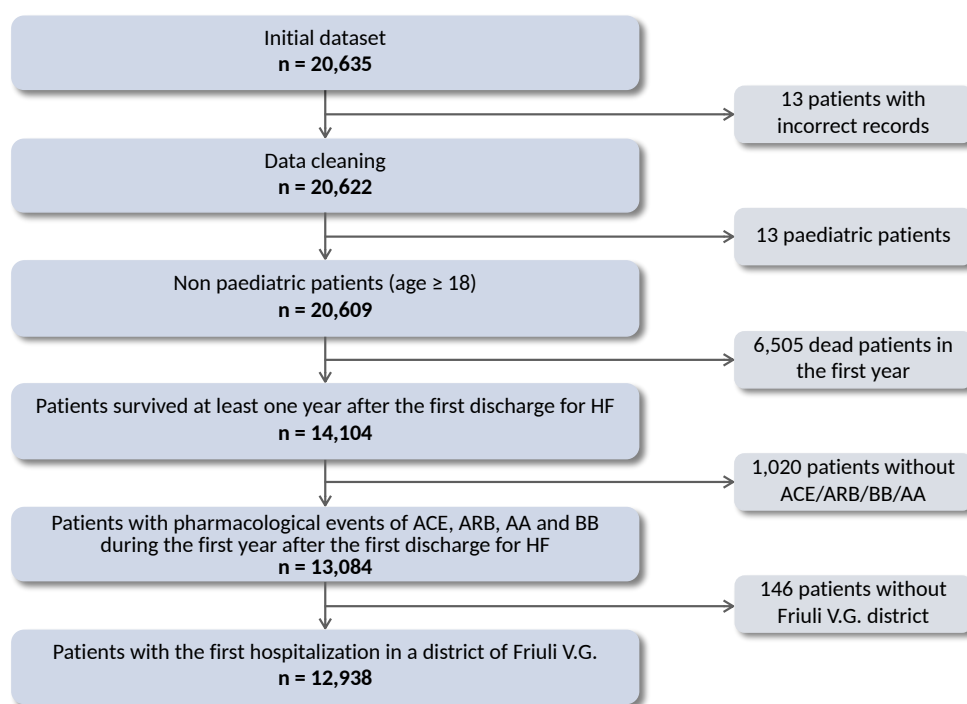


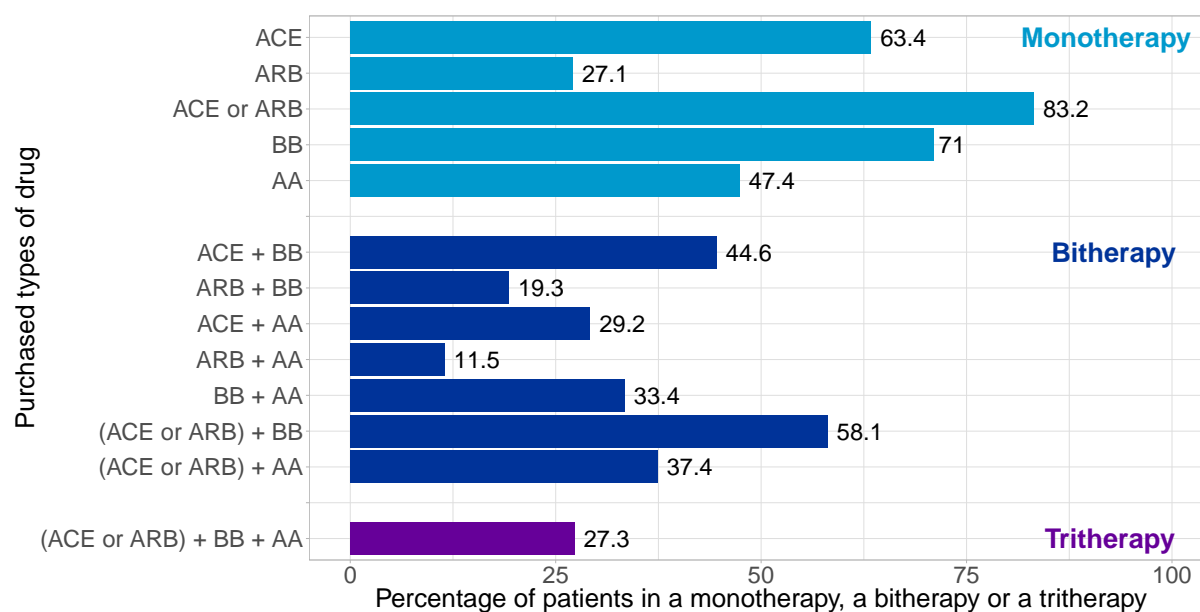
Figure 1.2. Flowchart of patient selection.

## 1. A new method for measuring adherence to polypharmacy

Overall, at index hospitalization (Table 1.2) mean age was 80 years with a substantial proportion of female patients (53.1%), high prevalence of *De Novo* patients (89.1%). Percentage of patients who have undergone at least one major procedure was 3.2%. Comorbidity burden was high (median of Charlson index 2; 46.8% of patients presenting Charlson index  $\geq 3$ ). The rate of discharge from Cardiological Ward (CW) was 10.3%. In the 24 months before the index hospitalization, 6,030 (46.6%) patients underwent a cardiological visit and 3,212 (24.8%) an echocardiogram.

Regarding pharmacological treatments, Figure 1.3 shows percentages of purchase of medications at discharge according to monotherapy, dual therapy or triple therapy. In monotherapy the most common purchased drugs were BB (71%) and the less ones were ARB (27.1%, light-blue columns); ACE or ARB (ACE/ARB) was purchased by 83.2% of patients. Regarding polypharmacy, the most common prescribed drugs were ACE or ARB and BB (58.1%) and the less frequent were ARB and AA (11.5%, blue columns). Finally, the triplet ACE or ARB, BB and AA was purchased by 27.3% of patients (purple column).

At the end of the observation period, i.e., one year after the index HF hospitalization (Table 1.3), mean age was 81 years and the median of Charlson index remained high (median of Charlson index 2; 47.4% with a Charlson index  $\geq 3$ ). Starting from the end of the observation period, during a median follow-up of 33 (IQR = [17.1; 55.1]) months, 7,752 (59.9%) patients died. In the 24 months before the last hospitalization of the observation period, 6,786 (52.5%) patients underwent a cardiological visit and 4,227 (32.7%) an echocardiogram. In the observation period, 53.6% patients were re-hospitalized



**Figure 1.3.** Barplots of percentages of patients in a monotherapy, a bitherapy or a tritherapy. Each column is related to the purchase of specific types of drug (i.e., ACE, ARB, ACE/ARB, BB, AA). On the top, light-blue columns show percentages about monotherapy, where ‘ACE or ARB’ means that a patient presents at least one purchase for ACE and/or ARB during the observation period. Central dark-blue columns show percentages about bitherapy and ‘ACE + BB’ means that a patient presents at least one purchase both for ACE and for BB during the observation period. On the bottom, the purple column shows the percentage about tritherapy and states that 27.3% of the whole cohort purchased ACE and/or ARB, BB and AA.

**Table 1.2.** Descriptive analysis of the whole cohort at index HF hospitalization.

<b>Variable characteristics at index HF hospitalization</b>	
<b>Study Cohort</b>	12,938 pts
<b>Age [year]</b>	
Mean (s.d.)	79.77 (9.62)
<b>Gender</b>	
<i>Female</i> (%)	6,875 (53.1%)
<i>Male</i> (%)	6,063 (46.9%)
<b>HF Condition</b>	
<i>De Novo</i> (%)	11,531 (89.1%)
<i>Worsening</i> (%)	1,407 (10.9%)
<b>Number of procedures*</b>	
0 (%)	12,440 (96.1%)
1 (%)	411 (3.2%)
2 (%)	62 (0.5%)
3 (%)	25 (0.2%)
≥ 4(%)	0 (0%)
<b>Charlson index</b>	
median (IQR)	2 (1; 4)
< 3	6,878 (53.2%)
≥ 3	6,060 (46.8%)
<b>Cardiological Ward</b>	
No (%)	11,602 (89.7%)
Yes (%)	1,336 (10.3%)
<b>Cardiological visit</b>	
No (%)	6,908 (53.4%)
Yes (%)	6,030 (46.6%)
<b>Echocardiogram</b>	
No (%)	9,726 (75.2%)
Yes (%)	3,212 (24.8%)
<b>Creatinine**</b>	
Median (IQR)	1.09 (0.89; 1.38)
Missing values (%)	1,599 (12.4%)
<b>Glycated haemoglobin**</b>	
Median (IQR)	6.6 (6.0; 7.5)
Missing values (%)	10,084 (77.9%)
<b>Haemoglobin**</b>	
Median (IQR)	12.3 (11.0; 13.6)
Missing values (%)	3,879 (30.0%)

Age, gender, number of procedures, Charlson index, laboratory tests and discharge from CW refer to the first event, the index hospitalization. Cardiological visit and echocardiogram refer to the 24 months before the index hospitalization. HF condition refers to the 5 years preceding the index admission.

\* Examined major procedures: coronary angiography, Percutaneous Transluminal Coronary Angioplasty (PTCA) (w/out implantation of stent in coronary artery), Coronary Artery Bypass Graft surgery (CABG), implantation of pacemaker, cardioverter defibrillator or Cardiac Resynchronization Therapy (CRT), Transcatheter Aortic Valve Implantation (TAVI) or percutaneous mitral valve repair with MitraClip device. For the descriptive of each procedure, see supplementary material of Spreafico *et al.* (2020) [187].

\*\* Laboratory tests: median values (if available) of creatinine, glycated haemoglobin and haemoglobin measured during the index hospitalization. Creatinine and glycated haemoglobin values were integrated to hospital diagnosis in the Charlson index computation.



**Table 1.3.** Descriptive analysis of the whole cohort at the beginning of follow-up period.

<b>Variable characteristics at index HF hospitalization</b>	
<b>Study Cohort</b>	12,938 pts
<b>Age [year]</b>	
Mean (s.d.)	80.77 (9.62)
<b>Follow-up time [months]</b>	
Median (IQR)	33 (17.1; 55.1)
<b>Death</b>	
0 (%)	5,186 (40.1%)
1 (%)	7,752 (59.9%)
<b>Charlson index*</b>	
median (IQR)	2 (1; 4)
< 3	6,801 (52.6%)
≥ 3	6,137 (47.4%)
<b>Cardiological visit**</b>	
No (%)	6,152 (47.5%)
Yes (%)	6,786 (52.5%)
<b>Echocardiogram***</b>	
No (%)	8,711 (67.3%)
Yes (%)	4,227 (32.7%)
<b>Number of all-cause re-Hospitalizations</b>	
0 (%)	6,006 (46.4%)
1 (%)	3,462 (26.8%)
2 (%)	1,775 (13.7%)
≥ 3 (%)	1,695 (13.1%)
<b>Number of HF re-Hospitalizations</b>	
0 (%)	10,422 (80.7%)
1 (%)	1,896 (14.7%)
2 (%)	437 (3.4%)
≥ 3 (%)	163 (1.2%)
<b>Number of ICU services</b>	
0 (%)	11,356 (87.7%)
1 (%)	1,305 (10.1%)
≥ 2 (%)	277 (2.2%)
<b>IHC activation</b>	
No (%)	8,718 (67.4%)
Yes (%)	4,220 (32.6%)
<b>Dead patients cohort</b>	7,752 pts
<b>HF Condition‡</b>	
<i>De Novo</i> (%)	6,563 (84.7%)
<i>Worsening</i> (%)	1,189 (15.3%)

ICU = Intermediate Care Unit, IHC = Integrated Home Care.

Age and gender refer to the end of the observation period (i.e., 365 days after the index hospitalization). Charlson index refers to the last hospitalization during the observation period. Cardiological visit and echocardiogram refer to the 24 months before the last hospitalization of the observation period. Number of re-hospitalizations, number of ICU and IHC activation refer to the observation period.

\* Wilcoxon test with respect to the index date: p-value < 0.0001

\*\* McNemar test on paired proportions with respect index date: p-value < 0.0001

\*\*\* McNemar test on paired proportions with respect index date: p-value < 0.0001

‡ Chi-square p-value < 0.0001

at least once for any-cause, 13.7% for two times and 13.1% more than two times. In particular, 19.3% of patients were re-hospitalized at least once for HF. Moreover, 12.3% of patients was admitted in ICU and IHC was activated at least once in 32.6% (4,220 patients) of the study cohort. Of note, for patients without any re-hospitalization during the observation period, the last hospitalization coincided with the index hospitalization.

### 1.3.2. Standardized Daily Dose

Figure 1.4 shows the standardized Daily Dose (sDD) of the main active principles under study. We investigated ramipril, enalapril and lisinopril among ACE; losartan, valsartan, olmesartan, telmisartan and candesartan among ARB; spironolactone, potassium canrenoate and canrenone among AA; bisoprolol, carvedilol, metoprolol and nebivolol among BB. Figure 1.4 shows distribution of sDD in our cohort by means of boxplots. To put the target dosages in evidence, we considered both 100% (blue lines) and 80% (orange lines). Considering 100% as the perfect target could be inappropriate, given the existence of some dynamical processes like up-titration of the drugs dosages that cannot be investigated through these data. So, we decided to consider also 80% as target in order to take into account these unknown processes and having a more realistic estimate of patients that actually reach target dosages. Percentages of patients with sDD > 80% were 11.6%, 28.9% and 32.3%, for ramipril, enalapril and lisinopril, respectively. Percentages of patients with sDD > 80% were 1.8%, 8.8%, 10%, 30.5% and 10.6%, for losartan, valsartan, olmesartan, telmisartan and candesartan, respectively. Percentages of patients with sDD > 80% were 34.2%, 43.7% and 29.5%, for spironolactone, potassium canrenoate and canrenone, respectively. Percentages of patients with sDD > 80% were 4.3%, 10.4%, 14.4% and 2.4%, for bisoprolol, carvedilol, metoprolol and nebivolol, respectively.

### 1.3.3. Patients' adherence measures

Using PDC, at the end of the observation period 47.2% of 8,199 ACE patients, 39.7% of 3,503 ARB patients, 22.6% of 9,183 BB patients, 18.3% of 6,137 AA patients and 48.5% of 10,759 of ACE or ARB patients were adherent to the corresponding treatment at the threshold of 80% (see Table 1.4). Using MPR measure, percentages were higher: 63% of ACE patients, 58.5% of ARB patients, 36% of BB patients, 31.5% of AA patients and 66% of ACE or ARB patients.

Descriptive statistics about adherence to polypharmacy indices of the study cohort are reported in Table 1.5. Using PDC, the following PAI values emerged: 47.2% (0: non-adherent patients), 11.1% (1/3), 20.5% (1/2), 5.1% (2/3) and 16.1% (1: fully adherent patients). Consequently, 41.7% of the patients had *good* percentage of adherence to polypharmacy (n = 5,393). Using MPR, the following PAI values were calculated: 29.1% (0: non-adherent patients, n = 3,758), 11.4% (1/3), 23.8% (1/2), 8.8% (2/3) and 26.9% (1: fully adherent patients). Consequently, 59.5% of the patients had *good* percentage of adherence to polypharmacy (n = 7,700). Of note, in Table 1.6 we provided information

## 1. A new method for measuring adherence to polypharmacy

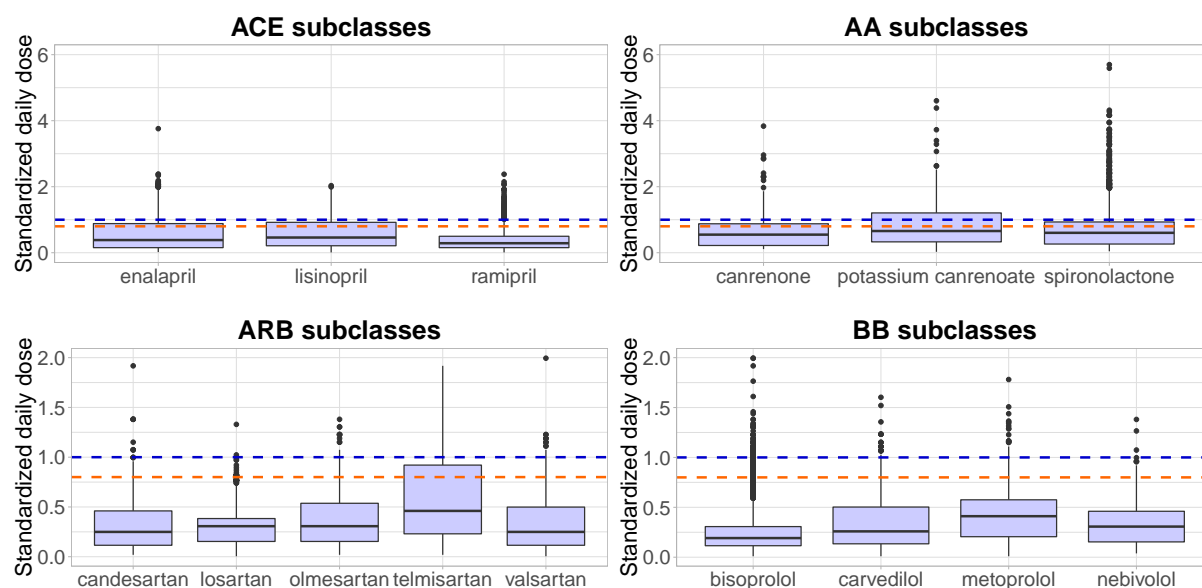
about the number of purchased drugs and the percentage of poly-adherent patients. As expected, we observe that the higher is the number of therapies, the lower is the poly-adherence (both p-values are less than 0.0001).

**Table 1.4.** Numbers and percentages of adherent patients to the corresponding treatment at the threshold of 80% using both PDC and MPR measures.

Adherent		ACE	ARB	BB	AA	ACE/ARB
<i>Cohort</i>	no. pts	8,199	3,503	9,183	6,137	10,759
<b>PDC</b>	No (%)	4,325 (52.8%)	2,112 (60.3%)	7,110 (77.4%)	5,015 (81.7%)	5,544 (51.5%)
	Yes (%)	3,874 (47.2%)	1,391 (39.7%)	2,073 (22.6%)	1,122 (18.3%)	5,215 (48.5%)
<b>MPR</b>	No (%)	3,030 (37.0%)	1,454 (41.5%)	5,874 (64.0%)	4,202 (68.5%)	3,661 (34.0%)
	Yes (%)	5,169 (63.0%)	2,049 (58.5%)	3,309 (36.0%)	1,935 (31.5%)	7,098 (66.0%)

**Table 1.5.** Descriptive analysis of Patient Adherence Indicators (PAIs) of the whole cohort.

PP Index	PP scale	PDC	MPR
<b>PAI</b>	0 (%)	6,107 (47.2%)	3,758 (29.1%)
	1/3 (%)	1,438 (11.1%)	1,480 (11.4%)
	1/2 (%)	2,653 (20.5%)	3,080 (23.8%)
	2/3 (%)	654 (5.1%)	1,139 (8.8%)
	1 (%)	2,086 (16.1%)	3,481 (26.9%)
<b>PAI group</b>	<i>good</i> (%)	5,393 (41.7%)	7,700 (59.5%)
	<i>poor</i> (%)	7,545 (58.3%)	5,238 (40.5%)



**Figure 1.4.** Boxplots of standardized Daily Dose (sDD) for the main active principles of each pharmacological class. Top-left panel reports ACE main subclasses: enalapril, lisinopril and ramipril. Top-right panel reports AA main subclasses: canrenone, potassium canrenoate and spironolactone. Down-left panel reports ARB main subclasses: candesartan, losartan, olmesartan, telmisartan and valsartan. Down-right panel report BB main subclasses: bisoprolol, carvedilol, metoprolol and nebivolol. Dashed blue lines (standardized daily dose = 100%) indicate that the mean purchased DD are equal to the respective target dosages recommended in the ESC Guidelines [139, 50] or according to clinical practice of AIFA's website [4]. Dashed orange lines (standardized daily dose = 80%) indicate that the mean purchased DD are equal to the 80% of the respective target dosages.

**Table 1.6.** Numbers and percentages of Poly-Adherence (PA), computed both with PDC and MPR, with respect to number of different types of purchased drugs (i.e., PI, Purchase Indicator).

		Poly-Adherence (PA) with PDC				
		no. pts	0	1	2	3
Purchase Indicator (PI)	1	3,331	2,068 (62.1%)	1,263* (37.9%)	Not possible	Not possible
	2	6,073	2,699 (44.4%)	2,653 (43.7%)	721 * (11.9%)	Not possible
	3	3,534	1,438 (37.9%)	1,438 (40.7%)	654 (18.5%)	102 * (2.9%)
		Poly-Adherence (PA) with MPR				
		no. pts	0	1	2	3
Purchase Indicator (PI)	1	3,331	1,577 (47.3%)	1,754 ** (52.7%)	Not possible	Not possible
	2	6,073	1,562 (25.7%)	3,080 (50.7%)	1,431 ** (23.6%)	Not possible
	3	3,534	619 (17.5%)	1,480 (41.9%)	1,139 (32.2%)	296 ** (8.4%)

\* Test for proportions of global Poly-Adherence (1263/3331, 721/6073, 102/3534): p-value < 0.0001.

\*\* Test for proportions of global Poly-Adherence (1754/3331, 1431/6073, 296/3534): p-value < 0.0001.

### 1.3.4. Multivariable Cox models for survival outcome

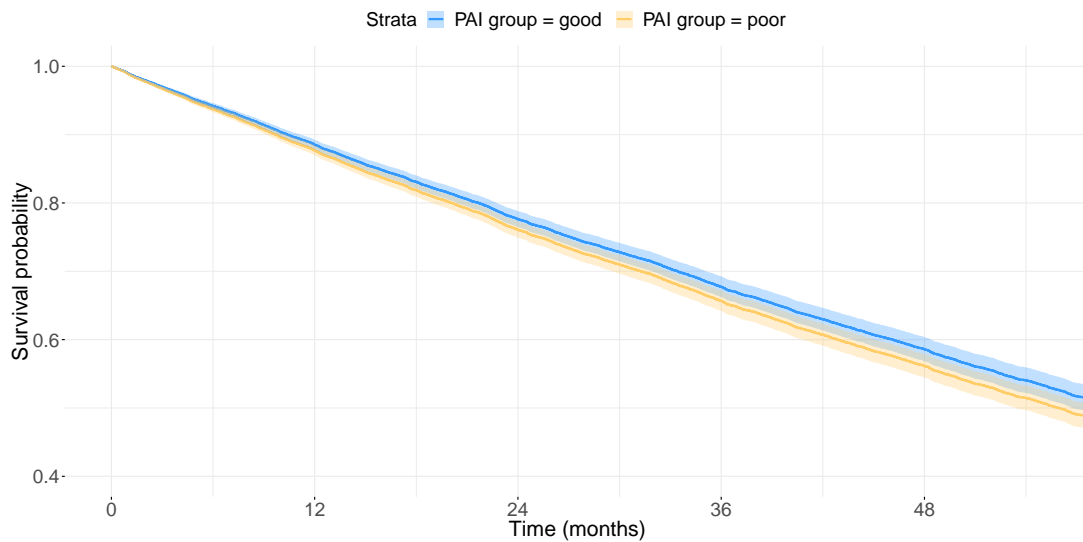
In Table 1.7 impact of covariates on survival for each Cox model is displayed. Among risk factors we identified: WHF, age, Charlson score, re-hospitalizations, ICU and IHC. Specifically, being a WHF patient with respect to a De Novo patient, being elder, having a higher Charlson index, being re-hospitalized more often, being admitted in ICU and the activation of IHC implied a higher risk of death. Conversely, among protective factors we identified: the discharge from a Cardiological Ward (CW) in the index hospitalization and a cardiological visit in the 24 months before the last hospitalization of the observation period. Regarding polypharmacy indices, both PAI (first and second model) and PAI group (third and fourth models) were significantly protective (HRs < 1). In particular, higher values of PAI and being labelled as *good* in case of PAI group were associated with a lower risk of death.

Figure 1.5 shows this result through the estimate of a survival curve stratified by *good* and *poor* levels in the case of PAI group computed using PDC (third model) for a hypothetical patient that should be representative of the studied cohort. Specifically, we considered a 82-years old, female, De Novo patient with a previous cardiological visit and a Charlson index (at the last hospitalization) equal to 2. Moreover, at index HF hospitalization this patient was not discharged from CW and during the observation period she was re-hospitalized only once and did not benefit of any ICU service or IHC activation. These values correspond to the medians of the continuous variables and the modes of the categorical variables measured in our cohort.

**Table 1.7.** Adjusted Hazard Ratios with 95% Confidence Intervals (CI) and p-values of each Cox's model. Each column corresponds to a different Cox regression models, one for each of the following polypharmacy indices: PAI and PAI group computed with both PDC and MPR adherence results.

	Model 1 – PDC		Model 2 – MPR	
	HR (95% CI)	p-value	HR (95% CI)	p-value
<b>HF Condition [WHF]</b>	1.24 [1.16; 1.32]	6.77e-11	1.24 [1.16; 1.32]	7.33e-11
<b>Age</b>	1.06 [1.06; 1.07]	< 2e-16	1.06 [1.06; 1.07]	< 2e-16
<b>Gender (M)</b>	1.32 [1.26; 1.39]	< 2e-16	1.32 [1.26; 1.39]	< 2e-16
<b>Charlson index</b>	1.11 [1.09; 1.12]	< 2e-16	1.11 [1.09; 1.12]	< 2e-16
<b>CW</b>	0.78 [0.71; 0.85]	1.97e-07	0.78 [0.71; 0.86]	2.39e-07
<b>Cardiological visit</b>	0.94 [0.89; 0.98]	0.00439	0.94 [0.90; 0.98]	0.00563
<b>Re-hospitalizations</b>	1.11 [1.10; 1.13]	< 2e-16	1.12 [1.10; 1.13]	< 2e-16
<b>ICU services</b>	1.14 [1.09; 1.20]	4.18e-08	1.14 [1.09; 1.20]	3.12e-08
<b>IHC activation</b>	1.27 [1.22; 1.34]	< 2e-16	1.28 [1.22; 1.34]	< 2e-16
<b>PAI</b>	0.91 [0.85; 0.97]	0.00270	0.94 [0.89; 0.99]	0.03819
	Model 3 – PDC		Model 4 – MPR	
	HR (95% CI)	p-value	HR (95% CI)	p-value
<b>HF Condition [WHF]</b>	1.24 [1.16; 1.32]	9.33e-11	1.24 [1.16; 1.32]	7.97e-11
<b>Age</b>	1.06 [1.06; 1.07]	< 2e-16	1.06 [1.06; 1.07]	< 2e-16
<b>Gender (M)</b>	1.32 [1.26; 1.39]	< 2e-16	1.32 [1.26; 1.39]	< 2e-16
<b>Charlson index</b>	1.11 [1.09; 1.12]	< 2e-16	1.11 [1.09; 1.12]	< 2e-16
<b>CW</b>	0.78 [0.71; 0.85]	1.63e-07	0.78 [0.71; 0.85]	1.87e-07
<b>Cardiological visit</b>	0.94 [0.89; 0.98]	0.00466	0.94 [0.90; 0.98]	0.00572
<b>Re-hospitalizations</b>	1.11 [1.10; 1.13]	< 2e-16	1.11 [1.10; 1.13]	< 2e-16
<b>ICU services</b>	1.14 [1.09; 1.20]	3.73e-08	1.14 [1.09; 1.20]	3.20e-08
<b>IHC activation</b>	1.27 [1.22; 1.34]	< 2e-16	1.28 [1.22; 1.34]	< 2e-16
<b>PAI group (good)</b>	0.93 [0.88; 0.97]	0.00119	0.93 [0.89; 0.98]	0.00354

Each model was adjusted for nine time-independent covariates: HF condition and discharge from CW at index hospitalization, cardiological visit in the 24 months before the last hospitalization of the observation period, number of re-hospitalizations, number of ICU services and IHC activation during the observation period, Charlson index at the last hospitalization, age and gender at the beginning of the follow-up period.



**Figure 1.5.** Estimated survival from the Cox model stratified by *good* and *poor* patients in the case of PAI group computed using PDC. Adjusting covariate values are the medians of the continuous variables and the modes of the categorical variables measured in our cohort: female-De Novo patient aged 82 years old with a Charlson index at the last hospitalization equal to 2; at index hospitalization this patient was not discharged from Cardiology Ward; during the observation period she was re-hospitalized only once time, she did not benefit of ICU service or IHC activation, and she underwent one cardiological visit.

## 1.4. Final remarks

The goal of Drug Utilization Research [220] is to facilitate the rational use of drugs in patient populations. A key point emerged from clinical trials is that the prescription of drugs should be in the “optimal” dose for the therapeutic indication. To date, few data exist regarding the adherence to drug therapies in a real world setting for HF patients. Indeed, most of previous data have been focused on physician’s prescription adherence to recommended medications in HF patients [108, 130]. In the present chapter, we took advantage of real-world pharmacological records about drugs purchases in order to have a proxy of patient’s adherence to polypharmacy.

Our study confirmed that, even when prescriptions of guideline-based HF treatment are high, there is evidence of frequent failures to reach target doses [130]. In fact, results showed high proportion of HF patients were treated with low dosages of recommended therapies: mean daily dosages purchased by patients were well below the target dosages for all the drugs considered. Data from quality surveys reported similar trend in the prescriptions, with less than one-third of patients on guideline-recommended target dosages [108]. Similarly, a recent European survey (BIOSTAT-CHF) conducted in 11 countries and enrolling 2,500 patients showed that only a minority of patients reached the target dose of ACE and BB [148]. These trends confirmed that simple calculation of the percentage of “treated” patients based on physician’s prescription might not be an adequate measure to indicate the quality of healthcare provided for HF patients.

In addition, patients’ adherence to oral treatment of HF medications was widely unsatisfactory, in particular taking into account the PDC approach (47% of non-adherent

## 1. A new method for measuring adherence to polypharmacy

patients, PAI index based on PDC). This is in contrast to physician's prescriptions of appropriate classes of therapy that instead has improved considerably over the past decade, from approximately a quarter of prescriptions in 2008 to nearly two thirds in 2016 [107]. In particular, we focused on patient's adherence in 1-year from an HF hospitalization. In this observation period, we found that treatment with the association of oral BB, ACE/ARB and AA was present in one-third of patients, and percentages of patients adherent to only one or two drugs out of three prescribed ranged from 16% to 20% (PAI index computed for 1/3 or 2/3 cases).

Our study indicated that the risk of death significantly decreased in presence of a good adherence to polypharmacy. Importantly, we reported adherence measures on effective purchases exploiting the potential of administrative healthcare databases. The proposed index PAI can be viewed as a modified version of the Guideline Adherence Indicator, GAI, [225] based only on physician's prescriptions at discharge and not on effective patient's purchases and adherence patterns. The significant PAI effect on survival suggests that medication non-adherence is associated with lower survival probability also in the case of polypharmacy therapy, so extending previous results about the effect of non-adherence to specific drugs classes [148].

The two methods of adherence calculations – based on PDC and MPR – showed some relevant differences in terms of percentages of adherent patients for the specific single-drug classes. These differences are due to the fact that adherence could be underestimated by measures which ignore overlaps (i.e., PDC) and overestimated by ones which count overlaps (i.e., MPR) during the observation period. In the current literature, PDC has been suggested as the preferred method to reflect adherence of patients who are prescribed multiple medications concurrently within a class [133] and recently a modified MPR calculation has been proposed in the context of polypharmacy [17]. Noteworthy, we did not observe relevant prognostic differences when PDC or MPR-based measures were combined in the PAI index. This could indicate that adherence to drugs combinations is prognostically more relevant, irrespective from the single-class measure adopted.

To the best of our knowledge, only one very recent paper published in 2018 tackled the issue of polypharmacy adherence in a cardiovascular setting and proposed a novel index [17]. Specifically, they introduced a new “daily polypharmacy possession ratio” (DPPR) based on Australian pharmaceutical benefits scheme database. This work [17] is of great interest, however it is not clear how much the obtained results are affected by the initial strong selection procedure; indeed, they excluded more than half of the cohort based on patients age. This procedure would lead to a not negligible selection bias in our cohort. In line with previous real-world studies focused on HF [59, 94, 93] our population included a high proportion of elderly patients and women, with high rates of non-cardiac comorbidities. Of note, this could also explain the large proportion of HF patients discharged from non-cardiological ward.

Some limitations of the present study have to be noted. First of all, in the healthcare administrative archives no socio-economic data and no clinical data about New York Heart Association (NYHA) class or Left Ventricular Ejection Fraction (LVEF) were available.

Therefore, it was not possible to stratify patients according to clinical HF severity; as a proxy, we used previous HF hospitalization in the patient history. Another limitation is the absence of a specific analysis on diuretics adherence. Although the topic is intriguing, patients' adherence to diuretics is hard to evaluate without clinical data, since their modulation to fluid overload. Since the lack of physician's prescriptions (at discharge and during follow-up) and the possible development of drugs adverse reactions, several concerns remain on the estimated rate of patient's non-adherence. Dedicated future studies are encouraged on this topic integrating clinical and administrative sources of data. Then, in the PDC and MPR computations, theoretical Defined Daily Doses (DDD) were used instead of Prescribed Daily Doses (PDD) and therefore a bias could be present in the estimated adherence if the underlying PDD/DDD ratio is different from 1 [220, 69]. Moreover, the cut-off of PDC or MPR greater than 80% to define patient adherence could be further examined in a sensitivity analysis in order to find if other values or a distribution of thresholds could better stratify patient's outcome. Finally, some technical improvements may be included into the PAI definition in order to provide a more elaborated formula which is able to reward more patients that are adherent to polypharmacy with respect to those adherent in monotherapy. In the present definition of PAI a patient scores 1 if he/she is fully adherent to only one drug or if he/she is fully adherent to a combination of drugs. For example, the PAI definition could be modified combining the terms with some weights, in order to maximize the predictive capacity of the model and producing a more refined grading score among patients.

To summarize, the main purpose of the present chapter was to describe guidelines compliance in a real-world HF community and evaluate the impact of combined drugs adherence on survival. Patients' adherence remained widely unsatisfactory especially taking into account attainment of target dosages and polypharmacy. Adjusting for patient's characteristics and intermediate events, good adherence to polypharmacy in the first year after HF hospitalization was associated with improved survival, irrespective of the specific measure of single drug class adherence used. Although the PAI index represents a first step forward in the assessment of adherence to polypharmacy using real-world data, it exploits the PDC/MPR measures which are *time-fixed* indexes computed at the end of the observation period, without taking into account changes in patient drug utilization behaviour over time. In addition to discarding valuable information, this can lead to selection bias due to the exclusion from the study cohort of patients who did not survive the first 1-year period. To overcome this problem, it is necessary to explore alternative statistical approaches that model adherence as a time-varying covariate, as we shall see in the next chapter.





# Joint modelling of time-varying adherence to medication and survival

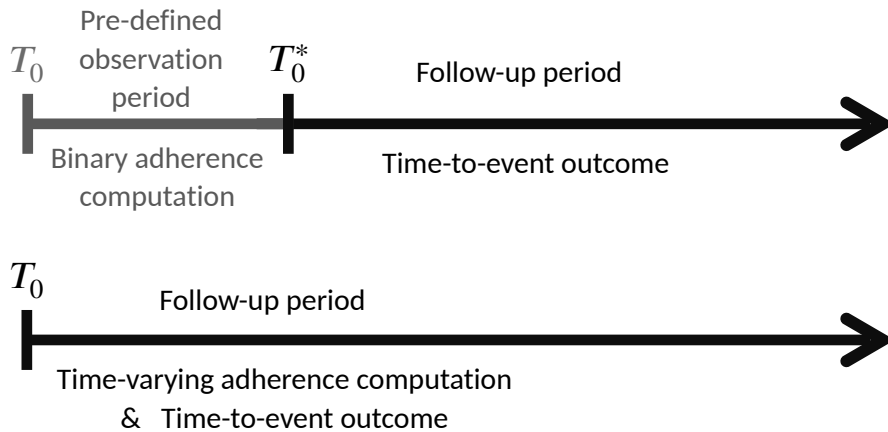
This chapter has been published in *Biometrical Journal*, 63(2):305–322, 2021 as M. Spreafico and F. Ieva “Dynamic monitoring of the effects of adherence to medication on survival in Heart Failure patients: a joint modelling approach exploiting time-varying covariates” [188].

In pharmacoepidemiology literature and current practice, the way adherence to medication is computed and accounted for into predictive models is far from being informative as it may be. As shown in the previous chapter, the most used adherence measures [14, 102] are computed over a pre-defined observation period over time and are usually included in classical survival models, such as Cox proportional hazard [46] or parametric survival [106] regressions, as a time-fixed baseline covariate considering as new origin event the end of the observation period (see *classical framework* in Figure 2.1). In this way, the dynamics of drug consumption over therapy are completely discarded. Moreover, patients need to survive for a period at least equal to the observation period, which leads to a possible bias due to exclusion of early dying patients. Both issues can be overcome modelling adherence as a time-varying covariate that jointly evolves with patient’s outcome, i.e., both starting from the origin event  $T_0$  as shown in the *time-varying framework* in Figure 2.1.

Bijlsma *et al.* (2016) [29] performed a first attempt to measure time-varying adherence using electronic records, proving that their time-varying method better distinguished an irregularly dosing patient from a stably dosing patient and better accounted for changes over time in drug utilization behaviour. However, through their method, time-varying adherence to medication has been computed over a time-period defined by three successive fills, time-lapse different from the global time-scale of the most studied clinical outcomes (e.g., time-to-event in survival analysis). Therefore, as Steiner (2016) [193] highlighted, to establish a relationship between time-varying adherence and clinical outcomes, it is fundamental that these two components are measured on the same time scale. In this way, it could be possible to investigate the effect of the longitudinal adherence on the clinical outcome.

In this chapter we propose an innovative method to represent and measure adherence as time-varying covariate exploiting administrative databases of *Regione Lombardia* [164].

## 2. Joint modelling of time-varying adherence to medication and survival



**Figure 2.1.** *Classical framework*: binary adherence is computed on a pre-defined observation period and time-to-event outcome refers to the follow-up period. *Time-varying framework*: time-varying adherence is computed jointly with the time-to-event outcome.  $T_0$  is the origin event,  $T_0^*$  is the end of the observation period.

Our method could be seen as an extension of the time-fixed Proportion of Days Covered [102, 14] and it is computed on the same time scale of our event of interest, i.e., the death of HF patients (see *time-varying framework* in Figure 2.1). In particular, we observed that the dynamics of consumption and adherence to medication can be reconstructed using secondary databases related to (i) patient admission to hospital (Hospital Discharge Charts - HSC), which contain data related to hospital admissions and time to death (or administrative censoring), and (ii) pharmaceutical purchases, which provide information on the number and times of drug purchases. Since data on drugs prescriptions are not publicly available neither accessible, the approximation of drug consumption with drug purchase is the only viable option. Examples and limitations of using this approach into a pharmacoepidemiological setting are discussed in [14, 80, 102, 118, 187].

Motivated by the clinical question regarding the association between adherence to medication and patients' survival, we compared two different time-varying covariates: the continuous *time-dependent cumulative months covered by drug consumption* and the dichotomous *time-dependent adherence to medication*. The first one represents the dynamic behaviour and shape of drug intake, whereas the second one reflects the patient's purpose of taking the medication during time. Once these dynamic indicators are computed, we plug them into joint models [167]. These models are used in follow-up studies where interest is in associating an endogenous time-dependent response [97] with an event time outcome. Since the data we came up with in our procedure were jointly determined with the responses of interest and may be intended as endogenous covariates, this framework enables their proper treatment. The flexibility and wide range applicability of joint models to clinical setting [83] allow for subject-specific predictions and construction of personalized medicine tools. In fact, the added value of our approach consists in performing an ongoing analysis and a quantification of adherence effect on patient's outcome that allow to carry out a real-time monitoring and profiling of patients as well as a personalised prediction about long-term prognosis.

The remaining part of this chapter is organized as follows. In Section 2.1 we describe the statistical methodologies. First, we introduce two novel time-varying representation methods for drug consumption and adherence to medication; then we model them into a joint modelling framework. Data extraction, inclusion criteria and representation of pharmacological time-varying covariates are described in Section 2.2. Key results from applying these methods to administrative data provided by *Regione Lombardia - Healthcare Division* within the project *HFData* [164] are presented in Section 2.3. In Section 2.4, we end with a discussion of the strengths and limitations of the current approach. All the analyses were carried out using the free software R [161], in particular *JMbayes* package [168]. Codes are available as Supplementary Material of [188].

## 2.1. Statistical Methodologies

### 2.1.1. Pharmacological time-varying covariates

In classical survival models, such as Cox's proportional hazard model [46] or parametric survival models [106], pharmacological consumption and adherence are usually considered as binary (or categorical) baseline covariates. One of the most used adherence measure is the Proportion of Days Covered (PDC) [102, 14], defined as in Equation (1.2). PDC measure is computed on a pre-defined observation period (see *classical framework* in Figure 2.1) and returns a number between 0 and 1. PDC is usually dichotomized to identify as adherent those patients that reach an established threshold. However, since the dynamics of drug intake changes during therapy depending on patient's health status, a time-varying representation could be more appropriate and informative. Therefore, starting from PDC definition in Equation (1.2), we define time-varying adherence to medication in two alternative ways:

- (i)  $y_i^{(C)}(t)$ : a continuous ( $C$ ) time-dependent variable which indicates the cumulative time covered by therapy consumption up to time  $t$  by the  $i$ -th subject,  $\forall i = 1, \dots, n$ ,
- (ii)  $y_i^{(D)}(t)$ : a dichotomous ( $D$ ) time-dependent variable which indicates if the  $i$ -th patient is adherent to therapy at time  $t$ ,  $\forall i = 1, \dots, n$

$$y_i^{(D)}(t) := \begin{cases} 1 & \text{if } PDC_i(t) = \frac{y_i^{(C)}(t)}{t} \geq \tau \\ 0 & \text{otherwise,} \end{cases}$$

where  $\tau$  is a pre-defined threshold and time  $t$  can be expressed in days, weeks, months or years, depending on the type of data and on the focus of the research.

Variable  $\mathbf{y}^{(C)}(t) = \{y_i^{(C)}(t), i = 1, \dots, n\}$  could be seen as an extension of PDC numerator in (1.2) in which we considered time-varying observation periods, i.e., periods that begins from our survival origin event (the index date, i.e., time  $T_0 = 0$ ) and ends up at different times  $t$ . Variable  $\mathbf{y}^{(D)}(t) = \{y_i^{(D)}(t), i = 1, \dots, n\}$  was a dichotomization of  $\mathbf{y}^{(C)}(t)$  to

identify as adherent those patients with a proportion of therapy consumption up to time  $t$  greater or equal to a pre-defined threshold  $\tau$ , i.e., the  $\tau \times 100\%$  of the observation period up to  $t$ . Using this approach, our covariates were measured on the same time-scale of the survival framework, since they both started at the origin event of the survival analysis, i.e., time  $T_0 = 0$ .

Variables  $\mathbf{y}^{(C)}(t)$  and  $\mathbf{y}^{(D)}(t)$  represent two different ways to include into a longitudinal framework the information related to patients' adherence to continuity in routine therapy assumption, leading to two different approaches that we want to compare. The dichotomous covariate  $\mathbf{y}^{(D)}(t)$  provided an "easy" time-dependent representation of adherence, since it represented the therapy assumption rate during time, i.e., the proportion of days covered at time  $t$   $PDC(t)$  dichotomized according to a certain threshold  $\tau$ , as usually done in the literature. This allowed us to distinguish patients with *good* ( $\mathbf{y}^{(D)}(t) = 1$ ) and *poor* ( $\mathbf{y}^{(D)}(t) = 0$ ) adherence continuity rates during time. On the contrary, the continuous covariates  $\mathbf{y}^{(C)}(t)$  was able to reflect how "compliant" they were in assuming therapy with continuity and in which periods they actually assumed the drug. In fact, while the value of  $\mathbf{y}^{(C)}(t)$  was a measure of the cumulative time on which the patient took the drug (i.e., how "compliant" it was), the slope of the longitudinal trajectory was able to provide information on modifications in patient's behaviours during different periods. Indeed, considering the longitudinal trajectory for a given patient between two consecutive times  $t$  and  $t + 1$ , we had that:

- (i) a slope equal to 0 indicated that the patient never took the drug during interval  $[t; t + 1]$ ;
- (ii) a slope equal to 1 indicated that the patient took the drug every day of interval  $[t; t + 1]$ ;
- (iii) a slope in in  $(0,1)$  indicated that the patient took the drug sometimes, but not every day.

Moreover, changes in the value of the trajectory slope over time reflected changes in patient's adherence continuity: an increasing slope indicated that patient's behaviour became more appropriate in terms of continuity in adherence to medication, a decreasing slope indicated that patient's behaviour became more inappropriate and a constant slope indicated that patient's behaviour remained unchanged (proper or improper according to the value of the slope). Therefore, since the continuous variable  $\mathbf{y}^{(C)}(t)$  was more descriptive and informative than the dichotomous  $\mathbf{y}^{(D)}(t)$ , we expected that the approach with  $\mathbf{y}^{(C)}(t)$  resulted more powerful in predicting and real-time evaluating the effect of the covariate on patient's survival status.

We finally underline that in clinical practice therapies are usually modified according to the disease progression. This aspect allowed us to consider covariates (2.8) and (2.9) as *endogenous* (or *internal*) time-dependent covariates, since their time paths were jointly determined with the responses of interest. Indeed, as Kalbfleisch and Prentice (2011) [97] stated, the key point of endogenous covariates is that their existence and future path are

directly related to the event status. Hence, since they were related to the behaviour of the individual over time, both the time-varying variables could be seen as endogenous.

### 2.1.2. Joint model specification

We now introduce the joint model approach proposed by Rizopoulos (2012) [167] for dealing with time-to-event and endogenous longitudinal covariates. The choice of a joint model is driven by the fact that, when the outcome processes are correlated, joint modelling has empirically demonstrated to reduce biases, improve efficiency and prediction and can be applicable to outcome surrogacy [83].

Let  $T_i^*$  denotes the true event time for the  $i$ -th subject,  $C_i$  the censoring time,  $T_i = \min(T_i^*, C_i)$  the corresponding observed event time and  $D_i = \mathbb{1}(T_i^* \leq C_i)$  the event outcome indicator, with  $I(\cdot)$  being the indicator function that takes the value 1 when  $T_i^* \leq C_i$ , and 0 otherwise. Let  $\mathcal{D}_n = \{T_i, D_i, \mathbf{y}_i; i = 1, \dots, n\}$  denote a sample from the target population, where  $\mathbf{y}_i$  denote the  $n_i \times 1$  longitudinal response vector for the  $i$ -th subject ( $\mathbf{y}_i^{(C)}$  or  $\mathbf{y}_i^{(D)}$  in our analyses), with element  $y_{il}$  denoting the value of the longitudinal process taken at time point  $t_{il}$ ,  $l = 1, \dots, n_i$ . The general form of joint models we used for our analysis is the following:

$$g[\mathbb{E}\{y_i(t)|\mathbf{b}_i\}] = \eta_i(t) = \mathbf{x}_i^T(t)\boldsymbol{\beta} + \mathbf{z}_i^T(t)\mathbf{b}_i \quad (2.1)$$

$$h_i(t|\mathcal{H}_i(t), \boldsymbol{\omega}_i) = h_0(t) \exp \{ \boldsymbol{\theta}^T \boldsymbol{\omega}_i(t) + f(\mathcal{H}_i(t), \mathbf{b}_i, \boldsymbol{\alpha}) \}, \quad t > 0. \quad (2.2)$$

The longitudinal process, given by equation (2.1), is a generalized linear mixed effects model in which  $g(\cdot)$  denotes a known one-to-one monotonic link function,  $y_i(t)$  denotes the value of the longitudinal process for the  $i$ -th subject at time point  $t$ ,  $\boldsymbol{\beta}$  is the vector of the unknown fixed effects parameters,  $\mathbf{b}_i$  is the vector of subject-specific random effects,  $\mathbf{x}_i(t)$  and  $\mathbf{z}_i(t)$  denote the time-dependent vectors for the fixed and random effect, respectively. The event process, given by equation (2.2), assumes that the risk  $h_i(\cdot)$  for an event depends on a function  $f(\cdot)$  of the subject-specific linear predictor  $\eta_i(t)$ . In particular,  $\mathcal{H}_i(t) = \{\eta_i(s), 0 \leq s < t\}$  denotes the history of the underlying longitudinal process up to time point  $t$ ,  $h_0(\cdot)$  denotes the baseline hazard function,  $\boldsymbol{\omega}_i(t)$  is a vector of exogenous, baseline or possibly time-varying, covariates with corresponding regression coefficients  $\boldsymbol{\theta}$ . The parameter  $\boldsymbol{\alpha}$  is the vector that quantifies the association between features of the marker process up to time  $t$  and the hazard of an event at the same time point. Moreover, the baseline hazard function  $h_0(\cdot)$  is modelled using a B-splines approach.

In `JMbayes` package, the estimation of the models parameters proceeds under a Bayesian approach, using MCMC algorithms. Details regarding Bayesian estimation of joint models can be found in [168, 88, 33].

### Longitudinal and event processes for pharmacological time-varying covariates

In this section we specify longitudinal and event processes that we used to properly model the pharmacological time-varying covariates introduced in Section 2.1.1. The choice of the longitudinal submodels was driven by the nature of the time-varying processes themselves: on one hand we needed a model able to capture the dynamic shape of drug consumption, on the other hand we used a model for binary values.

For the continuous time-varying variable  $\mathbf{y}^{(C)}(t) = \{y_i^{(C)}(t), i = 1, \dots, n\}$ , which indicates the cumulative time covered by therapy consumption up to time  $t$ , we postulated a linear mixed effect model as longitudinal submodel (2.1). Since the longitudinal trajectories were nonlinear for many patients, we included natural cubic splines in both the fixed and random effects parts in order to properly accounting for non-linearity, adjusting each trajectory for baseline covariates with fixed effects. The resulting longitudinal process was then of the following form:

$$y_i^{(C)}(t) = \eta_i(t) + \varepsilon_i(t) = (\beta_0 + b_{i0}) + \sum_{k=1}^4 (\beta_k + b_{ik})B_n(t, \lambda_k) + \tilde{\mathbf{x}}_i^T \tilde{\boldsymbol{\beta}} + \varepsilon_i(t) \quad (2.3)$$

where  $i$  is the patient's index,  $\{B_n(t, \lambda_k) : k = 1, 2, 3, 4\}$  denotes the B-spline basis matrix for a natural cubic spline of time  $t$  with three internal knots placed at 25th, 50th and 75th percentiles of the follow-up times,  $\tilde{\mathbf{x}}_i$  is the vector of baseline covariates with fixed effects with regression parameters  $\tilde{\boldsymbol{\beta}}$ ,  $\varepsilon_i(t) \sim \mathcal{N}(0, \sigma_\varepsilon^2 \mathbf{I}_{n_i})$  is the unknown vector of random errors and  $\mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{D})$  is the vector of the patient-specific random effects, with  $\mathbf{D}$  unstructured variance-covariance matrix. Therefore the time-dependent vectors for the fixed and random effects for the  $i$ -th patient in (2.1) were  $\mathbf{x}_i(t) = [1, B_n(t, \lambda_1), \dots, B_n(t, \lambda_4), \tilde{\mathbf{x}}_i^T]^T$  and  $\mathbf{z}_i(t) = [1, B_n(t, \lambda_1), \dots, B_n(t, \lambda_4)]^T$ , with relative vectors of regression coefficients  $\boldsymbol{\beta} = [\beta_0, \dots, \beta_4, \tilde{\boldsymbol{\beta}}^T]^T$  and  $\mathbf{b}_i = [b_{i0}, \dots, b_{i4}]^T$ .

For the dichotomous time-varying variable  $\mathbf{y}^{(D)}(t) = \{y_i^{(D)}(t), i = 1, \dots, n\}$ , which indicates adherence to therapy at time  $t$ , we postulated a logistic mixed effect model as longitudinal submodel (2.1), as follows:

$$\log \frac{\Pr [y_i^{(D)}(t) = 1]}{1 - \Pr [y_i^{(D)}(t) = 1]} = \eta_i(t) = \beta_0 + b_{i0} + (\beta_1 + b_{i1}) t + \tilde{\mathbf{x}}_i^T \tilde{\boldsymbol{\beta}} \quad (2.4)$$

where  $i$  is the patient's index,  $\tilde{\mathbf{x}}_i$  is the vector of baseline covariates with fixed effects with regression parameters  $\tilde{\boldsymbol{\beta}}$  and  $\mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{D})$  is the vector of the patient-specific random effects, with  $\mathbf{D}$  an unstructured variance-covariance matrix. Therefore the time-dependent vectors for the fixed and random effects for the  $i$ -th patient in (2.1) were  $\mathbf{x}_i(t) = [1, t, \tilde{\mathbf{x}}_i^T]^T$  and  $\mathbf{z}_i(t) = [1, t]^T$ , with relative vectors of regression coefficients  $\boldsymbol{\beta} = [\beta_0, \beta_1, \tilde{\boldsymbol{\beta}}^T]^T$  and  $\mathbf{b}_i = [b_{i0}, b_{i1}]^T$ .

For the event submodels, we wanted to focus on patient's cumulative and current adherence paths. On one hand, the current value of the subject-specific linear predictor gives

information about the pharmacological history of drug assumption of each patient, i.e., its cumulative path. On the other hand, the first derivative, i.e., the slope, is able to reflect changes in drug intake between different time periods, especially through the continuous covariate  $\mathbf{y}^{(C)}(t)$ , giving us information about the patient's current path. Therefore, we focused on three different forms of  $f(\mathcal{H}_i(t), \mathbf{b}_i, \boldsymbol{\alpha})$  in (2.2), corresponding to different meaning of the linear predictor:

- (i) the risk of death for the  $i$ -th patient depends on the current true value of the subject-specific linear predictor at time  $t$ :

$$h_i(t) = h_0(t) \exp \{ \boldsymbol{\theta}^T \boldsymbol{\omega}_i(t) + \alpha_1 \eta_i(t) \}; \quad (2.5)$$

- (ii) the risk of death for the  $i$ -th patient depends on both the current true value of the subject-specific linear predictor and its slope at time  $t$ :

$$h_i(t) = h_0(t) \exp \{ \boldsymbol{\theta}^T \boldsymbol{\omega}_i(t) + \alpha_1 \eta_i(t) + \alpha_2 \eta_i'(t) \}; \quad (2.6)$$

- (iii) the risk of death for the  $i$ -th patient depends on the slope of the subject-specific linear predictor at time  $t$ :

$$h_i(t) = h_0(t) \exp \{ \boldsymbol{\theta}^T \boldsymbol{\omega}_i(t) + \alpha_2 \eta_i'(t) \}. \quad (2.7)$$

In this way we were able to investigate the effects of (i) cumulative adherence, (ii) both cumulative and current adherence and (iii) current adherence on patients' long-term survival, adjusting for other baseline or time-varying exogenous characteristics in  $\boldsymbol{\omega}_i(t)$ .

## 2.2. Materials and Administrative data

### 2.2.1. Study setting

Between January 2000 and December 2012, non-paediatric (age  $\geq 18$  years) patients living in Lombardy (one of the biggest and most populated Italian region accounting for 10 million residents) hospitalized with a principal diagnostic code of HF were recruited (see Mazzali *et al.*, 2016 [136]). Enrolment occurred from the data of discharge of the first HF hospitalization (i.e., the index date). Among the disease-modifying drugs for HF patients mentioned in [139] and [154], we focused on Angiotensin-Converting Enzyme inhibitors (ACE) and Angiotensin II Receptor Blockers (ARB), which are drugs of routine use for HF [222] therefore they should be taken regularly by HF patients, regardless of the level of severity of their health status. In particular, patients who bought at least one medication of ACE or ARB during the first year of follow-up were selected.

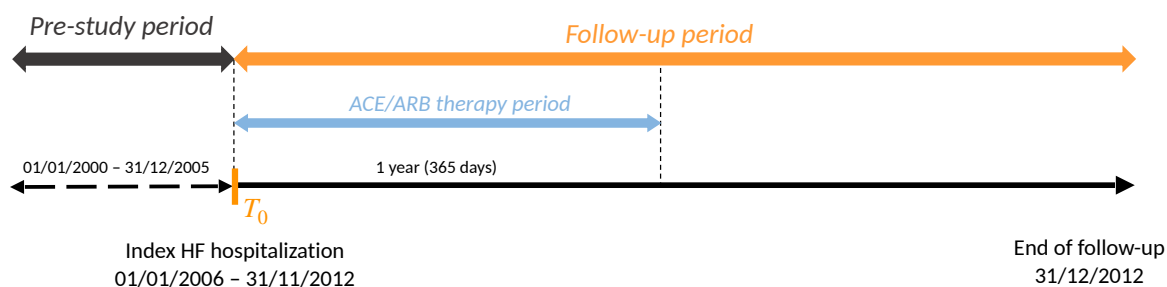


### 2.2.2. Administrative data sources

The project database was built for residents in Lombardy which were hospitalized for HF from 2000 to 2012. Data were provided by *Regione Lombardia - Healthcare Division*, within the research project *HFData* [HFData-RF-2009-1483329] [164]. In order to protect privacy, information retrieved from the different databases were linked via a single anonymous ID (identification) code. For further details regarding data extraction and selection see Mazzali *et al.* (2016) [136].

Each record in the dataset was related to an event, which could be an hospitalization or a drug purchase of a given patient. With regard to ordinary hospital admission, the date of discharge from hospital and the length of stay in hospital were retrieved. For drug purchases, identified by their Anatomical Therapeutic Chemical (ATC) codes [214], the date of purchase and the number of days of treatment covered by the prescription, based on the number of boxes and the Defined Daily Dose (DDD) [220] for that specific medicinal product, were retrieved.

In this work we focused on a representative sample of *HFData* related to patients with their first HF discharge between January 2006 to December 2012, excluding patients who died during the index hospitalization. A 5-years *pre-study period* from 2000 to 2005 (Figure 2.2) was used in order to consider only "incident" HF patients, i.e., patients with no contacts with healthcare system in the previous five years due to HF. This choice allowed us to reduce potential time-lag biases [155] due to different severity of the disease. To avoid a possible survival bias due to patient's critical conditions, we excluded those patients who died within 30 days from the index date. Moreover, we defined the *ACE/ARB therapy period* (Figure 2.2) to select purchases within 1 year of follow-up, since we were interested in the effect the time-varying adherence to the first year of ACE/ARB therapy. Therefore, only patients with at least one ACE/ARB purchase were included in the final study cohort. Demographics and comorbidities were considered to adjust models.



**Figure 2.2.** Study design. HF = Heart Failure, ACE = Angiotensin-Converting Enzyme inhibitors, ARB = Angiotensin II Receptor Blockers.

### 2.2.3. Pharmacological time-varying covariates for ACE/ARB therapy

As explained in Section 2.1.1, starting from PDC definition (1.2), we represented time-varying adherence to ACE/ARB therapy in two alternative ways:

- (i) a continuous time-dependent variable which indicates the cumulative months covered by ACE/ARB consumption up to time  $t$  for the  $i$ -th patient

$$\text{cum\_months}_i(t) = y_i^{(C)}(t) := \text{distinct coverage months up to time } t; \quad (2.8)$$

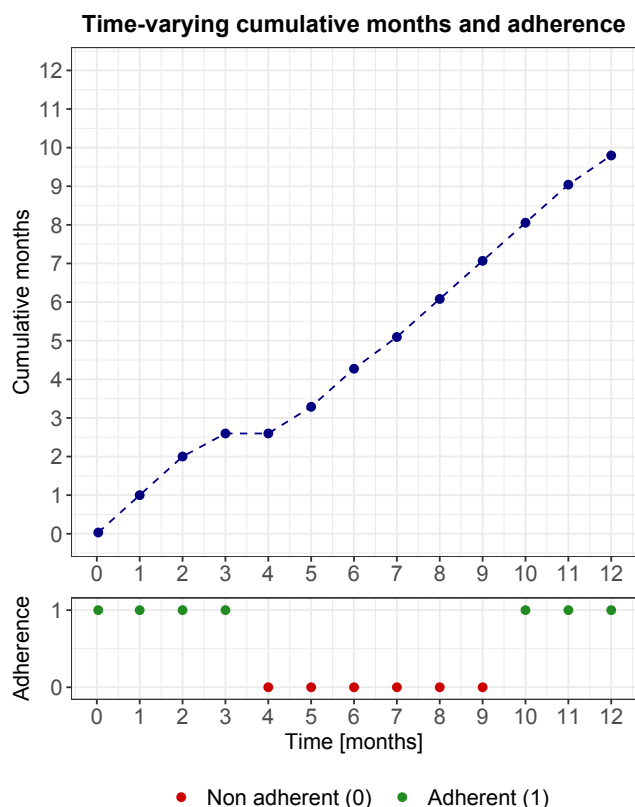
- (ii) a dichotomous time-dependent variable which indicates if the  $i$ -th patient is adherent to the ACE/ARB therapy at time  $t$

$$\text{adherence}_i(t) = y_i^{(D)}(t) := \begin{cases} 1 & \text{if } \frac{\text{cum\_months}_i(t)}{t} \geq 0.80 \\ 0 & \text{otherwise.} \end{cases} \quad (2.9)$$

In both cases, times  $t$  were expressed in months. In particular, we considered the first day of follow-up ( $t = 0.033$  months) and each months ( $t = 1, \dots, 12$  months) up to the end of the first year or up to the patient's death, if he/she died during the first year of follow-up. Using this approach, our covariate was measured on the same time-scale of our survival framework. For  $\text{cum\_months}_i(t)$  (2.8) computation we considered only the coverage of distinct periods, which means that, in case of overlapping of two subsequent purchases, we considered the period covered by the first purchase entirely and the second purchase only in those days that were not covered by the first one. Moreover, we assumed full adherence during re-hospitalization period [14], and we based our analysis on purchased drugs instead of prescribed drugs, as done in Spreafico *et al.* (2020) [187]. In particular, for each months we firstly computed the *cumulative coverage days* up to the current month  $t$ . Then, converting days into months, we obtained the continuous time-varying covariate  $\text{cum\_months}_i(t)$ , which indicates the cumulative months covered by ACE/ARB assumption up to time  $t$ .

Finally, variable  $\text{adherence}_i(t)$  (2.9) was a dichotomization of variable  $\text{cum\_months}_i(t)$  to identify as adherent those patients with a proportion of months covered by ACE/ARB consumption up to time  $t$  greater or equal to  $\tau = 0.8$ , i.e., the 80% of the observation period up to  $t$ . Therefore,  $\text{adherence}_i(t)$  was equal to 1 if  $\text{cum\_months}_i(t)/t \geq 0.8$  at time  $t$ , 0 otherwise.

This reconstruction process ended up with a long-format database with multiple rows for each patient, one for each time point of his/her time-varying covariates. In Figure 2.3 we reported an example of the final reconstruction of the covariates  $\text{cum\_months}_i(t)$  (top panel) and  $\text{adherence}_i(t)$  (bottom panel) for a random patient.



**Figure 2.3.** Example of time-varying consumption and adherence to ACE/ARB therapy,  $\text{cum\_months}_i(t)$  and  $\text{adherence}_i(t)$  respectively.

## 2.2.4. Outcome measure

Study outcome of interest was patient's death for any cause. Deaths were collected from the Hospital Discharge Forms Database (for in-hospital deaths) or Vital Statistics Regional Dataset (for out-hospital deaths). For the survival analysis, each patient was followed from the index date (i.e., the discharge from the index HF hospitalization,  $T_0 = 0$ ) until the end of the study or the date of death (see *follow-up period* in Figure 2.2). The administrative censoring date was December 31<sup>st</sup>, 2012.

## 2.3. Results

### 2.3.1. Study cohort

A representative sample cohort of 4,870 patients were identified with principal diagnostic code of HF during the period 2006-2012. Of these, we excluded 13 (0.3%) patients who died during the 30 days after the index hospitalization. Moreover, 883 patients (18.1%) were removed since they did not present any purchase of ACE or ARB in the first year after the index hospitalization. Thus, a total of 3,974 (81.6%) patients met study selection criteria (see Figure 2.4).

Overall, at index hospitalization mean age of the study cohort was 72.82 years ( $s.d. = 11.16$ ) with a percentage of male patients equal to 55.8% (2,219 patients). The mean number of comorbidities was 2.09 ( $s.d. = 1.08$ ) with 30.5% of patients presenting  $\geq 3$  comorbidities). The median time of *follow-up period* was 48.85 (IQR = [30.72; 66.94]) months. At administrative censoring date 1,012 patients (25.5%) were dead and 2,962 (74.5%) were censored. Moreover, at the end of *ACE/ARB therapy period* (i.e.,  $t = 12$  months), the percentage of living patients was 94.7% (3,764 patients), with a mean value of distinct coverage months, i.e., mean value of  $\text{cum\_months}_i(12)$ , equal to 8.77 months ( $s.d. = 3.04$ ) and only 2,039 patients (54.2%) with  $\text{adherence}_i(12) = 1$ .

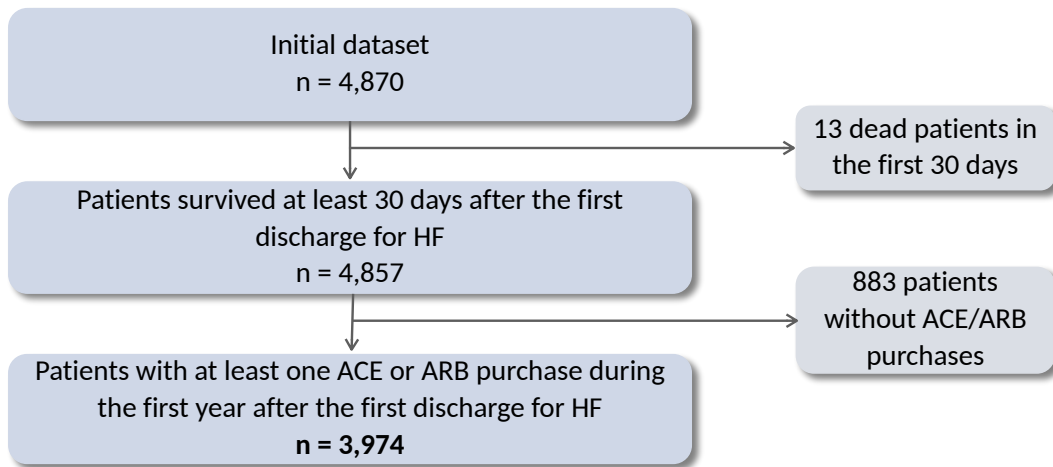


Figure 2.4. Flowchart of cohort selection.

### 2.3.2. Joint models for time-varying consumption and adherence to ACE/ARB therapy

In order to assess the role of time-varying consumption and adherence to ACE/ARB therapy with respect to the overall survival time of a patient, we estimated six different joint models.

Referring to the continuous time-varying consumption of ACE/ARB therapy  $\text{cum\_months}_i(t)$ , we considered three joint models, namely  $M_1$ ,  $M_2$  and  $M_3$ , with the same longitudinal subprocess (2.3) given by

$$\begin{aligned} \text{cum\_months}_i(t) &= \eta_i(t) + \varepsilon_i(t) \\ &= \beta_0 + b_{i0} + \sum_{k=1}^4 (\beta_k + b_{ik}) B_n(t, \lambda_k) + \beta_5 \text{n\_com}_i + \varepsilon_i(t) \end{aligned}$$

## 2. Joint modelling of time-varying adherence to medication and survival

and the event submodels (2.5)-(2.6)-(2.7) defined as follows

$$\begin{aligned} M_1 : h_i(t) &= h_0(t) \exp\{\theta_1 \text{age}_i + \theta_2 \text{gender}_i + \theta_3 \text{n\_com}_i + \alpha_1 \eta_i(t)\}; \\ M_2 : h_i(t) &= h_0(t) \exp\{\theta_1 \text{age}_i + \theta_2 \text{gender}_i + \theta_3 \text{n\_com}_i + \alpha_1 \eta_i(t) + \alpha_2 \eta'_i(t)\}; \\ M_3 : h_i(t) &= h_0(t) \exp\{\theta_1 \text{age}_i + \theta_2 \text{gender}_i + \theta_3 \text{n\_com}_i + \alpha_2 \eta'_i(t)\}. \end{aligned}$$

Referring to the dichotomous time-varying adherence to ACE/ARB therapy  $\text{adherence}_i(t)$ , we considered three joint models, namely  $M_4$ ,  $M_5$  and  $M_6$ , with the same longitudinal subprocess (2.4) given by

$$\log \frac{\text{Pr}[\text{adherence}_i(t) = 1]}{1 - \text{Pr}[\text{adherence}_i(t) = 1]} = \eta_i(t) = \beta_0 + b_{i0} + (\beta_1 + b_{i1}) t + \beta_2 \text{n\_com}_i$$

and the event submodels (2.5)-(2.6)-(2.7) defined as follows

$$\begin{aligned} M_4 : h_i(t) &= h_0(t) \exp\{\theta_1 \text{age}_i + \theta_2 \text{gender}_i + \theta_3 \text{n\_com}_i + \alpha_1 \eta_i(t)\}; \\ M_5 : h_i(t) &= h_0(t) \exp\{\theta_1 \text{age}_i + \theta_2 \text{gender}_i + \theta_3 \text{n\_com}_i + \alpha_1 \eta_i(t) + \alpha_2 \eta'_i(t)\}; \\ M_6 : h_i(t) &= h_0(t) \exp\{\theta_1 \text{age}_i + \theta_2 \text{gender}_i + \theta_3 \text{n\_com}_i + \alpha_2 \eta'_i(t)\}. \end{aligned}$$

In both cases, the longitudinal submodel was adjusted for the the total number of comorbidities at the index hospitalization. Therefore, in submodels (2.3) and (2.4) the vector of baseline covariates with fixed effects was given by  $\tilde{\boldsymbol{x}}_i = \text{n\_com}_i$ .

Moreover, each event submodel was adjusted for three baseline covariates: age, gender and total number of comorbidities at the index hospitalization. The choice of these covariates was driven by clinical relevance and availability from administrative data and were used to prevent as much as possible biases induced by the use of secondary database. Hence, in the event submodels (2.5)–(2.6)–(2.7) the vector of the exogenous baseline covariates was given by  $\boldsymbol{\omega}_i = (\text{age}_i, \text{gender}_i, \text{n\_com}_i)$ .

For the model fitting, we used version 3.6.2 of R software and version 0.8-85 of **JMbayes** package. We ran the MCMC sampler implemented in `jointModelBayes()` function for a total number of 36,000 iterations, discarding the first 3,000 as burn-in and other 3,000 as adaptation and thinning every 15 iterations; the final sample size was 2,000. Covariates age (`age`) and number of comobidities (`n_com`) have been standardized for easing convergence during parameters estimation.

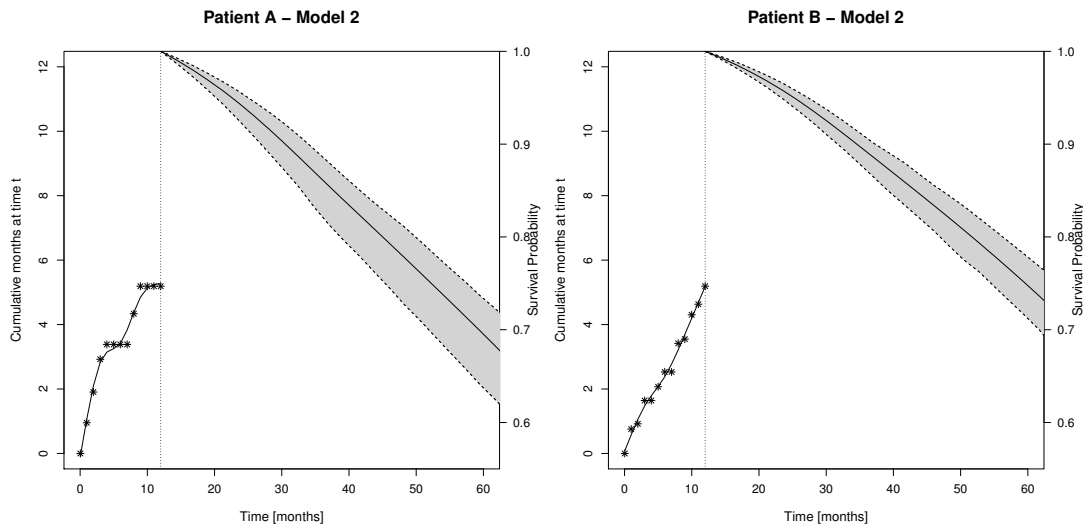
## Results

Considering the continuous time-varying variable  $\boldsymbol{y}^{(C)}(t) = \text{cum\_months}(t)$  as longitudinal process like in (2.3), we fitted the three different joint models  $M_1$ ,  $M_2$  and  $M_3$  introduced in Section 2.3.2. The results of the model parameter estimations are shown in Table 2.1, together with their deviance information criterion (DIC) values. Results from the different

models were similar, but model  $M_2$  presented a lower DIC value with respect to  $M_1$  and  $M_3$ . Therefore, the selected model for time-varying  $\text{cum\_months}(t)$  was the following:

$$M_2 : \begin{cases} \text{cum\_months}_i(t) = \eta_i(t) + \varepsilon_i(t) \\ \quad = \beta_0 + b_{i0} + \sum_{k=1}^4 (\beta_k + b_{ik}) B_n(t, \lambda_k) + \beta_5 \text{n\_com}_i + \varepsilon_i(t) \\ h_i(t) = h_0(t) \exp\{\theta_1 \text{age}_i + \theta_2 \text{gender}_i + \theta_3 \text{n\_com}_i + \alpha_1 \eta_i(t) + \alpha_2 \eta'_i(t)\}. \end{cases}$$

From parameter estimation of model  $M_2$  we observed that all the 95% credibility intervals did not contain 0, except for the parameters  $\beta_0$  and  $\alpha_1$ . From the longitudinal process, we observed that the number of baseline comorbidities negatively influenced the cumulative months covered by ACE/ARB assumption ( $\hat{\beta}_5 = -0.3414 < 0$  with 2.5%–97.5% CI =  $[-0.6453; -0.0495]$ ), probably reflecting that as comorbidities increased as the mix of drugs changed accordingly. In the event process, all the covariates were associated with the risk of death, except for the current level of the linear predictor. In particular, being younger or a female corresponded to a higher survival probability, whereas having a higher number of initial comorbidities corresponded to a lower survival probability, as it might be expected. Moreover, the slope of the linear predictor had a protective role: the HR related to the slope value is  $\exp(\hat{\alpha}_2) = \exp(-0.6301) = 0.533$ . Hence, a 1-unit increase in the value of the slope corresponded to a 0.533-fold decrease in the risk of death (2.5–97.5% CI =  $[0.380; 0.743]$ ). Figure 2.5 shows the survival probability plot for two male patients,  $A$  and  $B$ , aged 72 with two comorbidities and  $y_A(12) = y_B(12) = 5.191$ . From the figure, we observed that the patient with the higher slope of the linear predictor at time  $t = 12$  (patient  $A$  in right panel) had a higher survival probability during time. Hence, having a good adherence trend during time reflected a protective role on patients' survival.



**Figure 2.5.** Survival probability plots for two male patients aged 72 years and with two comorbidities at the index date using joint model  $M_2$ . The values of continuous time-varying covariate  $\text{cum\_months}(t)$  are reported in the left part of both panels (stars) with their linear predictors  $\eta_i(t)$  (line). In particular, at  $t = 12$  months  $\text{cum\_months}_A(12) = \text{cum\_months}_B(12) = 5.191$ , with relative linear predictors  $\eta_A(12) = 5.183$  and  $\eta_B(12) = 5.271$  and slopes  $\eta'_A(12) = 0.029$  and  $\eta'_B(12) = 0.503$ .

2. Joint modelling of time-varying adherence to medication and survival

**Table 2.1.** Means of parameter estimates with 95% credibility intervals (CI) and Deviance Information Criterion (DIC) under the joint modelling analyses  $M_1$ ,  $M_2$  and  $M_3$  for continuous time-varying covariate  $\mathbf{y}^{(C)}(t) = \text{cum\_months}(t)$ .

	$M_1$		$M_2$		$M_3$		
<b>Longitudinal Process</b>	<b>Mean</b>	<b>[2.5; 97.5]% CI</b>	<b>Mean</b>	<b>[2.5; 97.5]% CI</b>	<b>Mean</b>	<b>[2.5; 97.5]% CI</b>	
Intercept	$\beta_0$	-0.0254	[-0.4406; 0.3992]	-0.0213	[-0.4445; 0.3699]	-0.0240	[-0.4253; 0.3885]
$B_n(t, \lambda_1)$	$\beta_1$	4.4124	[4.3574; 4.4692]	4.4120	[4.3574; 4.4673]	4.4125	[4.3566; 4.4647]
$B_n(t, \lambda_2)$	$\beta_2$	6.0006	[5.9299; 6.0692]	5.9997	[5.9302; 6.0678]	6.0009	[5.9325; 6.0668]
$B_n(t, \lambda_3)$	$\beta_3$	9.8316	[9.7227; 9.9390]	9.8296	[9.7246; 9.9359]	9.8324	[9.7302; 9.9328]
$B_n(t, \lambda_4)$	$\beta_4$	7.6003	[7.5103; 7.6901]	7.5998	[7.5110; 7.6865]	7.6003	[7.5092; 7.6892]
Comorbidity	$\beta_5$	-0.3408	[-0.6370; -0.0448]	-0.3414	[-0.6453; -0.0495]	-0.3417	[-0.6446; -0.0523]
	$\sigma_\varepsilon$	0.1529	[0.1517; 0.1541]	0.1529	[0.1517; 0.1541]	0.1529	[0.1517; 0.1542]
<b>Event Process</b>	<b>Mean</b>	<b>[2.5; 97.5]% CI</b>	<b>Mean</b>	<b>[2.5; 97.5]% CI</b>	<b>Mean</b>	<b>[2.5; 97.5]% CI</b>	
Age	$\theta_1$	0.7391	[0.6494; 0.8280]	0.7332	[0.6483; 0.8129]	0.7365	[0.6525; 0.8207]
Gender (Male)	$\theta_2$	0.1824	[0.0548; 0.3107]	0.1922	[0.0687; 0.3193]	0.1918	[0.0567; 0.3218]
Comorbidity	$\theta_3$	0.1968	[0.1380; 0.2558]	0.1978	[0.1377; 0.2533]	0.1970	[0.1397; 0.2549]
Current Value	$\alpha_1$	-0.0090	[-0.0133; -0.0046]	0.0056	[-0.0032; 0.0146]	-0.4537	[-0.6131; -0.2785]
Slope	$\alpha_2$			-0.6301	[-0.9670; -0.2964]		
<b>DIC</b>		61439.06		61372.08		61431.62	

On the other hand, considering the dichotomous time-varying variable as longitudinal process  $\mathbf{y}^{(D)}(t) = \text{adherence}(t)$  like in (2.4), we fitted the three different joint models  $M_4$ ,  $M_5$  and  $M_6$  introduced in Section 2.3.2. The results of the model parameter estimations are shown in Table 2.2, together with their deviance information criterion (DIC) values. Results from the different models were similar, but model  $M_4$  presented a lower DIC value with respect to  $M_5$  and  $M_6$ . Therefore, the selected model for time-varying  $\text{adherence}(t)$  was the following:

$$M_4 : \begin{cases} \log \frac{\text{Pr}[\text{adherence}_i(t)=1]}{1-\text{Pr}[\text{adherence}_i(t)=1]} = \eta_i(t) = \beta_0 + b_{i0} + (\beta_1 + b_{i1}) t + \beta_2 \text{n.com}_i \\ h_i(t) = h_0(t) \exp\{\theta_1 \text{age}_i + \theta_2 \text{gender}_i + \theta_3 \text{n.com}_i + \alpha_1 \eta_i(t)\} \end{cases}$$

From parameter estimation of model  $M_4$  we observe that all the 95% credibility intervals did not contain 0. From the longitudinal process, we observed that, also in this case, the number of baseline comorbidities negatively influenced the probability of adherence to ACE/ARB assumption ( $\hat{\beta}_2 = -0.2955 < 0$  with 2.5%–97.5% CI =  $[-0.4166; -0.1753]$ ), probably reflecting a change in drugs mix according to the increased number of morbidities. In the event process, all the covariates were associated with the risk of death. In particular, being younger or a female corresponded to a higher survival probability, whereas having a higher number of initial comorbidities corresponded to a lower survival probability, as it might be expected. Moreover, the current value of linear predictor had a protective role: the HR for a 10-units increase in the current value is  $\exp(\hat{\alpha}_1 \cdot 10) = \exp(-0.0011 \cdot 10) = 0.989$ . Hence, a 10-units increase in the value of the predictor corresponded to a 0.989-fold decrease (2.5–97.5% CI =  $[0.981; 0.997]$ ) in the risk of death. Figure 2.6 shows the survival probability plot for two male patients,  $C$  and  $D$ , aged 72 with two comorbidities,  $y_C(12) = 1$  and  $y_D(12) = 0$ . From the figure, we observed that the adherent patient at time  $t = 12$  (patient  $C$  in left panel) had a higher survival probability during time. Patient  $C$  was the one with the higher value of the linear predictor at time  $t = 12$ . Indeed, the current values of their linear predictors at time  $t = 12$  were  $\eta_C(12) = 9.64$  and  $\eta_D(12) = -7.64$ . Therefore, also in this case, having a good adherence trend during time reflected a protective role on patients' survival.

Note that in the selected models, the risk of death depended on the slope of  $\text{cum\_months}(t)$  in  $M_2$  but on the current value of  $\text{adherence}(t)$  in  $M_4$ . This difference is due to the different meaning of the two covariates, as explained in Section 2.1.1. On one hand,  $\text{cum\_months}(t)$  is the cumulative months covered by ACE/ARB consumption and its slope reflects current adherence. On the other hand, the current value of  $\text{adherence}(t)$  is directly related to the rate of cumulative months covered by therapy assumption, since it represents its dichotomization in *good* and *poor* continuity using an 80% threshold.

### Comparison of the two approaches

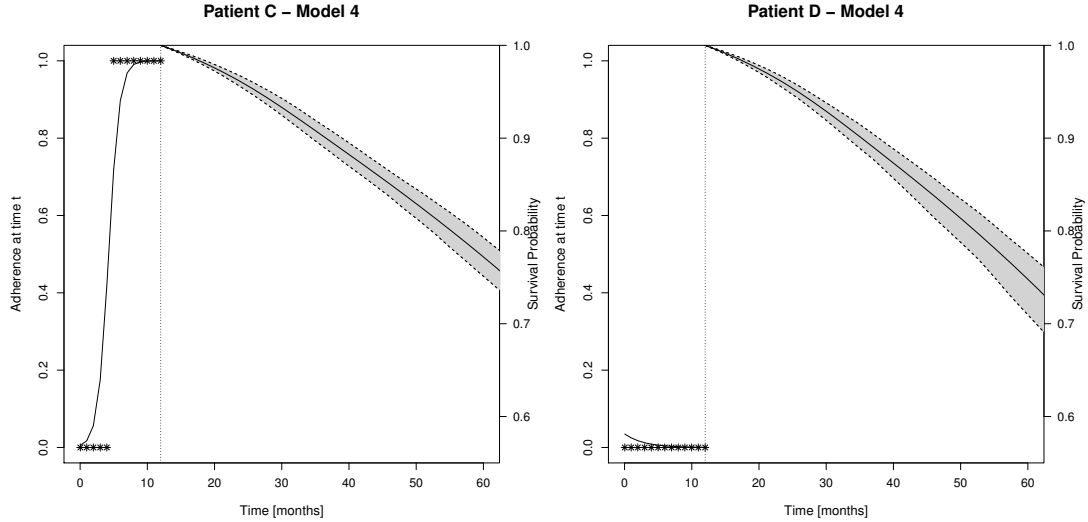
We finally compared models  $M_2$  and  $M_4$ . Both models led to similar considerations concerning adherence to medication. On one hand, they both indicated that patients with different number of comorbidities are characterized by different mix of drugs, suggesting that polytherapy, i.e., the use of multiple medications simultaneously, must be taken into



2. Joint modelling of time-varying adherence to medication and survival

**Table 2.2.** Means of parameter estimates with 95% credibility intervals (CI) and Deviance Information Criterion (DIC) under the joint modelling analyses  $M_4$ ,  $M_5$  and  $M_6$  for dichotomous time-varying covariate  $\mathbf{y}^{(D)}(t) = \text{adherence}(t)$ .

	$M_4$		$M_5$		$M_6$	
<b>Longitudinal Process</b>	<b>Mean</b>	<b>[2.5; 97.5]% CI</b>	<b>Mean</b>	<b>[2.5; 97.5]% CI</b>	<b>Mean</b>	<b>[2.5; 97.5]% CI</b>
Intercept	$\beta_0$	-2.0829 [-2.2211; -1.9405]	-2.0905 [-2.2243; -1.9604]	-2.0826 [-2.2120; -1.9528]		
Time $t$	$\beta_1$	0.8368 [0.7563; 0.9232]	0.8386 [0.7519; 0.9220]	0.8372 [0.7572; 0.9181]		
Comorbidity	$\beta_2$	-0.2955 [-0.4166; -0.1753]	-0.3079 [-0.4365; -0.1835]	-0.3072 [-0.4403; -0.1741]		
<b>Event Process</b>	<b>Mean</b>	<b>[2.5; 97.5]% CI</b>	<b>Mean</b>	<b>[2.5; 97.5]% CI</b>	<b>Mean</b>	<b>[2.5; 97.5]% CI</b>
Age	$\theta_1$	0.7394 [0.6570; 0.8210]	0.7342 [0.6443; 0.8176]	0.7354 [0.6517; 0.8224]		
Gender (Male)	$\theta_2$	0.1813 [0.0569; 0.3175]	0.1787 [0.0545; 0.3058]	0.1836 [0.0501; 0.3107]		
Comorbidity	$\theta_3$	0.1958 [0.1371; 0.2512]	0.1961 [0.1365; 0.2539]	0.1960 [0.1363; 0.2547]		
Current Value	$\alpha_1$	-0.0011 [-0.0019; -0.0003]	0.0000 [-0.0018; 0.0019]	-0.0511 [-0.0842; -0.0175]		
Slope	$\alpha_2$		-0.0530 [-0.1328; 0.0228]			
<b>DIC</b>		74869.36	74879.86	74879.11		



**Figure 2.6.** Survival probability plots for two male patients aged 72 years and with two comorbidities at the index date using joint model  $M_4$ . The values of dichotomous time-varying covariate  $\mathbf{adherence}(t)$  are in the left part of both panels (stars) with sigmoid transformation of their linear predictors,  $\exp(\eta_i(t))/(1+\exp(\eta_i(t)))$  (line). In particular, at  $t = 12$  months  $\mathbf{adherence}_C(12) = 1$  and  $\mathbf{adherence}_D(12) = 0$ , with relative linear predictors  $\eta_C(12) = 9.64$  and  $\eta_D(12) = -7.64$ .

account. On the other hand, they allowed us to confirm that non-adherence is commonly associated with adverse health conditions [102]. However, this was what we expected and it did not represent the key result of the study. In fact, the added value of our work consists in performing an ongoing analysis and quantification of adherence effect on patient's outcome that allowed to carry out a real-time monitoring and profiling of patients.

In this sense, we need to assess which of the two models allowed for a better dynamic monitoring of patient's status. We observed that DIC value of model  $M_2$  was lower than the one of  $M_4$  (61372.08 vs 74869.36), which suggested that joint model  $M_2$  outperformed  $M_4$ . Then, we performed a 10-fold cross validation to assess the predictive performances of the models in terms of calibration, i.e., how well the model predicts the observed event rates [181], and discrimination, i.e., how well can the model discriminate between patients who had experience the event from patients who did not [152]. In terms of calibration, we evaluated the accuracy of predictions of survival models through the integrated prediction error that accounts for censoring, introduced by Schemper and Henderson (2000) [181]. In particular, the integrated predictor at time  $u$  giving the longitudinal measurements up to time  $t$  is indicated by  $IPE(u|t)$  and it is a weighted average of the expected prediction errors over interval  $[t, u]$ , i.e.,  $\{PE(s|t), t < s < u\}$ . The index  $PE(s|t)$  measures the predictive accuracies at specific time points  $s$ , considering the longitudinal information up to time  $t$ . In particular, using the available longitudinal data up to two different time points  $t_1 = 3$  and  $t_2 = 12$  months, we focused on two different time points of medical relevance for HF: mid-term mortality,  $u_1 = 12$  (1 year), and long-term mortality,  $u_2 = 60$  months (5 years). On the other hand, to assess the discriminative capability of each model we used the dynamic concordance index  $C_{dyn}^{\Delta t}(u)$  introduced by Rizopoulos

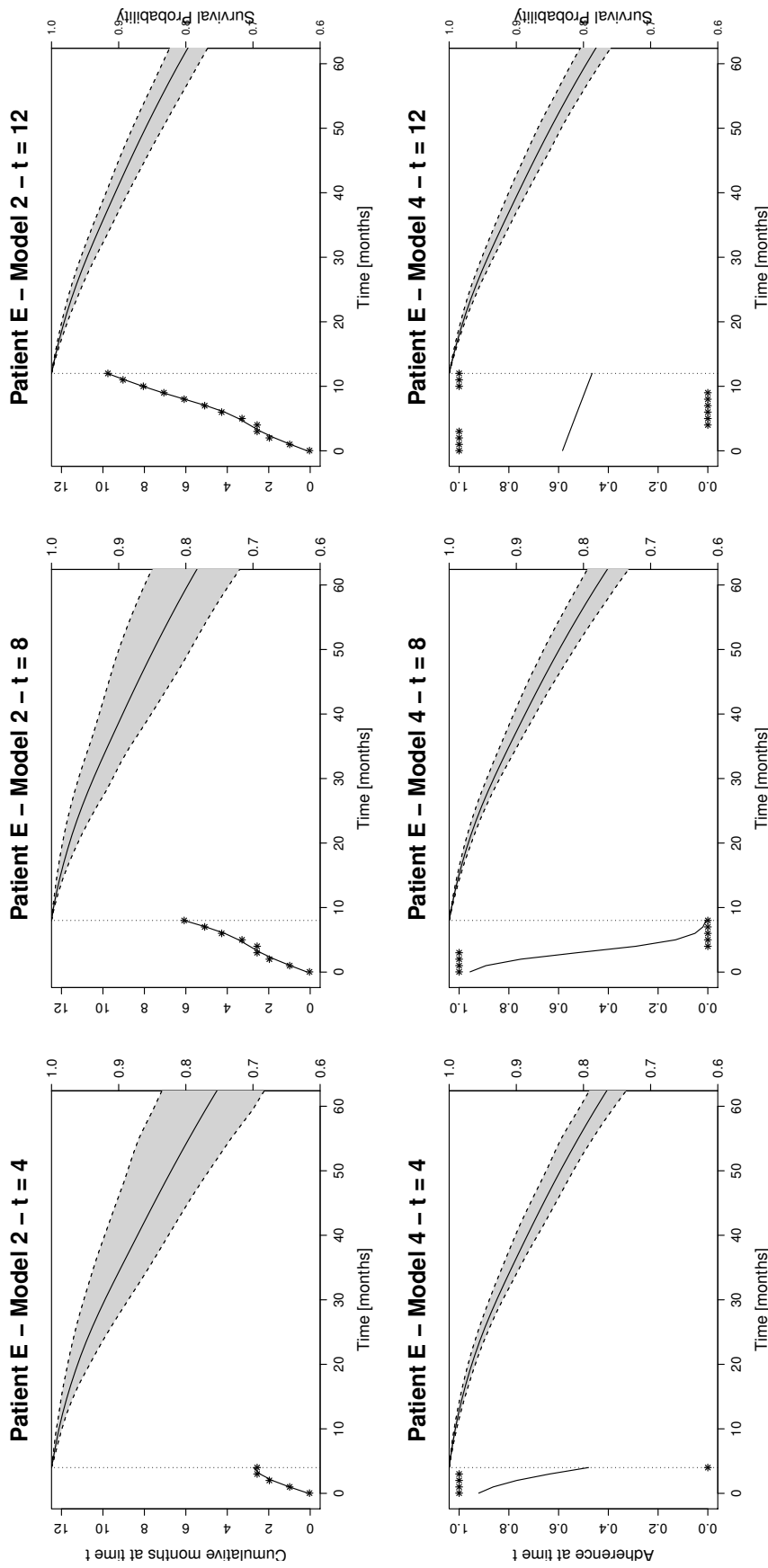
## 2. Joint modelling of time-varying adherence to medication and survival

**Table 2.3.** Estimated means along with standard deviations (s.d.) of the integrated prediction errors  $IPE(u|t)$  and the dynamic discrimination indexes  $C_{dyn}^{\Delta t}(u)$  computed through 10-fold cross-validation at time points  $u = 12$  months (1 year) and  $u = 60$  months (5 years) under the joint modelling analyses  $M_2$  and  $M_4$ .

	$M_2$	$M_4$
<b>Calibration</b>	<i>Mean (s.d.)</i>	<i>Mean (s.d.)</i>
$\widehat{IPE}(u = 12 t = 3)$	0.0095 (0.0052)	0.0095 (0.0052)
$\widehat{IPE}(u = 60 t = 12)$	0.1012 (0.0107)	0.1019 (0.0107)
<b>Discrimination</b>	<i>Mean (s.d.)</i>	<i>Mean (s.d.)</i>
$\widehat{C}_{dyn}^{\Delta t=1}(u = 12)$	0.7051 (0.0794)	0.6994 (0.0342)
$\widehat{C}_{dyn}^{\Delta t=6}(u = 60)$	0.6891 (0.0704)	0.6822 (0.0354)

(2016) [168], which is weighted average of the time-dependent areas under the receiver operating characteristic curves (AUCs) and takes into account the fact that not all the time points contribute equally, because at later time points less subjects are still available. In particular, we focused on the same time points  $u_1 = 12$  and  $u_2 = 60$  months, using one-month and six-months intervals  $\Delta t$ , respectively. For further details on integrated prediction error, dynamic concordance index and their estimates see Rizopoulos (2016) [168]. For each fold  $k$  with  $k \in \{1, \dots, 10\}$ , we computed the integrated prediction error  $IPE_k(u|t)$  and the dynamic concordance index  $C_{dyn,k}^{\Delta t}(u)$ , taking advantage of `prederrJM()` and `dynCJM()` functions implemented in the `JMbayes` package. Table 2.3 reports the means over the 10 folds along with standard deviations of the two indexes for both models  $M_2$  and  $M_4$ . We observed that the two models had comparable performances in terms of calibration and discrimination, but model  $M_2$  turned out slightly better (lower errors and higher concordances), confirming our suspects.

Finally, in Figure 2.7 we compared the survival probability plots for the same 72 year-old male patient  $E$  with two comorbidities at the index hospitalization and the pharmacological history shown in Figure 2.3. Top panels referred to model  $M_2$  and bottom panels to model  $M_4$ . We considered three different time points of the time-varying variables, i.e.,  $t \in \{4, 8, 12\}$  months (left, central and right panels, respectively). We observed that the two approaches led to two different behaviours of the survival probability plots during time. In particular, we noticed that the variability in survival predictions due to ongoing consumption was more informative and pronounced in model  $M_2$  than in model  $M_4$ , which was less able to capture and differently quantify the ongoing effect on patient's outcome. Indeed, looking at the ongoing behaviour of patient's ACE/ARB consumption, we observed that the patient assumed some drugs during the first three months but at time  $t = 4$  months he presented a non-adherence trend, with the long-term survival predictions showed in left panels. Then, he started to assume ACE/ARB again in order to improve his health status, but at time  $t = 8$  months he was still non-adherent to the therapy. That behaviour had a negligible impact on long-term prediction of model  $M_4$  (bottom-central panel), whereas the one of model  $M_2$  improved (top-central panel). He then continued to take the therapy, resulting adherent at time  $t = 12$  months. Also in that case, his behaviour had a negligible impact on long-term prediction of model  $M_4$  (bottom-right panel), whereas the one of model  $M_2$  further improved (top-right panel), also determining



**Figure 2.7.** Survival probability plots at different time points for the 72 year-old male patient  $E$  with two comorbidities at the index date and the pharmacological history shown in Figure 2.3 using joint model  $M_2$  (left panel) and  $M_4$  (right panel). The values of the time-varying covariates  $\text{cum\_months}_E(t)$  and  $\text{adherence}_E(t)$  are reported in the left part of the panels (stars) with the linear predictor for model  $M_2$  and the sigmoid transformation of linear predictor for model  $M_4$  (lines). Left panels are related to longitudinal process cut at time  $t = 8$  months and right panels at time  $t = 12$  months.

a reduction in the credibility intervals (and so in the uncertainty) of the survival prediction. This was probably due to the fact that dichotomous covariate `adherence(t)` is a poorer representation that only reflects the patient's purpose of taking the medication over time, whereas the continuous covariate `cum_months(t)` is able to capture the dynamic behaviour and shape of the consumption. Therefore the use of variable `cum_months(t)` in model  $M_2$  was preferable, since it provided a more detailed real-time monitoring of drug intake and of its effects on patient's outcome.

### 2.4. Final remarks

Since in pharmacotherapy practice the way adherence is usually computed discards valuable information related to the changes in patient drug utilization behaviour over time, in this chapter we proposed an innovative method to represent adherence to medication as time-varying covariate exploiting administrative database. In particular, we explored time-varying adherence to medication using two different representations: a continuous time-dependent variable, which indicated the cumulative months covered by drug assumption up to time  $t$ , and a dichotomous time-dependent variable, which indicates if the patient is adherent to the therapy at time  $t$ . For the computation, at each time-point  $t$  we took advantage of pharmacological records about drugs purchases collected in administrative databases, increasingly used for clinical and epidemiological purposes. These covariates were able to reflect the dynamics and the behaviour of adherence during the therapy, resulting more realistic and informative with respect to the commonly used baseline-fixed measures.

Once the covariates were determined, we applied the joint modelling technique in order to investigate how patients' time-to-event outcome was influenced by longitudinal data. We observed that modelling the drug intake process as time-varying covariates in a joint modelling setting represents an effective interpretative and forecasting approach for exploring the effects of adherence to medication on patients' survival, especially through the continuous time-varying representation. First of all, using both variables we confirmed that having a good adherence trend during time reflected a protective role on patients' survival, as we expected. Then, with a dynamic study of adherence, it was possible to real-time understand its effects on patient's health status directly monitoring the treatment, above all thanks to the use of the continuous time-dependent covariate able to satisfactorily capture the dynamic behaviour and shape of drug intake. A real-time monitoring and profiling of patients could allow to tailor therapeutic interventions and adjustments in order to prevent disease progression, leading to healthcare improvements, social benefits and economic utilities. In this sense, studying factors that could influence time-varying consumption, also through a deeper exploitation of administrative databases and a proper management of their population based massive records, could lead to interesting analysis and strong external validity. Furthermore, the use of a time-varying covariates into an appropriate survival framework, such as joint models, allowed to avoid the survival bias due to exclusion of early dying patients in the study cohort.

Some limitations of the present chapter have to be noted, mainly due to the use of secondary databases in the real case-study. First, the use of theoretical Defined Daily Dose (DDD) instead of Prescribed Daily Doses (PDD) could reflect a bias in the estimated adherence if the underlying PDD/DDD ratio is different from 1 [220, 69, 187], as mentioned in the previous chapter. It could be interesting to explore, whenever the linkage is possible, databases with information about dosages prescribed by doctors, in order to obtain a more realistic analysis of coverage periods. More in general, pharmacoepidemiology observational studies based on healthcare utilization databases are often characterized by potential biases, which can be divided in four categories according to [155]: confounding, selection bias, measurement bias and time-related biases. In particular, this study suffers from three main biases that usually occur in observational studies of pharmacoepidemiologic databases. First of all, HF patients are usually in a polytherapy, i.e., they usually take multiple drugs at the same time. Other treatments represent possible time-varying confounding factors, since they also influence the outcome of interest. The second issue concerned unmeasured confounding: our analysis was based on the information available in our dataset, and we could not control for other relevant not reported confounding factors, such as socio-economic or adverse drug reactions data. Finally, we could have biases related to the misclassification of exposure. Indeed, administrative data allowed to measure the effective consumption and adherence to medication with a big limitation: we were not able to assert if the patient was currently consuming the dispensed drug or if during re-hospitalizations period he/she actually received the treatment. These issues are related to the nature of administrative data: they address 'operational' goals, i.e., they are collected with no clinical question in mind and mainly for managerial and economic purposes [89], and the validity of using these kind of data is critically dependent on the reliability of the data [115, 180, 90]. Nevertheless, they are population based, comprehensive, capture real health system use, longitudinal and can be linked to other data, representing a valuable clinical research resource.

Despite the aforementioned limitations, this work opens doors for many further developments, both in the fields of statistical methods and clinical research. First of all, the considered models could be further improved (i) adding an autoregressive error in longitudinal submodel (2.3) in order to take into account the strong dependence of the value at previous time, (ii) exploring a more flexible longitudinal logistic mixed effects submodel (2.4) in which a nonlinear effect in time could allow for a better predictive ability of the model, and (iii) considering a nonlinear effect for demographics and comorbidity characteristics in the event submodels in order to allow for a better tailoring of predictions to different groups of patients. Nevertheless, such improvements present a number of issues in terms of convergence and patches to be added to the current version of *JMbayes* package (where autoregressive errors are not available), which go beyond the scope of the current work. For these reasons, point (i) was not implemented within this study, whereas points (ii) and (iii) were not pursued since their application encountered convergence issues. From a pharmacotherapy point of view, it will be necessary to simultaneously combine all the disease-modifying drugs for HF mentioned in [139] and [154] (ACE/ARB, Beta Blocking agents, Anti Aldosterone agents, Diuretics) since patients are usually in a polytherapy, as suggested by the decreasing mix of ACE/ARB drugs in case of increasing number of

## 2. *Joint modelling of time-varying adherence to medication and survival*

comorbidities. This would surely imply many issues related to the representation of the dynamic evolution of the multivariate time-varying datum and to include simultaneously all the treatments in a not trivial task. It could also be interesting to concomitantly analyse adherence to medication and, if available, other subject-specific measurements registered during follow-up, i.e., biomarkers. These measurements could be of clinical interest since they represent dynamic patterns that could reflect patient's disease progression, incorporating lots of information related to his health status and possibly leading to further improvements in subject-specific treatment and personalized medicine.

In summary, in this chapter we proposed a novel method to represent adherence to medication as time-varying covariate through administrative databases and we analysed its dynamic effect on patients' survival using a joint modelling framework. The developed approach is very flexible and can be generalized to many different settings. The main added value is the ongoing analysis and quantification of adherence effects on patient's outcome, which may allow researchers to proper modelling individual actual treatment, and clinicians to better target therapies for their patients. This study confirmed the importance of developing approaches to the representation of drugs consumption using a time-varying perspective, so that they are more realistic and informative than the commonly used time-fixed measures. In this sense, the modelling of time-varying covariates might be further exploited within the framework of functional data analysis [163, 162] or recurrent events theory [44]. In the next chapter, we propose an innovative methodology combining the exploitation of both as a first attempt to use these methods for an observational study in the pharmacotherapy field.

# Functional modelling of recurrent events on time-to-event processes

This chapter has been published in *Biometrical Journal*, 63(5):948–967, 2021 as M. Spreafico and F. Ieva “Functional modeling of recurrent events on time-to-event processes” [189].

In clinical practice many situations can be modelled in the framework of *recurrent events*, i.e., the repeated occurrence of the same type of events for the same patient over time. Chronic patients are usually involved in long-term therapies, that are often characterized by repeated situations like office visits, subsequent drug consumption, hospital admissions and many others. Examples include recurrences in breast cancer [174], asthma attacks [52], episodic relapses of follicular lymphoma [174], readmission after colorectal cancer [64, 36], epileptic seizures [215]. In patients with HF, two main types of events recur during treatment: (i) repeated consumption of multiple types of drugs and (ii) hospital readmissions [104, 21, 173]. Since models capable of simultaneously treating multiple drugs have not been well developed in pharmacotherapy, it could be interesting to concomitantly analyse more than one medication at the same time, along with re-hospitalizations events which usually herald a substantial worsening of patient’s survival prognosis. As discussed in Chapter 2, the natural and most appropriate way to look at these repeated events is to treat them as time-varying covariates, since their changing patterns over time could carry out information that may be related to patient’s health status and disease progression.

In biostatistical, epidemiological and medical literature, several approaches to analyse recurrent event data have been proposed and compared [202, 103, 44, 95, 106, 11, 149]. Different methods differ in the assumptions and in the interpretation of the results, but they all take into account the correlation between repeated events regarding the same individual. The most frequently applied method is the AG model by Andersen and Gill (1982) [12], which is an extension of the Cox proportional-hazard regression by Cox (1972) [46]. The AG model for recurrent events introduces the counting process formulation in terms of increments in the number of events along time. It assumes that the correlation between event times for an individual can be explained by past events, which share a common baseline hazard. In this way, the dependence could be captured by appropriate specification of time-varying covariates which are functions of the realisation of past events, such as the number of previous occurrences. This model is usually indicated for



analysing data when correlations among events for each individual are induced by measured covariate and the interest lies in the overall effect on the intensity of the occurrence of the event [11]. Two alternative approaches are the stratified Cox-type conditional [156] and marginal [213] models, which can incorporate both overall and event-specific effects for each covariate. The stratified conditional Prentice-Williams-Peterson (PWP) model analyses repeated events ordered by stratification, based on the prior number of events during the follow-up period. However, it can give unreliable estimates for higher order of events [11]. The stratified marginal Wei-Lin-Weissfeld (WLW) model ignores the order of occurrence of the events. Therefore, an individual is at risk for every event as long as he/she is under observation, even if no previous events occurred, leading to a ‘carry-over effect’ as explained by [103] and [149]. As a further alternative, Cox model can be also extended using frailty models [85, 202, 175, 176, 44, 106, 60], in which a random covariate that induces dependence among the event times is introduced. This approach assumes that recurrent event times are independent conditional on the covariates and the random effects, and it is used to model individual patients’ heterogeneity in the baseline hazards. Furthermore, approaches able to connect several event processes (recurrent and fatal/non-fatal ones) have been proposed. Among others, (copula-based) joint frailty models [177, 174, 54, 124, 125] allow the prediction of a terminal event time given recurrent event times. Alternatively, rate-based models [34, 197, 37, 224, 196] or multi-state models [13, 44] can be used in case of multiple types of recurrent events. The choice of the proper approach for the analysis of recurrent event data will therefore be determined by many factors, including among others, number and types of events, relationship between subsequent events and biological processes [11].

Aforementioned methods are used to analyse single or several event processes, possibly connecting them to another event of interest. However, none of these approaches has been used to extrapolate information from repeated events in the form of dynamic functional covariates, and then study how these covariates affect other specific events, such as patient’s death. In this framework, Baraldo *et al.* (2013) [21] proposed a method to model the realized trajectories of the cumulative hazard functions underlying a recurrent event process of interest (i.e., hospital readmissions in time). Estimated trajectories were treated as functional data and included into a generalized linear model to predict a binary telemonitoring outcome. However, the authors focused only on a counting process formulation for recurrent events, without considering further information about them. Indeed, many situations and events are characterized by both a location (in time or space) and a weight or other distinguish attribute, called *mark* [47]. For example, in HF treatment a longer period in hospital could reflect the aggravation of patient’s health condition, as well as a shorter drug coverage period could lead to nonadherence to therapy, commonly associated with adverse health conditions [102, 187, 188]. The development of models and methods able to deal with all these peculiar aspects is of statistical interest and of clinical relevance.

Motivated by the clinical question concerning the effect of re-hospitalizations and subsequent consumption of different drugs on survival in HF patients, in this chapter we proposed a new methodology that exploits recurrent events modelling [44], point processes

theory [113, 47] and Functional Data Analysis (FDA) [162] to represent time-varying events in terms of functions, plugging them into a suitable functional Cox model for overall survival. In order to take into account many aspects that could influence the events, our idea was to look at time-varying recurrent events as particular non-stationary stochastic counting processes which can depend on their marks, i.e., marked point processes [113, 47]. Starting from the idea by [21], we developed a *marked point process formulation for recurrent events* to compute the realized trajectories of the cumulative hazard functions (i.e., the *compensators*) underlying specific counting processes of interest, allowing the dependence on the marks. In particular, among the aforementioned methods to deal with recurrent events, we modelled the compensators through AG models [12], ending up with functional data that represent the dynamic evolution of the events risk. Then, we applied Functional Principal Component Analysis (FPCA) [162] in order to perform a dimensionality reduction and summarise information emerging from the functional compensators to a finite set of covariates, while losing a minimum part of the information. This information was finally included into a functional linear Cox regression model [109], extended to the case of multiple functional predictors.

The procedure presented in this chapter can hence be divided into two phases:

- (i) the representation of time-varying functional compensators,
- (ii) the modelling of such covariates in a time-to-event framework.

In doing so, we aimed to enrich the information available for modelling survival with relevant dynamic features, as well as to provide a new setting for quantifying the association between time-varying processes and patients' overall survival.

The remaining part of the chapter is organized as follows. In Section 3.1 we describe the real study design used in this work. In Section 3.2 we present the whole methodology. First, we focus on the main novelty introduced by the present work, i.e., the *marked point process formulation for recurrent events* to represent the compensators (Section 3.2.1). Then we introduce the functional linear Cox regression model for overall survival in case of multiple functional predictors (Section 3.2.2). In Section 3.3 we apply the proposed methodology to HF administrative database provided by *Regione Lombardia - Healthcare Division* [164]. Finally, Section 3.4 contains some concluding remarks, discussion of strengths and limitations of the proposed approach and opportunities for future work. Statistical analyses were performed in the R software environment [161]. Source code is available as Supporting Information of [189].

## 3.1. Materials and Administrative data

### 3.1.1. Administrative data sources

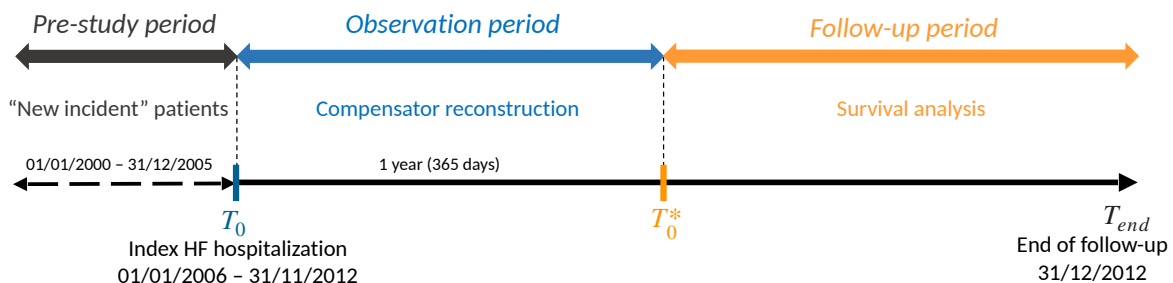
As in Chapter 2, in this work we focused on a representative sample of the real administrative *HFData* database [136] provided by *Regione Lombardia - Healthcare Division* [164]

related to non-paediatric patients living in Lombardy with their first HF discharge (*index event*) between January 2006 to December 2012. As explained in Section 2.2.2, patients' clinical history of hospitalizations or drug consumption could be reconstructed using secondary registry data related to (i) patient admission to hospital (i.e., date of discharge from hospital, length of stay in hospital) and (ii) pharmaceutical purchases (i.e., ATC code, date of purchase, number of treatment days covered by the prescription). Among the disease-modifying drugs for HF patients [138, 139, 154], we focused on polypharmacy treatment as a combination of Angiotensin-Converting Enzyme (ACE) inhibitors, Beta-Blocking (BB) agents and Anti-Aldosterone (AA) agents.

### 3.1.2. Study design and outcome measure

Figure 3.1 shows the study design. A 5-years *pre-study period* from 2000 to 2005 was used in order to consider only "incident" HF patients, i.e., patients with no contacts with healthcare system in the previous five years due to HF. The study-period started from the first discharge for HF (time  $T_0$  in Figure 3.1) and was divided into the *observation period* (365 days from the index date), used for the compensators reconstruction, and the *follow-up period*, used for the survival analysis, whose starting time was  $T_0^* = T_0 + 365$ . The modelling of the compensators related to the stochastic processes of interest regarded the time interval  $[T_0; T_0^*]$  in Figure 3.1. Therefore, only patients alive at the end of the *observation period* were selected in the study cohort and followed up to observe survival outcomes. We underline that this choice, necessary for the reconstruction of compensator trajectories, could imply a survival bias in case of the exclusion of too many early dying patients (that is not our case since only 6.8% of patients died during the observation period).

Study outcome of interest was patient's death for any cause. Deaths were collected from the Hospital Discharge Forms Database (for in-hospital deaths) or Vital Statistics Regional Dataset (for out-hospital deaths). Overall survival was measured from the end of



**Figure 3.1.** Study design for a HF patient of the study cohort. The *pre-study period* is used to define "incident" HF patients. The *observation period* is used for the selection of patient's clinical history and the compensators reconstruction. The *follow-up period* is used for survival analysis.  $T_0$  is the time instant the patient is discharged by her/his first hospitalization and enrolled into the current study.  $T_0^* = T_0 + 365$  is the starting time of the follow-up.  $T_{end}$  is the minimum between the death or the administrative censoring (December 31<sup>st</sup>, 2012).

the *observation period* ( $T_0^*$  in Figure 3.1) to the date of death or to the administrative censoring date (December 31<sup>st</sup>, 2012). Outcome  $(T_i, D_i)$  denotes the observed time-to-death data of patient  $i \in \{1, \dots, n\}$ , where  $T_i = \min(T_i^*, C_i)$  is the observed event time,  $T_i^*$  is the true event time,  $C_i$  is the censoring time and  $D_i = \mathbb{1}(T_i^* \leq C_i)$  is the event indicator, with  $I(\cdot)$  being the indicator function that takes the value 1 when  $T_i^* \leq C_i$ , and 0 otherwise. Independent censoring between true death and censoring times was assumed.

## 3.2. Statistical Methodologies

We now introduce the methodology developed and then applied on the case study of interest in Section 3.3. In Section 3.2.1 we focus on the main novelty introduced by the present work, i.e., the *marked point process formulation for recurrent events*. In Section 3.2.2 we introduce the functional linear Cox regression model for overall survival in case of multiple functional compensators.

### 3.2.1. Marked point process formulation for recurrent events

A recurrent event process is characterized by an increasing sequence of *events times*, where each element denotes the time of the corresponding event [44]. To this sequence of times could be associated (i) a *counting process* that at time  $t$  records the cumulative number of events occurred up to  $t$  [44] and (ii) other random elements, called *marks*, containing further information about the events [113, 47]. Marks can also be thought of as the size, weight or magnitude related to the jumps of the counting process. Extending the approach by [21], we now introduce the *marked point process formulation for recurrent events* to compute the realized trajectories of the compensators underlying a specific counting process of interest, allowing the dependence on the marks.

Let us consider a set  $\mathcal{M}$  of recurrent events for a set of  $n$  individuals as stochastic processes. For each patient  $i \in \{1, \dots, n\}$ , let  $\{t_{i,j}^{(m)}, j = 0, 1, \dots, n_i^{(m)}\}$  be the increasing sequence of event times related to recurrent event process  $m$ , where  $n_i^{(m)}$  is the total number of events of type  $m$  experienced by the  $i$ -th subject,  $t_{i,j}^{(m)}$  denotes the time of the  $j$ -th event and  $t_{i,0}^{(m)} = 0 \forall i, m$ . Let  $\mathbf{w}_i^{(m)}$  be the vector of marks elements, where each *jump mark*  $w_{i,j}^{(m)}$  is the magnitude of the information associated to each *jump time*  $t_{i,j}^{(m)}$ . The observations (possibly censored) may be considered as the realisation of  $N_1^{(m)}, \dots, N_n^{(m)}$  processes, where  $N_i^{(m)}$  is the stochastic process which counts the observed events (or jumps) of the process  $m$  in the *observation period* related to the  $i$ -th individual. According to the Doob–Meyer (D-M) decomposition theorem [142], each counting process  $N_i^{(m)}(t)$ , adapted to the filtration  $\{\mathcal{F}_{t,i}^{(m)}, t \geq 0\}$  representing the history of realisations of the process itself, can be seen as:

$$N_i^{(m)}(t) = M_i^{(m)}(t) + \Lambda_i^{(m)}(t) = M_i^{(m)}(t) + \int_0^t \lambda_i^{(m)}(s) ds \quad (3.1)$$

### 3. Functional modelling of recurrent events on time-to-event processes

where  $M_i^{(m)}(t)$  is a zero-mean uniformly integrable martingale which represents the residual of the process, and  $\Lambda_i^{(m)}(t) = \int_0^t \lambda_i^{(m)}(s)ds$  is a unique predictable, non-decreasing, *cadlag* (right-continuous with left limits) and integrable process, i.e., the *compensator* (or *cumulative hazard*). Process  $\lambda_i^{(m)}(t)$  is the *conditional intensity function*, in which we omitted the conditioning with respect to the history  $\mathcal{F}_{t,i}^{(m)}$  for ease of notation, and represents the infinitesimal risk of occurrence of an event  $m$  at time  $t$ , given the history, i.e.,  $\lambda_i^{(m)}(t) = \lim_{\Delta t \rightarrow 0} \mathbb{E} \left[ N_i^{(m)}(t + \Delta t) - N_i^{(m)}(t) | \mathcal{F}_{t,i}^{(m)} \right] / \Delta t$ . The compensator  $\Lambda_i^{(m)}(t)$  may be thought of as a positive non-decreasing  $L^2$ -function over the temporal domain and will be the core of our modelling effort.

A counting process where jumps may have different size can be modelled as a marked point process, assuming that a given distribution regulates the size of the jumps. A marked point process is the couple of processes describing the behaviour of jump times and marks modelled through the *conditional intensity function*  $\lambda_i^{(m)}(t, \mathbf{w}_i^{(m)})$ , i.e., the infinitesimal risk of occurrence of event  $m$  at time  $t$  with marks  $\mathbf{w}_i^{(m)}$  given the history:

$$\lambda_i^{(m)}(t, \mathbf{w}_i^{(m)}) = \lambda_{ig}^{(m)}(t) f_i^{(m)}(\mathbf{w}_i^{(m)}) \quad (3.2)$$

where  $\lambda_{ig}^{(m)}$  is the intensity process of the counting process, also called *ground intensity*, and  $f_i^{(m)}$  is the multivariate density of the marks  $\mathbf{w}_i^{(m)}$ . Using this formulation, conditional independence of jump times and marks is assumed. Note that, if  $\lambda_i^{(m)}(t, \mathbf{w}_i^{(m)})$  is properly modelled, the D-M decomposition in (3.1) is still valid in the marked point process framework considering Equation (3.2) as conditional intensity process.

To handle recurrent events and allow predictors to change over time, we use the counting process formulation for recurrent events introduced by [12], also called AG model for recurrent events, assuming a particular distribution for the marks in order to ease computations. In particular, we assume that the density  $f_i^{(m)}$  depends on some time-dependent features related to the marks  $\mathbf{w}_i^{(m)}$ . Under these hypotheses, for each event  $m$  the conditional intensity function  $\lambda_i^{(m)}(t, \mathbf{w}_i^{(m)})$  in Equation (3.2) related to patient  $i$  takes the form:

$$\begin{aligned} \lambda_i^{(m)}(t, \mathbf{w}_i^{(m)}) &= Y_i^{(m)}(t) \lambda_0^{(m)}(t) \exp \left\{ \boldsymbol{\beta}^{(m)T} \mathbf{x}_i^{(m)}(t) \right\} \exp \left\{ \boldsymbol{\gamma}^{(m)T} \mathbf{z}_i^{(m)}(t) \right\} \\ &= Y_i^{(m)}(t) \lambda_0^{(m)}(t) \exp \left\{ \boldsymbol{\beta}^{(m)T} \mathbf{x}_i^{(m)}(t) + \boldsymbol{\gamma}^{(m)T} \mathbf{z}_i^{(m)}(t) \right\} = \lambda_i^{(m)}(t) \end{aligned} \quad (3.3)$$

where  $\mathbf{x}_i^{(m)}(t)$  and  $\mathbf{z}_i^{(m)}(t)$  are the possibly time-dependent vectors of covariates of the  $i$ -th individual, the latter related to the marks  $\mathbf{w}_i^{(m)}$ . Parameters  $\boldsymbol{\beta}^{(m)}$  and  $\boldsymbol{\gamma}^{(m)}$  are fixed vectors of coefficients,  $\lambda_0^{(m)}$  is the baseline hazard function shared across patients and  $Y_i^{(m)}$  is a predictable process taking values in  $\{0, 1\}$ . Whenever  $Y_i^{(m)} = 1$ , the  $i$ -th individual is under observations, i.e.,  $Y_i^{(m)}$  takes the role of the censoring variable.

Parameters  $\boldsymbol{\beta}^{(m)}$  and  $\boldsymbol{\gamma}^{(m)}$  are estimated maximizing the partial likelihood function constructed given the history, using a counting process approach [12]. The baseline cumulative

hazard  $\Lambda_0^{(m)}(t) = \int_0^t \lambda_0^{(m)}(s) ds$  can be estimated  $\forall m \in \mathcal{M}$  using the *Breslow estimator* [32]  $\hat{\Lambda}_0^{(m)}(t)$ , which returns a step-function. However, since true underlying functions  $\Lambda_0^{(m)}(t)$  are absolutely continuous, we smooth the estimates using the approach adopted in [21], obtaining regularised version of  $\Lambda_0^{(m)}(t)$ , namely  $\tilde{\Lambda}_0^{(m)}(t)$ .

Let us now consider the sequence  $0 = t_{i,0}^{(m)} < t_{i,1}^{(m)} < \dots < t_{i,N_i^{(m)}(\tau)}^{(m)}$  of realised jump times related to process  $N_i^{(m)}(t)$ , with  $\tau$  equal to the censoring time (possibly equal for all individuals or not) and  $n_i^{(m)} = N_i^{(m)}(\tau) \forall m, i$ . In our case,  $\tau$  is the censoring time of the *observation period*, i.e.,  $T_0$  in Figure 3.1. We can express the realisations of each compensator  $\Lambda_i^{(m)}(t)$  for the process  $m$  of the  $i$ -th patient as a function of  $\Lambda_0^{(m)}(t)$ ,  $\beta^{(m)}$  and  $\gamma^{(m)}$ :

$$\begin{aligned} \Lambda_i^{(m)}(t) &= \int_0^t \lambda_i^{(m)}(s) ds = \int_0^t Y_i^{(m)}(s) \lambda_0^{(m)}(s) \exp \left\{ \beta^{(m)T} \mathbf{x}_i^{(m)}(s) + \gamma^{(m)T} \mathbf{z}_i^{(m)}(s) \right\} ds \\ &= \sum_{j=1}^{N_i^{(m)}(t)} \int_{t_{i,j-1}^{(m)}}^{\min(t_{i,j}^{(m)}, t)} \lambda_0(s) \exp \left\{ \beta^{(m)T} \mathbf{x}_i^{(m)}(t_{i,j-1}) + \gamma^{(m)T} \mathbf{z}_i^{(m)}(t_{i,j-1}) \right\} ds \\ &= \sum_{j=1}^{N_i^{(m)}(t)} \exp \left\{ \beta^{(m)T} \mathbf{x}_i^{(m)}(t_{i,j-1}) + \gamma^{(m)T} \mathbf{z}_i^{(m)}(t_{i,j-1}) \right\} \left[ \Lambda_0^{(m)} \left( \min \left( t_{i,j}^{(m)}, t \right) \right) - \Lambda_0^{(m)} \left( t_{i,j-1}^{(m)} \right) \right]. \end{aligned} \quad (3.4)$$

An estimate of the compensator in Equation (3.4) can be then obtained as:

$$\hat{\Lambda}_i^{(m)}(t) = \sum_{j=1}^{N_i^{(m)}(t)} \exp \left\{ \hat{\beta}^{(m)T} \mathbf{x}_i^{(m)}(t_{i,j-1}) + \hat{\gamma}^{(m)T} \mathbf{z}_i^{(m)}(t_{i,j-1}) \right\} \left[ \tilde{\Lambda}_0^{(m)} \left( \min \left( t_{i,j}^{(m)}, t \right) \right) - \tilde{\Lambda}_0^{(m)} \left( t_{i,j-1}^{(m)} \right) \right] \quad (3.5)$$

where  $\hat{\beta}^{(m)}$  and  $\hat{\gamma}^{(m)}$  are the estimated vectors of coefficients and  $\tilde{\Lambda}_0^{(m)}(t)$  is the smoothed estimate of the cumulative baseline hazard.

To check the fitting of  $\hat{\Lambda}_i^{(m)}(t)$ , we have to verify whether the estimates of martingale residuals  $M_i^{(m)}(t)$  involved in the D-M decomposition (3.1), i.e., the residuals [203] given by

$$\hat{M}_i^{(m)}(t) = \hat{\Lambda}_i^{(m)}(t) - N_i^{(m)}(t), \quad (3.6)$$

may be effectively considered as realisations of zero-mean processes. In order to do so, we can plot the residuals evaluated in the whole *observation period* and check if the average residual curve  $\bar{M}^{(m)}(t) = \frac{1}{n} \sum_{i=1}^n \hat{M}_i^{(m)}(t)$  is approximately close to 0 over time.

This formulation extends the one proposed in [21], allowing the counting processes to depend on their marks and setting up a framework for multiple processes to be considered. In fact, applying this procedure  $\forall m \in \mathcal{M}$ , we end up with a multivariate time-dependent data  $\left\{ \Lambda_i^{(m)} \right\}_{m \in \mathcal{M}}$  for each patient  $i$ , characterizing her/his recurrent events dynamics during the *observation period*  $[T_0; T_0^*]$ . These compensator trajectories may be thought of as patient-specific time-varying covariates and, mathematically speaking, as positive non-decreasing  $L^2$ -functions over the temporal domain  $[T_0; T_0^*]$ .

### 3.2.2. Functional linear Cox regression model with multiple functional compensators

To include the functional compensators into a survival model, the functional linear Cox regression model introduced by Kong *et al.* (2018) [109] can be extended to the case of multiple functional predictors, i.e., Multivariate Functional Linear Cox Regression Model (MFLCRM). For each patient  $i$ , let  $\{\Lambda_i^{(m)}\}_{m \in \mathcal{M}}$  be the realizations of the  $|\mathcal{M}|$ -variate compensators related to a set  $\mathcal{M}$  of recurrent events. The functional compensators are included in the hazard function of Cox model [46] as:

$$h_i \left( t | \boldsymbol{\omega}_i, \left\{ \Lambda_i^{(m)} \right\}_{m \in \mathcal{M}} \right) = h_0(t) \exp \left\{ \boldsymbol{\theta}^T \boldsymbol{\omega}_i + \sum_{m \in \mathcal{M}} \int_{S_m} \Lambda_i^{(m)}(s) \alpha^{(m)}(s) ds \right\} \quad (3.7)$$

where  $h_0(t)$  is the baseline hazard function,  $\boldsymbol{\omega}_i$  is the vector of scalar (non functional) covariates with regression parameters  $\boldsymbol{\theta}$ . The realizations  $\{\Lambda_i^{(m)}\}_{m \in \mathcal{M}}$  are defined over the temporal domains  $S_m = [T_0; T_0^*] \forall m$ . Parameters  $\alpha^{(m)}(s)$  denote the functional regression coefficients.

By applying Functional Principal Component Analysis (FPCA) [162], each functional compensator  $\Lambda_i^{(m)}(s)$  can be approximated with a finite sum of  $K_m$  orthonormal basis  $\{\xi_1^{(m)}, \dots, \xi_{K_m}^{(m)}\}$ :

$$\Lambda_i^{(m)}(s) \approx \mu^{(m)}(s) + \sum_{k=1}^{K_m} f_{ik}^{(m)} \xi_k^{(m)}(s) \quad (3.8)$$

where  $\mu^{(m)}(s)$  is the functional compensator mean and  $f_{ik}^{(m)}$  is the FPC score of individual  $i$  related to the  $k$ -th orthonormal base  $\xi_k^{(m)}$  and  $K_m$  is the truncation parameter, representing the number of FPCs to be considered. In particular, the score  $f_{ik}^{(m)}$  represents the projection of the  $i$ -th functional observation  $\Lambda_i^{(m)}(t)$  related to event  $m$  along the direction of the  $k$ -th principal component  $\xi_k^{(m)}(t)$ . From (3.8) the integrals in (3.7) can be approximated considering:

$$\begin{aligned} \int_{S_m} \left[ \Lambda_i^{(m)}(s) - \mu^{(m)}(s) \right] \alpha^{(m)}(s) ds &\approx \int_{S_m} \sum_{k=1}^{K_m} f_{ik}^{(m)} \xi_k^{(m)}(s) \alpha^{(m)}(s) ds \\ &= \sum_{k=1}^{K_m} f_{ik}^{(m)} \int_{S_m} \xi_k^{(m)}(s) \alpha^{(m)}(s) ds = \sum_{k=1}^{K_m} f_{ik}^{(m)} \alpha_k^{(m)} \end{aligned} \quad (3.9)$$

where  $\alpha_k^{(m)}$  is the scalar representing the quantity  $\int_{S_m} \xi_k^{(m)}(s) \alpha^{(m)}(s) ds$ . Introducing approximation (3.9) in Equation (3.7), the hazard function becomes:

$$\begin{aligned} h_i \left( t | \boldsymbol{\omega}_i, \left\{ \Lambda_i^{(m)} \right\}_{m \in \mathcal{M}} \right) &= h_0(t) \exp \left\{ \boldsymbol{\theta}^T \boldsymbol{\omega}_i + \sum_{m \in \mathcal{M}} \left[ \int_{S_m} \mu^{(m)}(s) \alpha^{(m)}(s) ds + \sum_{k=1}^{K_m} f_{ik}^{(m)} \alpha_k^{(m)} \right] \right\} \\ &= h_0^*(t) \exp \left\{ \boldsymbol{\theta}^T \boldsymbol{\omega}_i + \sum_{m \in \mathcal{M}} \sum_{k=1}^{K_m} f_{ik}^{(m)} \alpha_k^{(m)} \right\} \end{aligned} \quad (3.10)$$

where  $h_0^*(t) = h_0(t) \exp \left\{ \sum_{m \in \mathcal{M}} \int_{S_m} \mu^{(m)}(s) \alpha^{(m)}(s) ds \right\}$  is the baseline hazard function and  $\alpha_k^{(m)} = \int_{S_m} \xi_k^{(m)}(s) \alpha^{(m)}(s) ds$  is the regression parameter related to the  $k$ -th FPC score of the functional compensator of event  $m$ . Therefore, defining the following quantities:

$$\tilde{\boldsymbol{\theta}} = \left[ \boldsymbol{\theta}^T, \left\{ \left( \alpha_1^{(m)}, \dots, \alpha_{K_m}^{(m)} \right) \right\}_{m \in \mathcal{M}} \right]^T$$

$$\tilde{\boldsymbol{\omega}}_i = \left[ \boldsymbol{\omega}_i^T, \left\{ \left( f_{i1}^{(m)}, \dots, f_{iK_m}^{(m)} \right) \right\}_{m \in \mathcal{M}} \right]^T$$

and substituting them in Equation (3.10), through FPCA the MFLCRM can be expressed as Cox model with hazard function

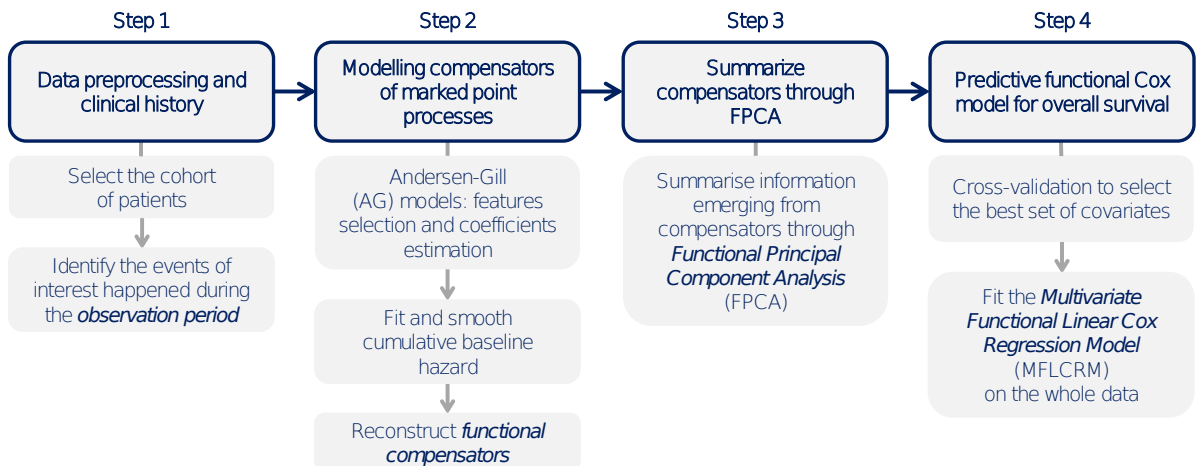
$$h_i(t|\tilde{\boldsymbol{\omega}}_i) = h_0(t) \exp \left\{ \tilde{\boldsymbol{\theta}}^T \tilde{\boldsymbol{\omega}}_i \right\}.$$

All the properties of the Cox model still hold in this framework and the vector of coefficients  $\tilde{\boldsymbol{\theta}}$  can be estimated by maximising the partial likelihood function [46]. In R software [161] the MFLCRM can be fitted through `coxph` function of package `survival` by [201].

In this analysis, the truncation parameters  $K_m$ , representing the number of FPCs to be considered for each event  $m$ , are chosen through a 10-fold cross validation procedure to select the best set of covariates among patients' baseline characteristics  $\boldsymbol{\omega}_i$  and scores  $f_{ik}^{(m)}$ , according to the highest Concordance Index [151].

The entire procedure may be resumed in four steps, as shown in Figure 3.2:

- Steps 1 and 2 are devoted to reconstruct the compensators of suitable marked point processes as time-varying (functional) covariates;
- Steps 3 and 4 set up a suitable framework for including such time-varying covariates in a time-to-event model.



**Figure 3.2.** Summary of the entire methodological procedure presented in Section 3.2.



### 3.3. Data application

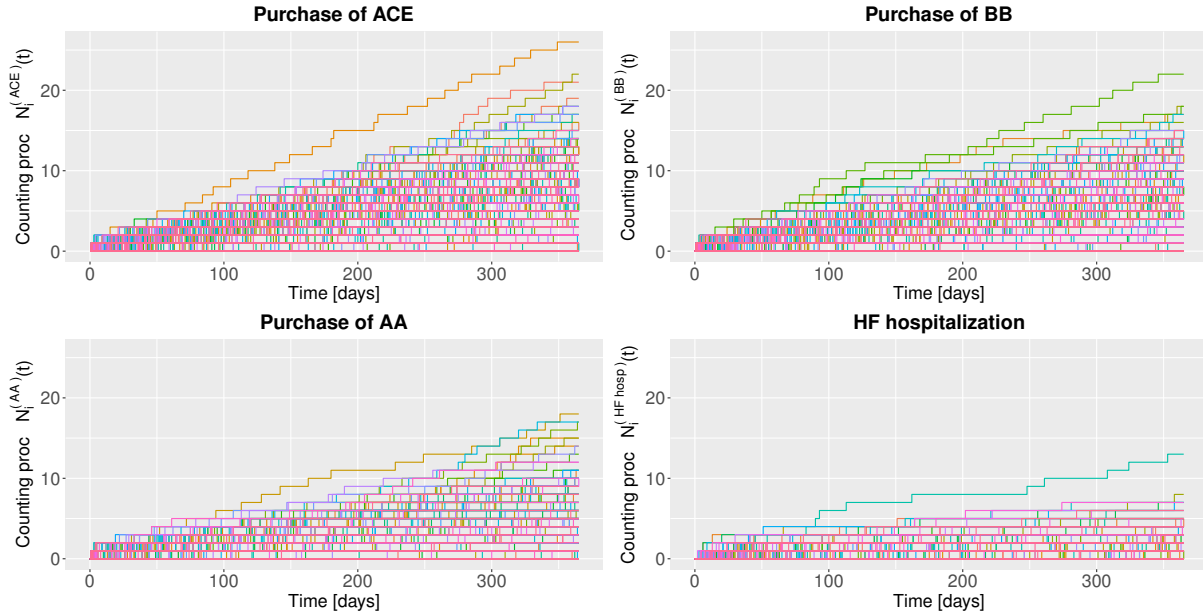
We now proceed with the application of the methodology described in Section 3.2/Figure 3.2 to the administrative database of *Lombardy Region*, in order to study how re-hospitalizations and multiple drugs consumption processes affect overall survival in HF patients. R source code is available as Supporting Information of [189].

#### 3.3.1. Step 1: Data preprocessing & clinical history

We focused on a representative sample of the administrative database of Lombardy Region related to 4,872 patients with their first HF discharge between January 2006 to December 2012. Excluding patients who died during the *observation period*, a final cohort of  $n = 4,541$  (93.2%) patients was selected. Overall, at index hospitalization, mean age of the study cohort was 73.98 years ( $s.d. = 11.37$ ) with a percentage of male patients equal to 54.4% (2,466 patients). The median value of overall survival was 37.32 (IQR = [20.53; 54.93]) months. At administrative censoring date 1,200 patients (26.4%) were dead and 3,341 (73.6%) were censored.

We identified four stochastic processes of interest: hospitalizations due to HF, purchases of ACE, BB and AA drugs, identified by their ATC codes. Hence, the set of recurrent events of interest was  $\mathcal{M} = \{m : ACE, BB, AA, HF\ hosp\}$ . In particular, we selected only events within the 1-year *observation period* (censoring time  $\tau = T_0^*$ ). For each patient  $i \in \{1, \dots, n = 4,541\}$ , repeated events of process  $m$  were modelled as a marked point process  $N_i^{(m)}(t)$ , with *jump times*  $t_{i,j}^{(m)}$  equal to event times (i.e., date of  $j$ -th admission in hospital or date of  $j$ -th drug purchase) and *jump marks*  $w_{i,j}^{(m)}$  equal to the length of stay in hospital or the duration of drug coverage respectively, where  $j \in \{0, 1, \dots, N_i^{(m)}(\tau)\}$ . Figure 3.3 shows the counting processes  $N_i^{(m)}(t)$  describing ACE purchase (top-left panel), BB purchase (top-right panel), AA purchase (bottom-left panel) and HF hospitalization (bottom-right panel) for a sample of 500 HF patients belonging to the administrative database. Overall, at the end of the observation period (time  $t = \tau = T_0^*$ ), the most frequent events were ACE and BB purchases: 2,916 patients (64.2%) purchased ACE at least once with a median of 4 purchases (IQR = [0;8]), and 2,890 patients (63.6%) purchased BB at least once with a median of 4 purchases (IQR = [0;7]), where the median number of events  $m$  at time  $\tau$  is given by  $median_{i \in \{1, \dots, n\}} N_i^{(m)}(\tau)$ . Purchase of AA and hospitalization due to HF were less frequent: 2,007 patients (44.2%) purchased AA at least once with a median of no purchases (IQR = [0;4]) and 2,699 patient (59.4%) were re-hospitalized due to HF, with a median of 1 HF hospitalization (IQR = [0;2]).

In order to proceed with the analyses, we reformatted the administrative data as explained in Appendix A.1.



**Figure 3.3.** Representation of counting processes  $N_i^{(m)}(t)$  related to purchases of ACE inhibitors (top-left panel), BB agents (top-right panel), AA (bottom-left panel) and of HF hospitalizations (bottom-right panel) during the observation period for a sample of 500 HF patients belonging to the administrative database. Each non-decreasing step function is related to a different patient.

### 3.3.2. Step 2: Modelling compensators of marked point processes

We can now reconstruct the compensators of the marked point processes for recurrent events, as explained in Section 3.2.1. For each process  $m \in \{ACE, BB, AA, HF hosp\}$ , we first select the best set of features for the AG model for recurrent events in Equation (3.3) using 10-fold cross validation and we estimate the selected coefficients on the whole dataset. Then, we fit and smooth cumulative baseline hazard using the constrained B-spline smoothing algorithm introduced by [73]. Finally, we reconstruct the compensator trajectories as functions of the estimated coefficients and of the smoothed estimate of the cumulative baseline hazard through Equation (3.5).

#### Features selection and coefficients estimation

For each process  $m \in \{ACE, BB, AA, HF hosp\}$ , we used as covariates  $\mathbf{z}_i^{(m)}(t)$  of patient  $i$ : the time-dependent variable *enum* which indicates the number of events related to process  $m$  occurred in the past and the time-dependent variable *marks* representing the sum of the corresponding marks. Also the logarithmic transformations (shifted away from 0) of the same variables, i.e.,  $\log(enum+1)$  and  $\log(marks+1)$ , and respective interactions, were considered. Adjustments for *age* and *gender* at baseline were performed. The vector of all the covariates considered for the model is indicated by  $\mathbf{x}_i^{(m)}(t)$ . In particular, for each process  $m$  we performed a 10-fold cross-validation to determine the best sets of features according to the lowest Mean Absolute Martingale Residual (MAMR) (see Appendix A.2 for details). Once covariates were selected, we fitted four AG models in Equation (3.3),

### 3. Functional modelling of recurrent events on time-to-event processes

one for each process  $m$ , using the selected features on the entire dataset to estimate coefficients  $\hat{\beta}^{(m)}$  and  $\hat{\gamma}^{(m)}$ .

In Table 3.1 selected features, hazard ratios and corresponding 95% CI are reported. Among all the models tested through the cross-validation procedure, features related to *enum*, *marks* and their interaction were selected and their coefficients were always significantly different from 0. In particular, the procedure selected the original features for HF hospitalization process ( $m = HF\ hosp$ ) and their logarithmic transformations for drug purchases ( $m \in \{ACE, BB, AA\}$ ). This was probably due to the fact that hospitalizations were rarer than drug purchases, so they might have a greater effect in increasing the risk of experiencing a new event. The signs of the fitted coefficients relative to these three types of features were consistent throughout the four processes, allowing us to give similar interpretations. HRs related to the number of past events *enum* and to the sum of the past marks *marks* were greater than 1. This could be interpreted as a “self-exciting” behaviour: many events (drug purchases or hospitalizations) in the past and higher marks (the purchase of big quantities of drug or having spent longer periods in hospital) both increase the risk of a new event. HR related to the interaction terms were lower than 1, meaning that (i) in case of the same number of events, the increase in the risk of experiencing a new event is softened by higher *marks*, or (ii) in case of the same cumulative marks, it is softened by an higher number of events *enum*. Furthermore, males [HR > 1] were more likely to buy medications or being re-hospitalized than females, except for AA purchases [HR < 1], and elder patients were more likely to be re-hospitalized than younger ones [HR > 1].

**Table 3.1.** Selected features, Hazard Ratios (HRs) and corresponding 95% Confidence Intervals (CIs) of the AG models for recurrent events for the stochastic processes describing the purchase of ACE inhibitors, BB agents, AA agents and the HF hospitalizations.

Process $m$	Selected features	HR	[2.5; 97.5]% CI
<i>ACE</i>	<i>gender (Male)</i>	1.0586	[1.0309; 1.0871]
	$\log(enum + 1)$	4.5271	[4.1674; 4.9178]
	$\log(marks + 1)$	1.1026	[1.0862; 1.1192]
	$\log(enum + 1) \times \log(marks + 1)$	0.9148	[0.9033; 0.9265]
<i>BB</i>	<i>gender (Male)</i>	1.0612	[1.0333; 1.0898]
	$\log(enum + 1)$	5.4270	[5.1195; 5.7529]
	$\log(marks + 1)$	1.1404	[1.1206; 1.1606]
	$\log(enum + 1) \times \log(marks + 1)$	0.8332	[0.8213; 0.8454]
<i>AA</i>	<i>gender (Male)</i>	0.9435	[0.9073; 0.9811]
	$\log(enum + 1)$	9.8781	[8.6116; 11.3310]
	$\log(marks + 1)$	1.2023	[1.1722; 1.2332]
	$\log(enum + 1) \times \log(marks + 1)$	0.7780	[0.7561; 0.8005]
<i>HF hosp</i>	<i>age</i>	0.9957	[0.9934; 0.9979]
	<i>gender (Male)</i>	1.1510	[1.0854; 1.2207]
	<i>enum</i>	1.4319	[1.3809; 1.4848]
	<i>marks</i>	1.0083	[1.0051; 1.0116]
	$enum \times marks$	0.9976	[0.9968; 0.9985]

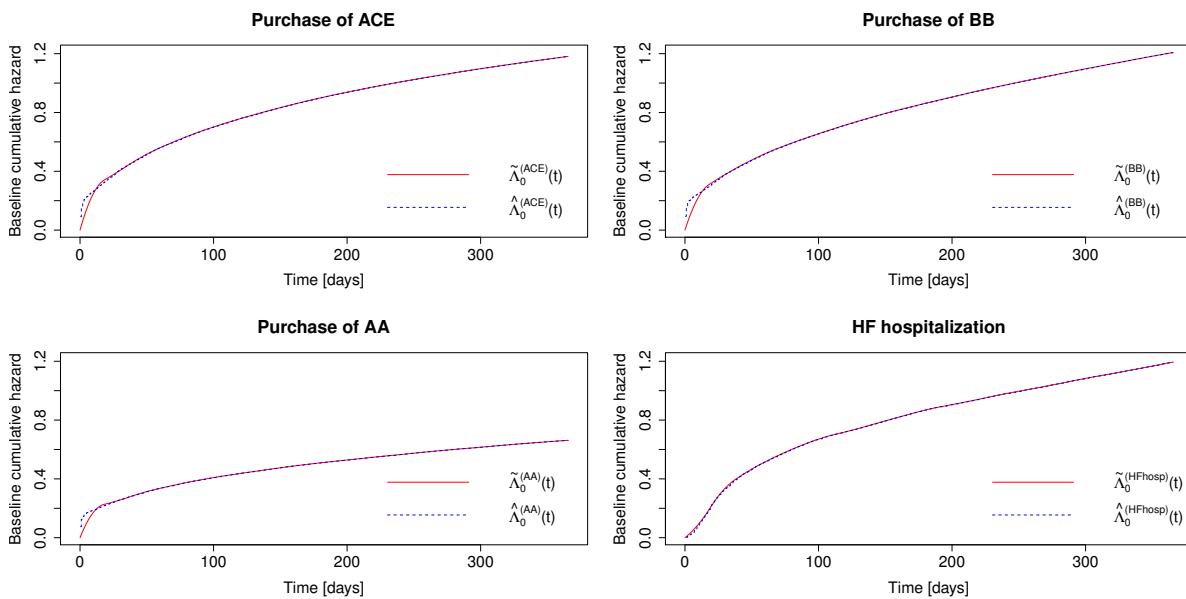
### Fit and smooth cumulative baseline hazard

Once we estimated the coefficients  $\hat{\beta}^{(m)}$  and  $\hat{\gamma}^{(m)}$  of each AG model for recurrent events of type  $m$ , we computed the estimated cumulative baseline hazards  $\hat{\Lambda}_0^{(m)}(t)$  using the Breslow estimator. We smoothed them through the use of constrained B-splines [73] with increasing monotone constraints and no roughness penalties. In particular, we used 20 knots for the B-spline basis and we assumed that they took value 0 at time  $t = 0$ .

Figure 3.4 shows both the estimates obtained with the Breslow estimator  $\hat{\Lambda}_0^{(m)}(t)$  (dashed blue lines) and the corresponding smoothed estimates  $\tilde{\Lambda}_0^{(m)}(t)$  (solid red lines) for the four stochastic processes describing ACE purchase (top-left panel), BB purchase (top-right panel), AA purchase (bottom-left panel) and HF hospitalization (bottom-right panel). We observed that  $\forall m \in \mathcal{M}$  we obtained monotonically increasing estimates  $\tilde{\Lambda}_0^{(m)}(t)$  of the cumulative baseline hazards with  $\tilde{\Lambda}_0^{(m)}(0) = 0$ .

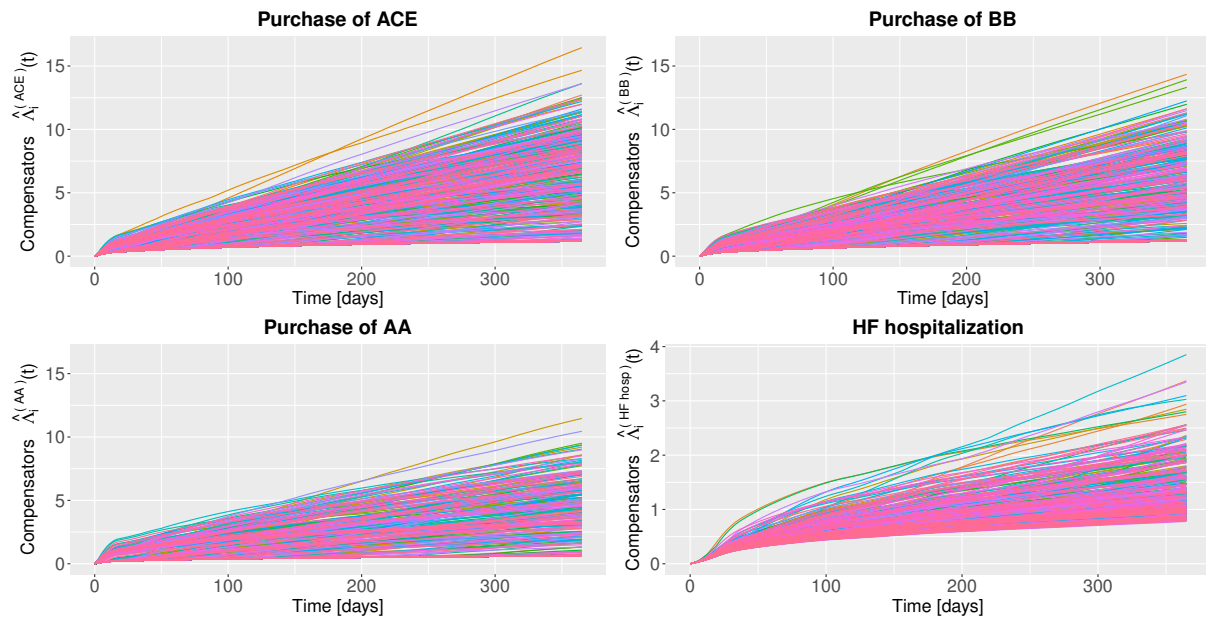
### Reconstruct compensators

At this point, we could reconstruct the trajectories of the compensators  $\hat{\Lambda}_i^{(m)}(t)$  of the four considered stochastic processes for all the patients, exploiting Equation (3.5). The trajectories of compensators  $\hat{\Lambda}_i^{(m)}(t)$  constitute our functional data. Figure 3.5 shows the compensators of the stochastic processes describing ACE purchase (top-left panel), BB purchase (top-right panel), AA purchase (bottom-left panel) and HF hospitalization (bottom-right panel) of the same sample of 500 HF patients mentioned above. We observed that the trajectories  $\hat{\Lambda}_i^{(m)}(t)$  are monotonically non-decreasing and take value 0 at time  $t = 0$ , as did the smoothed baseline cumulative hazards  $\tilde{\Lambda}_0^{(m)}(t)$ . For each patient



**Figure 3.4.** Cumulative baseline hazards of the Cox models for recurrent events describing the stochastic processes of purchases of ACE inhibitors (top-left panel), BB agents (top-right panel), AA (bottom-left panel) and of HF hospitalizations (bottom-right panel), fitted with the Breslow estimator  $\hat{\Lambda}_0^{(m)}(t)$  (dashed blue lines) and smoothed  $\tilde{\Lambda}_0^{(m)}(t)$  according to the procedure described in [21] (solid red lines).

### 3. Functional modelling of recurrent events on time-to-event processes

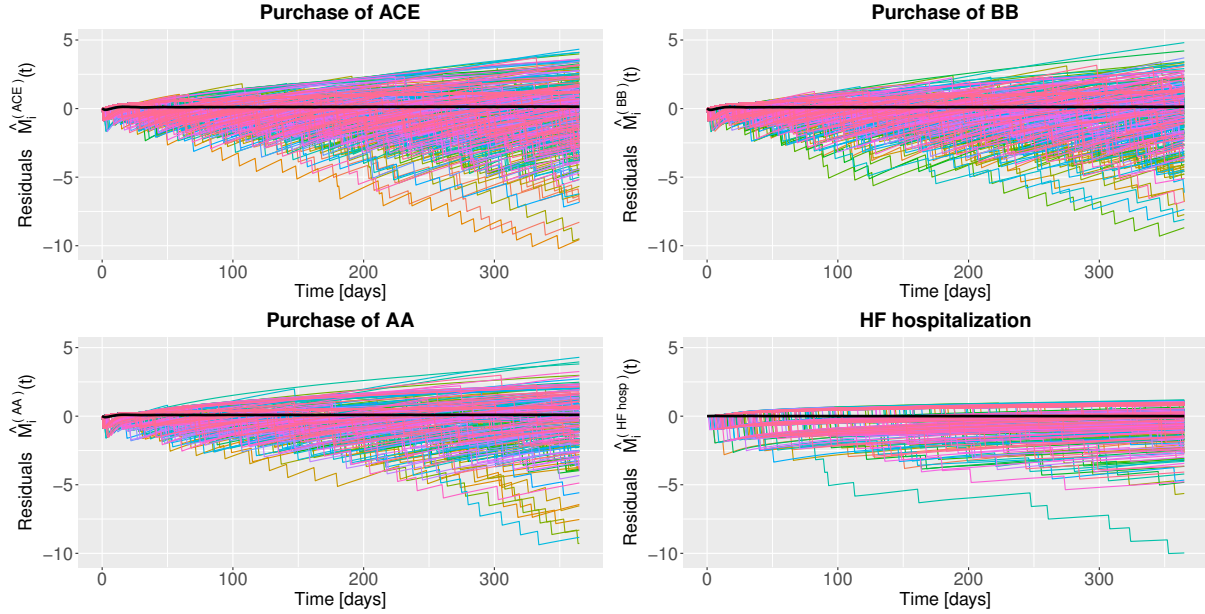


**Figure 3.5.** Compensators  $\hat{\Lambda}_i^{(m)}(t)$  of the marked counting processes of purchases of ACE inhibitors (top-left panel), BB agents (top-right panel), AA (bottom-left panel) and of HF hospitalizations (bottom-right panel) fitted using Equation (3.5) for a sample of 500 HF patients belonging to the administrative database. Each line is related to a different patient. Note that in HF hospitalizations the ordinate axis range is smaller than the other ones due to less number of hospitalization events with respect to drugs purchases.

$i$ , the compensator curve  $\hat{\Lambda}_i^{(m)}(t)$  represents the expected number of events by time  $t$  given the covariates, i.e., the dynamic evolution of the events risk. This means that for a patient with a higher curve the cumulative risk of a new event (i.e., drug purchases or re-hospitalizations) is higher over time compared to a patient with a less steep curve. The large variability of the compensators across different patients reflects the variability of the realizations of their recurrent events times and marks.

Finally, we had to check for adequate fitting of the procedure. In order to do so, for each process of interest, we plotted the residuals evaluated in the whole observation period and we checked graphically that their means  $\bar{M}^{(m)}(t)$  were approximately equal to 0. Figure 3.6 show the fitted residuals  $\hat{M}_i^{(m)}(t)$  for each process for the sample of the 500 patients mentioned above (*ACE*: top-left; *BB*: top-right; *AA*: bottom-left; *HF hosp*: bottom-right). The black line in each panel corresponds to the temporal average residual curve  $\bar{M}^{(m)}(t)$ , computed using all the  $n = 4,541$  patients. From the figure we observed that the time-varying means were approximately constant lines equal to zero for all the considered processes. Hence, we might conclude that we succeeded in fitting the compensators of the stochastic processes.

For each patient  $i \in \{1, \dots, 4,541\}$ , we ended up with a four-variate time-varying data given by the compensator trajectories  $\left\{ \hat{\Lambda}_i^{(m)}(t) \right\}_{h \in \mathcal{M}}$  with  $\mathcal{M} = \{ACE, BB, AA, HF hosp\}$ , which could be thought of as positive non-decreasing  $L^2$ -functions over the temporal domain  $[T_0; T_0^*]$ .



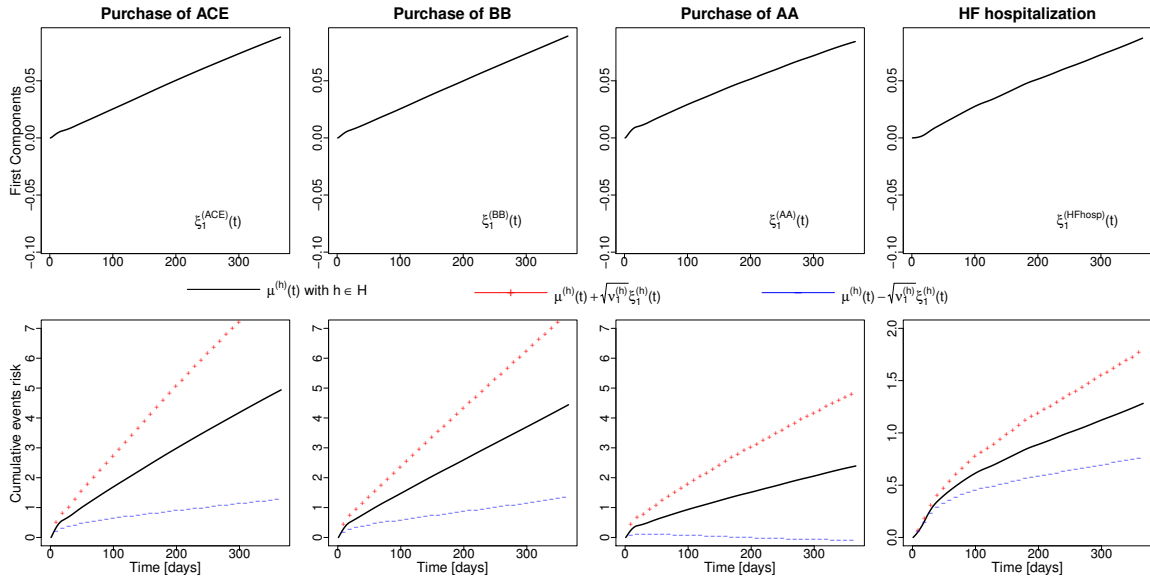
**Figure 3.6.** Residuals  $\hat{M}_i^{(m)}(t)$  of the compensators of the stochastic process describing the purchase of ACE inhibitors (top-left panel), BB agents (top-right panel), AA (bottom-left panel) and of HF hospitalizations (bottom-right panel) for a sample of 500 HF patients belonging to the administrative database, computed according to Equation (3.6). Each line is related to a different patient. Solid black lines represent the temporal average residual curve  $\bar{M}^{(m)}(t)$  computed using all the  $n = 4,541$  patients.

### 3.3.3. Step 3: Summarize compensators through Functional Principal Component Analysis

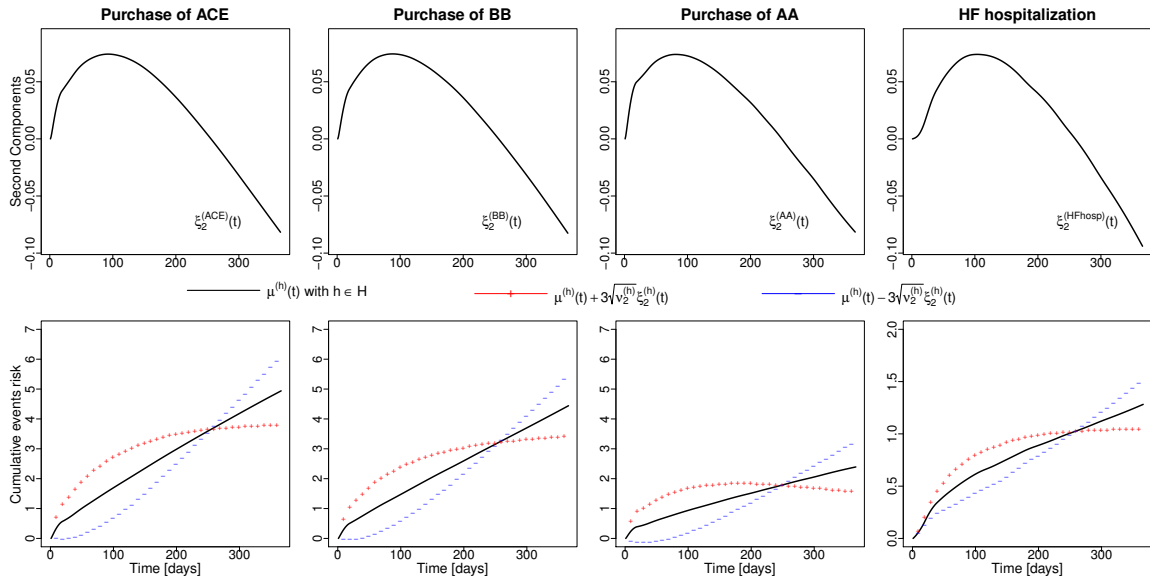
Once we computed the functional trajectories of the compensators  $\hat{\Lambda}_i^{(m)}(t)$ , we performed Functional Principal Component Analysis (FPCA) [162] in order to summarise information emerging from the time-varying compensators to a finite set of covariates while losing a minimum part of the information. Although it was no longer guaranteed that the functions reconstructed through FPCA were positive and non-decreasing, for each process  $m$  we observed that two Principal Components (PCs) were enough to have a  $L^2$ -reconstruction error lower than 1%.

Figure 3.7 and Figure 3.8 show results of FPCA on functional compensators and are related to first and second PCs, respectively. In both figures, each column is related to a different type of process (*ACE*: first column; *BB*: second column; *AA*: third column; *HF hosp*: fourth column). Top panels show that first and second PCs, i.e.,  $\xi_1^{(m)}(t)$  and  $\xi_2^{(m)}(t)$ , across the four processes types  $m \in \{ACE, BB, AA, HF hosp\}$  have similar shapes. Bottom panels report the plots of compensators as perturbation of the mean [162]. In particular, the black lines constitute the average compensators curves  $\mu^{(m)}(t) = \frac{1}{n} \sum_{i=1}^n \hat{\Lambda}_i^{(m)}(t)$ , also denoting subjects with null FPC scores. Red plus and blue minus curves represent the perturbations  $\mu^{(m)}(t) \pm c_k \sqrt{\nu_k^{(m)}} \xi_k^{(m)}(t)$  (red '+' and blue '-' respectively), where  $\nu_k^{(m)}$  is the eigenvalues related to the  $k$ -th component and  $c_k$  are constants chosen in order to let the values lie within one ( $c_k = 1$ ) or three ( $c_k = 3$ ) standard deviations (i.e., square roots of  $\nu_k^{(m)}$ ).

### 3. Functional modelling of recurrent events on time-to-event processes



**Figure 3.7.** First functional Principal Components (PCs) of the compensators of the stochastic processes describing the purchase of ACE (first column), BB (second column), AA (third column) and HF hospitalization (fourth column). Upper panels show the first PCs  $\xi_1^{(m)}(t)$  with  $m \in \mathcal{M} = \{ACE, BB, AA, HFhosp\}$ . Lower panels report the average compensators curves  $\mu^{(m)}(t) = \frac{1}{n} \sum_{i=1}^n \hat{\Lambda}_i^{(m)}(t)$  (black lines) and  $\mu^{(m)}(t) \pm \sqrt{\nu_1^{(m)}} \xi_1^{(m)}(t)$  (red '+' and blue '-' respectively) where  $\nu_1^{(m)}$  are the eigenvalues related to the first components. Note that in HF hospitalizations the ordinate axis range is smaller than the other ones due to less number of hospitalization events with respect to drugs purchases.



**Figure 3.8.** Second functional Principal Components (PCs) of the compensators of the stochastic processes describing the purchase of ACE (first column), BB (second column), AA (third column) and HF hospitalization (fourth column). Upper panels show the second PCs  $\xi_2^{(m)}(t)$  with  $m \in \mathcal{M} = \{ACE, BB, AA, HFhosp\}$ . Lower panels report the average compensators curves  $\mu^{(m)}(t) = \frac{1}{n} \sum_{i=1}^n \hat{\Lambda}_i^{(m)}(t)$  (black lines) and  $\mu^{(m)}(t) \pm 3\sqrt{\nu_2^{(m)}} \xi_2^{(m)}(t)$  (red '+' and blue '-' respectively) where  $\nu_2^{(m)}$  are the eigenvalues related to the second components. Note that in HF hospitalizations the ordinate axis range is smaller than the other ones due to less number of hospitalization events with respect to drugs purchases.

In Figure 3.7 we observe that the first components  $\xi_1^{(m)}(t)$  distinguish patients with different events risks. In particular, positive scores related to the first PC (red plus curve) reflect higher curves with respect to negative ones (blue minus curve), indicating that a patient with a high score on the first component is likely to experience more events than a patient with a low score. Figure 3.8 shows that the second components  $\xi_2^{(m)}(t)$  distinguish patients with different time distribution of the events. In particular, a patient with a high score (red plus curve) on the second PC is likely to experience more events in the first months of the *observation period* and less events in the last months than a patient with a low score (blue minus curve), indicating different events timing.

### 3.3.4. Step 4: Predictive functional Cox model for overall survival

At this point we wanted to quantify the association between time-varying processes and patients' overall survival through a Multivariate Functional Linear Cox Regression Model (MFLCRM) in Equation (3.10). First, we applied 10-fold cross validation to select the best set of covariates among possible combinations of patients' baseline characteristics *age*, *gender* and truncation parameters  $K_m$  of FPCA with  $m \in \{ACE, BB, AA, HF hosp\}$ , according to the highest median Concordance Index [151]. The selected MFLCRM, given by

$$h_i \left( t | \boldsymbol{\omega}_i, \left\{ \Lambda_i^{(m)} \right\}_{m \in \mathcal{M}} \right) = h_0^*(t) \exp \left\{ \theta_1 age_i + \theta_2 gender_i + \right. \\ \left. \alpha_1^{(ACE)} f_{i1}^{(ACE)} + \alpha_1^{(BB)} f_{i1}^{(BB)} + \alpha_1^{(AA)} f_{i1}^{(AA)} + \right. \\ \left. \alpha_1^{(HF hosp)} f_{i1}^{(HF hosp)} + \alpha_2^{(HF hosp)} f_{i2}^{(HF hosp)} \right\}, \quad (3.11)$$

was then fitted on the whole data to quantify the association between functional compensators and overall survival.

Table 3.2 reports the summary of fitted model (3.11). All the covariates resulted statistically significant at confidence level 5%, except for  $f_1^{(AA)}$ . Elder patients coherently have a higher risk of dying [HR = 1.067] and being a male corresponds to 1.25-times faster experience of the event. The HR relative to the scores of the first PCs for *ACE* and *BB* processes, i.e.,  $f_1^{(ACE)}$  and  $f_1^{(BB)}$ , are lower than 1, indicating that a proper *ACE/BB* drug intake is correlated to longer life expectancy. On the contrary, the HR related to  $f_1^{(HF hosp)}$  is greater than 1, standing as a proxy of patients' critical conditions: patients experiencing many hospitalizations in the past present a higher risk of dying. Interestingly, even if the second PC of compensators related to *HF hosp* process concerned only the 2% of the total explained variance of the original data,  $f_2^{(HF hosp)}$  is strongly significant with HR = 0.773 < 1 (95% CI = [0.725; 0.825]). This means that patients with many hospitalizations at the beginning of the *observation period* and few hospitalizations in the end have higher survival probability, since they probably correspond to the ones who had already experienced a critical phase of the disease and survived from it.



### 3. Functional modelling of recurrent events on time-to-event processes

**Table 3.2.** Hazard ratios (HRs) along with 95% Confidence Intervals (CIs) of the final multivariate functional linear Cox regression model (MFLCRM) for overall survival fitted on the whole cohort using the covariates selected through 10-fold cross-validation.

Covariates	HR	[2.5; 97.5]% CI	p-value
<i>gender (Male)</i>	1.2540	[1.1080; 1.4194]	< 0.001
<i>age</i>	1.0670	[1.0592; 1.0748]	< 0.001
$f_1^{(ACE)}$	0.9977	[0.9963; 0.9992]	0.003
$f_1^{(BB)}$	0.9964	[0.9945; 0.9982]	< 0.001
$f_1^{(AA)}$	1.0006	[0.9986; 1.0026]	0.550
$f_1^{(HF hosp)}$	1.0157	[1.0049; 1.0266]	0.004
$f_2^{(HF hosp)}$	0.7733	[0.7251; 0.8247]	< 0.001

## 3.4. Final remarks

In this chapter, a novel approach to reconstruct the compensators of suitable marked point processes of interest as time-varying covariates has been proposed. This approach was exploited to enrich information to be included into a survival model. The development of this procedure is due to the need of effectively describing and resuming information from dynamic processes affecting an outcome of interest, with the purpose of obtaining deeper insight on the patient's health status using administrative databases. This methodology extends the one proposed in [21], allowing the counting processes to depend on their marks and moving towards the multivariate setting.

From the study on the administrative database of *Regione Lombardia*, we observed that modelling patient's clinical history in terms of compensators of suitable stochastic processes as time-varying covariates and plug them into a survival model represents an effective, interpretative and forecasting approach for exploring the effects of these processes on patients' survival. The marked point process formulation is a natural way of representing the occurrence of hospitalizations or drugs purchases over time. The use of FPCA allowed to extract additional information contained in the functions, representing a powerful exploratory and modelling technique for highlighting trends and variations in the shape of the processes over time. The introduction of this novel way to account for time-varying variables by means of compensators allowed for modelling self-exciting behaviours, for which the occurrence of events in the past increases the probability of a new event. This enabled us to include a large piece of information contained in the administrative data to describe the patient's clinical history. Furthermore, our approach was able to take into account the fact that HF patients usually consume different types of drugs at the same time, representing a novelty for clinical and pharmacological research in the direction of properly treating multimorbidity patients and polypharmacy. To the best of our knowledge, our approach represents the first attempt in literature of merging potential of FDA, recurrent events theory and survival analysis.

Thanks to its flexibility, the proposed methodology could be extended and generalized to many different settings, adapting the procedure to the different biological and clinical aspects of the specific application. In particular, alternative ways to get the trajectories

related to the  $L^2$  functional compensators could be considered. The AG model for recurrent events in Equation (3.3) represents only one of the possible approaches to express the conditional intensity function. Alternative methods or distributions for the marks could be considered according to many factors, among others number of events, relationship between subsequent events and intrinsic characteristics of the processes. For example, our case study was also analysed considering a shared gamma-frailty model [175], in which the intensity function in Equation (3.3) was assumed to partly depend on an unobservable random variable that acted multiplicatively on it. In that case, the compensator trajectories were expressed as functions of estimated coefficients, smoothed cumulative baseline hazard and estimated frailties. Obtained results were comparable to the ones shown in the paper in terms of both estimated effects on patients' overall survival and clinical implications. In case of a limited number of events, stratified Cox models for recurrent events, such as the Prentice-Williams-Peterson [156] or the Wei-Lin-Weissfeld [213] model, could be used modifying Equation (3.4) in order to consider the proper strata of the cumulative baseline hazards. As a further alternative, in case of multiple events with cyclical occurrence, the best choice would be to account for seasonality in the model through cyclic functions, such as in the rate model with multiple event types by [196]. In that case, the  $L^2$  functions could be obtained by smoothing the cumulative rate functions. Therefore, thanks to its adaptability, the presented methodology can be generalized and applied to the study of many different pathologies characterized by complex data sources.

Some limitations of the present study have to be mentioned. Firstly, the use of a pre-defined *observation period* could lead to survival bias due to cohort selection. Indeed, it is necessary that patients survived for a period at least equal to the length of the period used to compute the functional compensators trajectories. This could imply a survival bias in case of the exclusion of too many early dying patients. This is softened if low-rate short-term mortality diseases are considered. In the present work, the final choice for a pre-defined *observation period* of 1 year after the index hospitalization was made under clinical indication, once performed a sensitivity analysis to evaluate the robustness of our method using two different clinically acceptable periods of 6 months and 1 year whose results led to common conclusions. From a modelling point of view, the assumption of independence between jump times and marks in Equation (3.2) could in general be relaxed, but this could lead to several issues [132]. In fact, considering re-hospitalization process, it is difficult to conjecture a mathematical relation of length of stay in hospital with time of hospitalization. The same is valid for drug purchases. Moreover, there could be computational limitations in terms of modelling a temporary dependence. Since dependence is harder to be dealt with due both to computational and modelling issues, we limited our analysis to the independence case, which was considered a clinically acceptable assumption. The development of proper statistical tools to test this hypothesis can be of great help for our topic, since existing techniques for testing independence are rather complex to apply and customize to the current context. Furthermore, FPCA was performed in  $L^2 [T_0; T_0^*]$  and not in the subspace of positive non-decreasing  $L^2$ -functions. In this way, we obtained a good reconstruction of compensators approximated using PCs but it was no longer guaranteed that these functions were positive and non-decreasing.

Other limitations are mainly due to the use of secondary databases in the real case-study, as in the previous chapters. First, not being able to ascertain whether the patient was currently consuming the dispensed drug remains the major limitation of using drug purchases as a proxy for drug intake, which is the only possible way through administrative data. Second, the use of theoretical Defined Daily Dose (DDD) instead of Prescribed Daily Doses (PDD) could reflect a bias in the computation of coverage days, i.e., of jump marks, if the underlying PDD/DDD ratio is different from 1 [187, 220]. In future analysis, it could be interesting to explore, whenever the linkage is possible, databases with information about dosages prescribed by doctors, in order to obtain a more realistic analysis of coverage periods. Since administrative data are collected with no clinical question in mind and mainly for managerial and economic purposes [89], the validity of using these kind of data is critically dependent on the reliability of the data [115, 180, 90]. Nevertheless, in the last decade significant improvements have been gained through administrative data sources, and their use in clinical biostatistics has become an accepted practice, representing a great challenge for statistics and related modelling [90].

Despite the aforementioned limitations, our approach opens doors for many further developments, both in the fields of statistical methods and clinical research. The proposed predictive models could be enriched by considering other relevant clinical information as covariates, and enlarging the cohort of patients. For example, it could be of clinical interest to further extend the study of polypharmacy by considering also drug-drug interaction terms, which could be included in the model through compensator-compensator interaction terms. However, a compensator-compensator interaction term involves the modelling of bivariate (or more in general multivariate) marked point processes, which represents a non-trivial task beyond the scope of the present work.

In summary, the presented methodology, involving database integration, marked point process modelling of critical events and FDA techniques, enabled a manageable and relatively simple analysis of the results, describing complex dynamics in an easily interpretable form. Both parts of the procedure represent flexible approaches that can be used to quantify personal behaviours and to investigate their effect of on survival. On one hand, the developed marked point processes formulation could be applied in many different clinical contexts characterized by recurrent occasions. On the other, the use of FPCA to extract additional information contained in the functions and to include them into a MFCLRM can be easily applied to all settings where the time-varying characteristics of interest are adequately reconstructed by FDA, as we will see in Chapter 4 for the case of biomarkers and chemotherapy dose in osteosarcoma patients. Its possible generalization to many different contexts, combined with cooperation with medical staff, could therefore lead to improvements in the definition of useful tools for health care assessment and treatment planning.

## A. Appendix to Chapter 3

### A.1. Data Preparation

Once selected the cohort of patients being part of the analysis and identified the events related to each patient's clinical history (Section 3.3 – Step 1 of the procedure), we had to reformat the administrative data building four different datasets, one for each process  $m \in \mathcal{M} = \{m : ACE, BB, AA, HF hosp\}$ , in the form required by `coxph` function for recurrent events of `survival` R package by [201]. Table 3.3 shows an example of reformatted dataset related to ACE purchases process for a hypothetical patient with four *ACE* events during the observation period. In the Table, *start* indicates the time of the patient's previous event (equal to 0 for the index date), *stop* is the time of the current event (equal to 365.5 if it is the censoring event), *status* is the event indicator (0 if censored, 1 otherwise), *enum* is the number of events related to process  $m$  occurred in the past and *marks* is the sum of the corresponding marks. In particular, the choice to consider the time limit at 365.5 was made in order to not have events at censoring time  $t = 365$ . Moreover, it could also happen that a patient  $i$  experienced the first event of type  $m$  during the index day. In that specific case, we considered jump time equal to 0.5, i.e.,  $t_{i,1}^{(m)} = 0.5$ , in order to not have events at time  $t = 0$ . Hence, for each process  $m$  we ended up with a long-format dataset with multiple rows for each patient (specifically the number of patient's events of type  $m$  during the observation period plus one). In particular, in the first row of each patient *enum* and *marks* are always 0 and in the last one *status* is always equal to 0.

**Table 3.3.** Example of reformatted dataset related to ACE purchases process for a hypothetical patient with four ACE events during the 1-year *observation period*.

<i>ID</i>	<i>start</i>	<i>stop</i>	<i>status</i>	<i>gender</i>	<i>age</i>	<i>enum</i>	<i>marks</i>
<i>id</i>	0	0.5	1	<i>Female</i>	87	0	0
<i>id</i>	0.5	83	1	<i>Female</i>	87	1	56
<i>id</i>	83	91	1	<i>Female</i>	87	2	70
<i>id</i>	91	215	1	<i>Female</i>	87	3	98
<i>id</i>	215	365.5	0	<i>Female</i>	87	4	112

### A.2. Mean Absolute Martingale Residual

Given two or more Andersen-Gill (AG) models for recurrent events in Equation (3.3) fitted using different sets of covariates, we need a metric to evaluate the goodness of fit of each model and select the best set of features. Since we are dealing with stochastic processes and recurrent events, we cannot rely on standard regression metrics, like mean squared error. A possible way is given by functions of the residuals in Equation (3.6): smaller residuals correspond to a greater predictive power of the model. Therefore, to compare models fitted with different features, for each process  $m$  we would like to use the

Mean Absolute Martingale Residual (MAMR):

$$MAMR^{(m)} = \sum_{i=1}^n \frac{\int_0^T |\hat{M}_i^{(m)}(s)| ds}{T} \quad (3.12)$$

where  $T$  represents the length of the observation period. Using this indicator, the smaller the MAMR the better the model.

To correctly compute the MAMR, we should first compute the compensators using Equation (3.5) and then evaluate the residuals on a grid of points. Since we want to use this quantity only to rank models fitted with different sets of predictors, to avoid high computational costs we decided to rely on the following estimate:

$$\widehat{MAMR}^{(m)} = \frac{1}{\sum_{i=1}^n n_i^{(m)}} \sum_{i=1}^n \sum_{j=1}^{n_i^{(m)}} |\hat{M}_i^{(m)}(t_{i,j}^{(m)})| \quad (3.13)$$

where  $i$  and  $m$  are respectively the patient and event indices,  $\hat{M}_i^{(m)}(t)$  is the residual obtained by fitting the compensator without smoothing the baseline hazard, i.e., using  $\hat{\Lambda}_0^{(m)}(t)$  instead of  $\tilde{\Lambda}_0^{(m)}(t)$  in Equation (3.5),  $n_i^{(m)}$  is the total number of events of type  $m$  experienced by the  $i$ -th patient and  $t_{i,j}^{(m)}$  is the time instant in which patient  $i$  experienced the  $j$ -th event of type  $m$ .

This estimate is not accurate since the residuals are evaluated only when events happen (rather than on the continuous interval corresponding to the one year observation period) and because the estimate is done by reconstructing the compensators without the smoothing of the baseline hazard. However, it allows to rank models while limiting computational needs.

## **PART II**

---

# **Chemotherapy in Osteosarcoma**



## CHAPTER 4

---

# Modelling time-varying covariates effect on survival via Functional Data Analysis

This chapter has been published in *Statistical Methods & Applications* 2022 as M. Spreafico, F. Ieva and M. Fiocco “Modelling time-varying covariates effect on survival via Functional Data Analysis: application to the MRC BO06 trial in osteosarcoma” [192].

Osteosarcoma is a malignant bone tumour mainly affecting children and young adults with an annual incidence of 3-4 patients per million [185]. Multidisciplinary management including neoadjuvant and adjuvant chemotherapy with aggressive surgical resection [166] or intensified chemotherapy has improved clinical outcomes although the overall 5-year survival rate has remained unchanged in the last 40 years at 60-70% [15]. Therefore, it is extremely important to provide an effective tool to evaluate the prognosis for osteosarcoma and to guide the diagnosis.

Time-varying (or time-dependent) covariates are often of interest in clinical and epidemiological research: patients are followed during the study and subject-specific measurements are recorded at each visit. Well-known examples include biomarkers which change during follow-up or cumulative exposure to medications [18], such as chemotherapy. Depending on patients' treatment history or development of toxicity, biomarkers values change and chemotherapy treatment is modified by delaying a course or reducing the dose intensity. To study the association between time-varying responses with time-to-event outcome (e.g., death) is a challenging task which could offer new insights into the direction of personalised treatment.

In osteosarcoma treatment, patients usually undergo assessment of haematologic and serum biochemical parameters [119], such as white blood cell (WBC) counts and alkaline phosphatase (ALP). The role of ALP as tumour marker for osteosarcoma has not been established, although several studies suggested that high ALP level is associated with poor overall or event-free survival and presence of metastasis [165, 71]. Chemotherapy is usually modelled by different allocated regimens, i.e., by Intention-To-Treat (ITT) analysis [70]. ITT ignores anything that happens after randomization, such as protocol deviations or changes in drug intake over time, i.e., delays or dose reduction [110]. Lancia *et al.* (2019) [111] showed that there is mismatch between target and achieved dose of chemotherapy and the impact of dosis on patients' survival is still unclear. For these reasons, in this



chapter a novel method to study received chemotherapy dose and biomarkers as time-varying variables is proposed. This approach has never been applied to osteosarcoma treatment and provides new insight in understanding the effect of chemotherapy dosis intensity on sarcoma in childhood cancer. Moreover, as will be clear in the following, the application is inspiring from a statistical modelling perspective.

Models for time-to-event data which are able to deal with the dynamic nature of time-varying responses during follow-up are not well developed. One approach for using time-varying covariate data is the Time-Varying covariate Cox Model (TVCM) [202, 97], that is an extension of the Cox proportional hazard model [46] accounting for covariates that can change value during follow-up. Since time-dependent observations are only available at the time of measurements, TVCM uses the last-observation-carried-forward (LOCF) approach [206], which leads to the pitfall of introducing bias due to the continuous nature of the process underlying the data, and fails to account for possible measurement errors [16]. Joint models address these issues by modelling simultaneously longitudinal and time-to-event data using shared random effects [76, 206, 38, 49, 167, 168, 65, 157, 81, 82]. As seen in Chapter 2, they are parametric models that allow for the inference on the association between the hazards characterizing the event outcome and the longitudinal processes. However, they require additional strong assumptions over TVCM that need to be carefully validated to avoid biased estimates [16]. Their benefits are hence strictly linked to the correct specification of longitudinal trajectories and baseline hazard function. In addition, inference computations could become prohibitive, especially for approaches developed in a Bayesian framework.

During the past two decades, Functional Data Analysis (FDA) has been increasingly used to analyse, model and predict dynamic processes [163, 162, 144, 223, 56, 128, 207, 92, 91, 134, 189]. The idea behind FDA and functional models is to express discrete observations arising from time series, i.e., longitudinal time-varying observations, in the form of functions [163, 162]. Functional representation incorporates trends and variations in the evolution of the process over time [207]. Since functional data are infinite-dimensional covariates, some dimensionality reduction methods are needed to summarize and select a finite dimensional set of elements representing the most important features of each covariate. This information can then be included into time-to-event models. To model the relationship between survival outcomes and a set of finite and infinite dimensional predictors Functional Linear Cox Regression Models (FLCRM) have been recently proposed [62, 116, 159, 109, 121]. In case of an infinite dimensional process, Kong *et al.* (2018) [109] characterized the joint effects of both functional and scalar predictors on time-to-event outcome employing Functional Principal Component Analysis (FPCA). FPCA is one of the most popular dimensionality reduction method in FDA and it is used to summarise each function to a finite set of covariates through FPC scores, while losing a minimum part of the information. An extended version of the FLCRM by [109] to the case of multiple functional predictors – named Multivariate FLCRM (MFLCRM) – was introduced in the previous chapter to model recurrent events effect on long-term survival [189]. However, since the main focus of Chapter 3 was to develop a methodology for effectively modelling time-varying recurrent events in terms of the functional compensators underlying

the processes of interest, we have neither compared MFLCRM with other survival models, nor considered its predictive performances over time. In case of multiple longitudinal processes, Li and Lou (2019) [121] exploited the multivariate FPCA approach by [72] to extract the FPC scores from the multiple longitudinal trajectories in order to make personalized dynamic predictions. However, the authors did not focus on the smoothing and functional representation aspects of the processes realized by the observed longitudinal data, on the clinical interpretation of the FPC scores and on their association with overall survival. Since it is often the changing patterns of the functional trajectories rather than the actual values that affects patients' survival, FDA provides a novel modelling and prediction approach, with a great potential for many applications in public health and biomedicine [207].

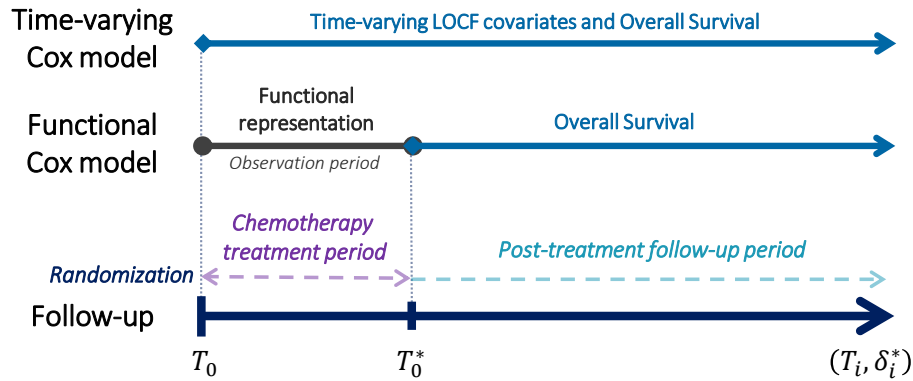
Motivated by a clinical question concerning the effect of biomarkers and dose variations during treatment on survival for osteosarcoma patients, in this chapter a FDA-based approach, named *Functional covariate Cox Model* (FunCM), is proposed and compared to a standard TVCM. In FunCM, FDA techniques are first exploited to represent time-varying processes and their derivatives over time in terms of functional data. Unlike joint models, FDA approach does not make assumptions on the distributions of longitudinal processes being computationally advantageous [121]. Then, additional information contained into the evolution of the functions over time are included into MFLCRMs for overall survival through FPCA. A cross-validation method is implemented to compare MFLCRMs and standard TVCM in terms of their predictive performances at different time horizons. Three novelties of this work are listed here: (i) application of advanced statistical techniques to deal with time-varying covariates in the field of osteosarcoma treatment; (ii) reconstruction of the functional representations for biomarkers and chemotherapy dose values, and their rates of change, to retrieve information on the progression of processes over time; (iii) comparison between TVCM and FunCM in terms of both clinical interpretability and time-dependent predictive performances. This novel approach provides more information about the effect of individualized treatment adaption on survival for osteosarcoma patients.

The rest of this chapter is organized as follows. In Section 4.1 TVCM and FunCM to represent time-varying covariates by means of FDA and to include them into survival models are discussed. MRC BO06/EORTC 80931 Randomized Controlled Trial [119] and longitudinal representations of time-varying covariates are described in Section 4.2. Results are presented in Section 4.3. Section 4.4 ends with a discussion of strengths and limitations of the current approach, identifying some developments for future research.

## 4.1. Statistical Methodologies

### 4.1.1. Time-varying covariates and survival frameworks

A time-varying (or time-dependent) process is a covariate whose value can change over the duration of follow-up (e.g., time-varying biomarkers, current use of medication, and



**Figure 4.1.** Follow-up periods. Time-varying (LOCF/functional) representation and Overall Survival (OS) for Time-Varying covariate Cox Model (TVCM) and Functional covariate Cox Model (FunCM).  $T_0$  is the time of randomization.  $T_0^* = T_0 + 180$  days is the end of the 6-months chemotherapy treatment period. LOCF = last-observation-carried-forward.

cumulative dose of drugs). In this study, the main interest is in analysing the association between patient's survival and variations during treatment of his/her multiple time-varying characteristics. The focus is hence on patients who had completed the entire chemotherapy treatment protocol in a pre-defined and clinically acceptable timing period.

Follow-up starts from date of randomization  $T_0$  and is divided into a pre-defined 6-months *chemotherapy treatment period*  $[T_0; T_0^*]$  – also called *observation period* – considered for chemotherapy treatment completion, and a *post-treatment follow-up period* from  $T_0^*$  onwards (see Figure 4.1).

Under the TVCM framework, the Overall Survival (OS) is measured from randomization ( $T_0$ ) to the date of death or last follow-up date, and the time-varying covariates can be defined over the entire follow-up period. Let  $\mathcal{M}$  be a set of time-varying processes. Let  $\mathbf{z}_i^{(m)} = \{z_{il}^{(m)} = z_i^{(m)}(t_{il}), l = 1, \dots, n_i^{(m)}\}$  be the vector of longitudinal values related time-varying process  $m \in \mathcal{M}$  for each patient  $i$ , where  $t_{il}$  is the time of the  $l$ -th measurement,  $z_i^{(m)}(t_{il})$  is the value of the process at time  $t_{il}$  and  $n_i^{(m)}$  is the number of different measurements.

Under the FunCM framework, the *observation period*  $[T_0; T_0^*]$  is used to reconstruct the functional representations of time-varying covariates. OS is then measured from the end of the *observation period* ( $T_0^*$ ) to the date of death or last follow-up date. Only patients still alive at  $T_0^*$  are included in the study cohort. To reconstruct the functional covariates, only measurements registered during the *observation period* (i.e., up to  $T_0^*$ ) are considered, namely vector  $\bar{\mathbf{z}}_i^{(m)} = \{z_{il}^{(m)} = z_i^{(m)}(t_{il}), l = 1, \dots, \bar{n}_i^{(m)}\} \subseteq \mathbf{z}_i^{(m)}$ , where  $\bar{n}_i^{(m)}$  denotes the index of last measurement of type  $m$  for patient  $i$  in  $[T_0; T_0^*]$ , with  $\bar{n}_i^{(m)} \leq n_i^{(m)}$  and  $t_{i\bar{n}_i^{(m)}} \leq T_0^* < t_{i\bar{n}_i^{(m)}+1}$ .

In both cases, the observed time-to-death outcome for patient  $i \in \{1, \dots, N\}$  can be denoted as  $(T_i, D_i)$ , where  $T_i = \min(T_i^*, C_i)$  is the observed event time (measured from  $T_0$  or  $T_0^*$  according to the framework),  $T_i^*$  is the true event time,  $C_i$  is the censoring time

and  $D_i = \mathbb{1}(T_i^* \leq C_i)$  is the event indicator, with  $\mathbb{1}(\cdot)$  being the indicator function that takes the value 1 when  $T_i^* \leq C_i$ , and 0 otherwise.

### 4.1.2. Time-Varying covariate Cox Model

Starting from vector of longitudinal values  $\mathbf{z}_i^{(m)}$ , a time-varying covariate  $z_i^{(m)}(t)$  can be defined over the entire follow-up period, according to the LOCF approach [206]:

- when  $z_i^{(m)}(t)$  is not observed at time  $t \in [T_0; t_{in_i^{(m)}}]$ , the most updated value is used:  $z_{il}^{(m)} = z_i^{(m)}(t_{il})$  with  $t_{il} \leq t < t_{il+1}$ ;
- from  $t_{in_i^{(m)}}$  onwards, the last available measurement  $z_i^{(m)}(t_{in_i^{(m)}})$  is considered.

The TVCM is an extension of the proportional hazard model by [46] accounting for covariates that can change value during follow-up [202, 97]. Under TVCM, the proportional hazards model for patient  $i$  has the form

$$h_i(t|\boldsymbol{\omega}_i, \mathbf{z}_i(t)) = h_0(t) \exp \{ \boldsymbol{\theta}^T \boldsymbol{\omega}_i + \boldsymbol{\alpha}^T \mathbf{z}_i(t) \} \quad (4.1)$$

where  $h_0(t)$  is the baseline hazard function,  $\boldsymbol{\omega}_i$  and  $\mathbf{z}_i(t) = (z_i^{(1)}(t), \dots, z_i^{(M)}(t))$  are the vectors of baseline and time-varying covariates with regression parameters  $\boldsymbol{\theta}$  and  $\boldsymbol{\alpha}$ , respectively. Inference for coefficients  $(\boldsymbol{\theta}, \boldsymbol{\alpha})$  is based on maximizing the partial likelihood [97].

TVCM can also be stratified to allow for control by "stratification" of a predictor that does not satisfy the proportional hazard assumption [97]. Under stratified TVCM, the hazard function  $h_{ig}(t|\boldsymbol{\omega}_i, \mathbf{z}_i(t))$  contains also a subscript  $g$  that indicates the  $g$ -th stratum, as well as the baseline hazards  $h_{0g}(t)$ , where the strata are different categories of the stratification variable. Notice that the baseline hazard functions are different in each stratum.

### 4.1.3. Functional covariate Cox Model

The FunCM approach consists of four parts: Steps 1 and 2 are devoted to the reconstruction of functional trajectories; Steps 3 and 4 provide a suitable framework for including these time-varying covariates in a time-to-event model. Specifically, once the data have been pre-processed and longitudinal time-varying characteristics during the *observation period* have been identified (Step 1), the corresponding functional trajectories and their derivatives are reconstructed by applying FDA techniques (Step 2). FPCA is then applied to perform dimensionality reduction and summarise the information from the functional predictors into a finite set of FPC scores (Step 3). Finally, once the best set of covariates and number of principal components have been selected through cross-validation, the MFLCRM is estimated to quantify the association between functional processes and patients' overall survival (Step 4).

### From longitudinal to functional representation

To model the continuous longitudinal vectors  $\bar{z}_i^{(m)}$  defined over  $[T_0; T_0^*]$  as functions  $\tilde{x}_i^{(m)}(t)$ , FDA techniques can be exploited, as discussed by [163, 162]. The observed data  $z_{il}^{(m)}$  are assumed as noisy measurements of the latent processes  $\tilde{X}_i^{(m)}(t)$ , where time  $t \in [T_0; T_0^*]$  and  $i$  is the patient's index.

For each process  $m$ , first the time-scale  $t \in S_m \subseteq [T_0; T_0^*]$  is chosen. There are no restrictions on the choice of unit of measurement for  $t$ , though the specific choice can simplify the computational process. According to the type of observed data (i.e., periodic or open-ended data) and the number of measurements  $\bar{n}_i^{(m)}$ , the basis function system  $\phi_i^{(m)}(t)$  (e.g., polynomials, B-spline, Fourier, wavelets) is selected, with a number of basis less or equal to  $\bar{n}_i^{(m)}$ . Functional data objects are usually expressed by a general functional form as linear combination of the basis functions  $W_i^{(m)}(t) = \phi_i^{(m)}(t)^T \mathbf{c}_i^{(m)}$ , where  $\mathbf{c}_i^{(m)}$  is the vector of coefficients for patient  $i$ . Other functional forms can be used to take into account the nature of the process itself (e.g., positive, increasing, decreasing). For example, for an increasing process, the functional data object can be defined using the monotone functional form  $W_i^{(m)}(t) = \beta_{0i} + \beta_{1i} \int_{t_0}^t \exp[\phi_i^{(m)}(u)^T \mathbf{c}_i^{(m)}] du$  [162]. Once selected the type of basis functions and the functional form, data can be smoothed by regression analysis minimizing the (penalized) sum of squared errors, obtaining functions  $\tilde{x}_i^{(m)}(t) = \hat{W}_i^{(m)}(t)$ .

In the presence of constrain due to the specific application, data can be alternatively smoothed by regression analysis using the transformation  $g(x) = \log \frac{x-L_m}{U_m-x}$ , where  $L_m$  and  $U_m$  denote the lower and upper bounds respectively. For each patient  $i$  the customized functional predictor  $m$  is defined as:

$$\tilde{x}_i^{(m)}(t) = \frac{L_m + U_m \cdot \exp[\hat{W}_i^{(m)}(t)]}{1 + \exp[\hat{W}_i^{(m)}(t)]}. \quad (4.2)$$

Starting from the customized functional datum, the FDA approach also allows to reconstruct its derivative  $d\tilde{x}_i^{(m)}(t)$  as function of the derivatives of the basis functions  $d\phi_i^{(m)}(t)$ . The derivative of the functional process, indicated as  $\tilde{x}_i^{(dm)}(t)$ , represents the rate of change of process values over time. Both functional data  $\tilde{x}_i^{(m)}(t)$  and derivatives  $\tilde{x}_i^{(dm)}(t)$  can be incorporated as functional predictors into a functional Cox regression model for overall survival by taking into account that they are correlated.

### Multivariate functional linear Cox regression model

As shown in Chapter 3, MFLCRM extends the functional Cox regression model by [109] to the case of multiple functional predictors [189]. Let  $\{\tilde{x}_i^{(m)}\}_{m \in \mathcal{M}}$  be the set of realizations of the  $|\mathcal{M}|$ -variate functional predictors for individual  $i$ . MFLCRM includes the multiple functional predictors in the classical Cox model [46] as:

$$h_i \left( t | \boldsymbol{\omega}_i, \left\{ \tilde{x}_i^{(m)} \right\}_{m \in \mathcal{M}} \right) = h_0(t) \exp \left\{ \boldsymbol{\theta}^T \boldsymbol{\omega}_i + \sum_{m \in \mathcal{M}} \int_{S_m} \tilde{x}_i^{(m)}(s) \alpha^{(m)}(s) ds \right\} \quad (4.3)$$

where  $h_0(t)$  is the baseline hazard function,  $\omega_i$  is the vector of scalar (non functional) covariates with regression parameters  $\theta$ .  $\alpha^{(m)}(s)$  are the functional regression parameters. Sets  $S_m \subseteq [T_0; T_0^*]$  are compact sets in  $\mathbb{R}$  and can be different (both in period length and time scale) among between different types  $m$  of functional predictors.

As shown in Section 3.2.2, by applying FPCA, each functional trajectory  $\tilde{x}_i^{(m)}(s)$  can be approximated with a finite sum of  $K_m$  orthonormal basis  $\{\xi_1^{(m)}, \dots, \xi_{K_m}^{(m)}\}$ , i.e., the principal components, and the hazard function in Equation (4.3) becomes:

$$h_i \left( t | \omega_i, \left\{ \tilde{x}_i^{(m)} \right\}_{m \in \mathcal{M}} \right) = h_0^*(t) \exp \left\{ \theta^T \omega_i + \sum_{m \in \mathcal{M}} \sum_{k=1}^{K_m} f_{ik}^{(m)} \alpha_k^{(m)} \right\} \quad (4.4)$$

where  $h_0^*(t) = h_0(t) \exp \left\{ \sum_{m \in \mathcal{M}} \int_{S_m} \mu^{(m)}(s) \alpha^{(m)}(s) ds \right\}$  is the baseline hazard function with functional means  $\mu^{(m)}(s)$ . The FPC score of individual  $i$  related to the  $k$ -th orthonormal base  $\xi_k^{(m)}$ , representing the projection of the  $i$ -th functional compensator related to process  $m$  along the direction of the  $k$ -th principal component, is denoted by  $f_{ik}^{(m)}$ . Parameters  $K_m$  and  $\alpha_k^{(m)}$  are the truncation and the  $k$ -th score regression parameters related to process  $m$ , respectively, with  $\alpha_k^{(m)} = \int_{S_m} \xi_k^{(m)}(s) \alpha^{(m)}(s) ds$ .

Therefore, through FPCA, MFLCRM can be expressed as Cox model with vector of regression coefficients  $\tilde{\theta} = \left[ \theta^T, \left\{ \left( \alpha_1^{(m)}, \dots, \alpha_{K_m}^{(m)} \right) \right\}_{m \in \mathcal{M}} \right]^T$  that can be estimated by maximising the partial likelihood function [46]. For details related to MFLCRM formulation see Section 3.2.2.

To select the truncation parameters  $K_m$ , representing the number of FPCs to be considered, in Chapter 3 we chose the model with the highest Concordance index [151], that is an overall measure of discrimination in survival analysis. In this work, the truncation parameters  $K_m$  are selected in terms of predictive discrimination and calibration performances at different time horizons through the cross-validation procedure introduced in the next Section.

### Selection of truncations parameters

The truncation parameters  $K_m$  in Equation (4.4) can be chosen in different ways: (i) the Proportion of Variance Explained (PVE) [162], (ii) Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC) or (iii) data-adaptive methods, such as cross-validation [223]. In this analysis, a combination of these three methods is used. Let the sets of baseline and functional predictors be fixed. First, different combinations of increasing values of the truncation parameters  $K_m$  for different time-varying processes  $m$  are considered and the best models according to both AIC and BIC criteria are selected. Then, models according to five different thresholds for PVE ( $K_m$  such that  $\text{PVE} \geq 80, 85, 90, 95, 99\%$ ) are identified. Finally all the selected models are compared in terms of their predictive performances at different time horizons through cross-validation to identify the best one.

The predictive performance of the models is assessed in terms of discrimination and calibration. Discrimination is assessed through the time-dependent area under the curve (AUC), estimated through the nonparametric method by [122]. Calibration is assessed by the weighted version of the Brier score under the assumption of independent censoring [66]. Higher AUC and lower Brier score indicate better discrimination and calibration, respectively.

## 4.2. MRC BO06 randomized clinical trial data

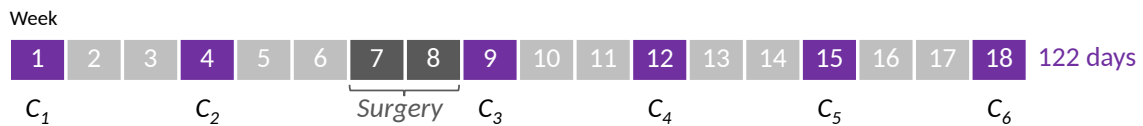
MRC BO06/EORTC 80931 randomized controlled trial (*International Standard Randomised Controlled Trial Number*: <https://www.isrctn.com/ISRCTN86294690>, ISRCTR 86294690) was funded by the Medical Research Council (MRC) (<https://www.ukri.org/councils/mrc/>) and the European Organisation for Research and Treatment of Cancer (EORTC) (<https://www.eortc.org>). BO06 Randomized Controlled Trial (RCT) compared the effectiveness combination chemotherapy and surgery in operable osteosarcoma using the conventional European Osteosarcoma Intergroup (EOI) treatment of doxorubicin (DOX) and cisplatin (CDDP) versus a dose-intensified regimen of DOX and CDDP supported by granulocyte colony-stimulating factor (G-CSF). Results of the primary analyses can be found in Lewis *et al.* (2007) [119].

### 4.2.1. Trial protocol

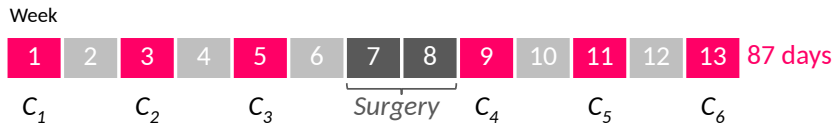
Newly diagnosed patients with non-metastatic high-grade operable osteosarcoma were recruited between 1993 and 2002. BO06 RCT randomised patients between conventional treatment with DOX and CDDP given every 3 weeks (*Reg-C*) versus a dose-intense regimen of the same two drugs given every 2 weeks supported by G-CSF (*Reg-DI*). Chemotherapy was administered for six cycles (a cycle is a period of either 2 or 3 weeks depending on the allocated regimen), before and after surgical removal of the primary osteosarcoma. Surgery to remove the primary tumour was scheduled at week 6 after starting treatment in both arms, i.e., after 2 cycles ( $2 \times [\text{DOX}+\text{CDDP}]$ ) in *Reg-C* and after 3 cycles ( $3 \times [\text{DOX}+\text{CDDP}]$ ) in *Reg-DI*. Postoperative chemotherapy was intended to resume 2 weeks after surgery in both arms. Planned total cumulative dose was  $1,050 \text{ mg/m}^2$  in both regimens, i.e.,  $25 \text{ mg/m}^2/\text{day}$  for 3 days of DOX plus  $100 \text{ mg/m}^2$  of CDDP as a continuous 24-h infusion on cycle-day 1 were given at each cycle. Planned treatment time from beginning first cycle was 122 and 87 days for *Reg-C* ( $5 \text{ cycles} \cdot 3 \text{ weeks/cycle} \cdot 7 \text{ days/week} + 14 \text{ days of surgery period} + 3 \text{ days of last cycle}$ ) and *Reg-DI* ( $5 \text{ cycles} \cdot 2 \text{ weeks/cycle} \cdot 7 \text{ days/week} + 14 \text{ days of surgery period} + 3 \text{ days of last cycle}$ ), respectively. Figure 4.2 shows the trial design.

Patients baseline characteristics (age, sex, allocated chemotherapy regimen, site and location of the tumour) were registered at randomization. Treatment-related factors (administered dose of chemotherapy, cycles delays, haematological and biochemical parameters,

**Regimen-C:** 6 cycles of DOX+CDDP every 3 weeks (DOX: 75 mg/m<sup>2</sup>/cycle; CDDP: 100 mg/m<sup>2</sup>/cycle)



**Regimen-DI:** 6 cycles of DOX+CDDP every 2 weeks (DOX: 75 mg/m<sup>2</sup>/cycle; CDDP: 100 mg/m<sup>2</sup>/cycle)



**Figure 4.2.** Patients are randomized at baseline to one of the two regimens, with the same anticipated cumulative dose (DOX: 25 mg/m<sup>2</sup>/d for 3 days + CDDP: 100 mg/m<sup>2</sup> as a continuous 24-h infusion on day 1) but different duration (3-weekly vs 2-weekly cycles, i.e., 122 vs 87 days).

chemotherapy-induced toxicity) were collected prospectively at each cycle of chemotherapy [119]. The resected specimen was examined histologically to assess response to pre-operative chemotherapy. Haematological and biochemical laboratory tests were usually performed before each cycle of chemotherapy (for blood count also at the expected nadir of the course, that is day 10 of the cycle in *Reg-C* or day 8 in *Reg-DI*) in order to monitor patient's health status and the development of toxicities or adverse events. Delays or chemotherapy dose reductions during treatment were possible in case of toxicity. Non-haematological chemotherapy-induced toxicity for nausea/vomiting, mucositis, neurological toxicity, cardiac toxicity, ototoxicity and infection were graded according to the Common Terminology Criteria for Adverse Events Version 3 [208] (see next chapters). Additional details related to the trial protocol can be found in [119].

#### 4.2.2. Sample cohort selection and baseline characteristics

BO06 trial dataset included 497 eligible patients; 19 patients who did not start chemotherapy (13) or reported an abnormal dosage of one or both agents (6) were excluded. Motivated by the clinical research question concerning the effect of doses intensity on survival, only patients who completed all six cycles within 180 days (i.e.,  $T_0^*$  of the *observation period*) were included in the analyses. The final cohort of 377 patients (75.9% of the initial sample) is shown in Figure 4.3. Among them, one subject presented  $T_i < T_0^*$  and was excluded from the FunCM cohort (376 patients – 75.7% of the initial sample).

Follow-up starts from date of randomization ( $T_0$ ) and the *observation period*  $[T_0; T_0^*]$  is given by the first 180 days after randomization (i.e., the 6-months chemotherapy treatment period). Patients' characteristics at baseline are provided in Table 4.1. Three age groups were defined according to [43]: *child* (male: 0-12 years; female: 0-11 years), *adolescent* (male: 13-17 years; female: 12-16 years) and *adult* (male: 18 or older; female: age 17 years or older). Among 377 patients, 229 (60.7%) were males and *Reg-DI* was allocated in 52.3% of the patients (197). Median age was 15 years (IQR = [11; 18]) with 40.9% of *adolescents* (154) and 30.2% of *adults* (114). Median follow-up time, computed using the



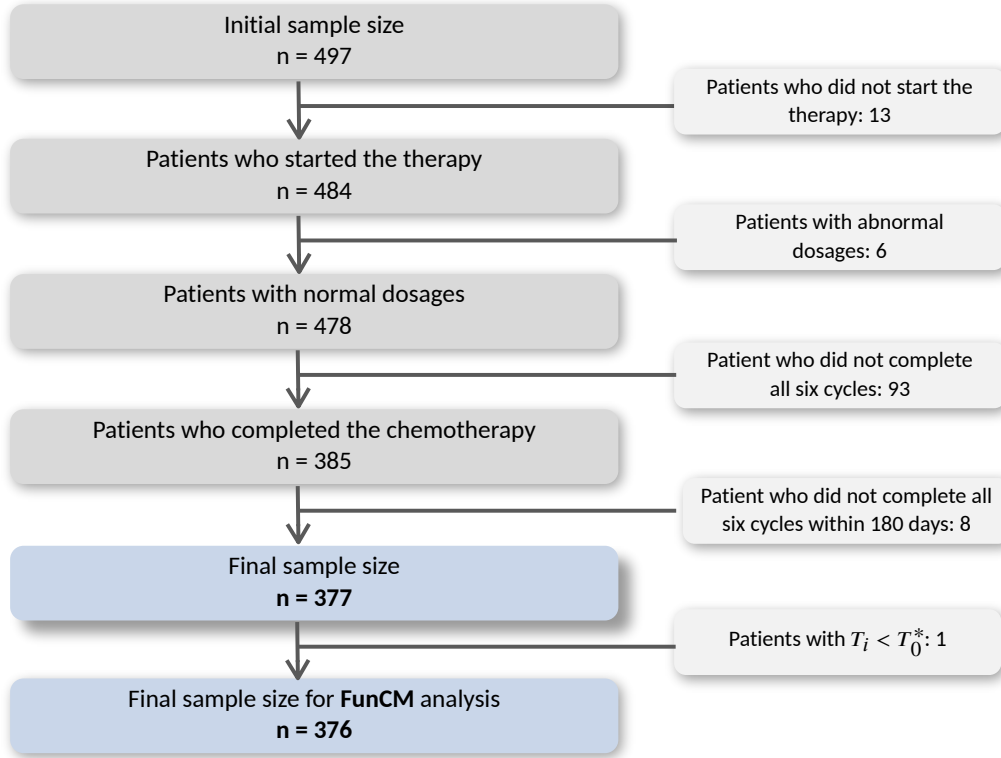


Figure 4.3. Flowchart of cohorts selection.

reverse Kaplan-Meier method by [182], was 62.19 months (IQR = [38.93; 87.46]) and 245 patients (65%) were alive at the last follow-up visit.

## 4.3. Results

Since the role of received chemotherapy dose, ALP and WBC biomarkers on patient's survival is still unclear for osteosarcoma [165, 71, 111], a new time-varying/functional perspective may help in understanding their relationship, providing new insights for childhood cancer. In this regard, the methodologies proposed in Section 4.1 were applied to the BO06 trial. Statistical analyses were performed in the R-software environment [161]. R code is provided here: <https://github.com/mspreafico/BO06-FunCM>.

### 4.3.1. Time-varying characteristics

Due to the skewed nature of the longitudinal trajectories of both ALP and WBC biomarkers, their logarithmic transformations shifted by one were considered. The vectors of longitudinal values of ALP and WBC measurements for patient  $i$  are given as

$$\mathbf{z}_i^{(ALP)} = \left\{ z_i^{(ALP)}(t_{il}), l = 1, \dots, n_i^{(ALP)} \right\} \quad (4.5)$$

$$\mathbf{z}_i^{(WBC)} = \left\{ z_i^{(WBC)}(t_{il}), l = 1, \dots, n_i^{(WBC)} \right\} \quad (4.6)$$

**Table 4.1.** Patients' characteristics at randomization and histological responses.

<b>Baseline characteristic</b>	
<b>Patients</b>	377
<b>Age*</b> [years]	
Median (IQR)	15 (11; 18)
Minimum/maximum	3/40
<i>Child*</i>	109 (28.9%)
<i>Adolescent*</i>	154 (40.9%)
<i>Adult*</i>	114 (30.2%)
<b>Sex</b>	
<i>Female</i>	148 (39.3%)
<i>Male</i>	229 (60.7%)
<b>Allocated treatment</b>	
<i>Regimen-C</i>	180 (47.7%)
<i>Regimen-DI</i>	197 (52.3%)
<b>Site of tumour</b>	
<i>Femur</i>	227 (60.2%)
<i>Fibula</i>	22 (5.8%)
<i>Humerus</i>	37 (9.8%)
<i>Radius</i>	3 (0.8%)
<i>Tibia</i>	87 (23.1%)
<i>Ulna</i>	1 (0.3%)
<b>Location of tumour</b>	
<i>Distal</i>	217 (57.6%)
<i>Mid-shaft</i>	11 (2.9%)
<i>Proximal</i>	148 (39.2%)
Missing (NA)	1 (0.3%)
<b>Histological Response**</b>	
<i>Poor</i>	186 (49.3%)
<i>Good</i>	144 (38.2%)
Missing (NA)	47 (12.5%)
<b>White Blood Count‡</b> [ $\times 10^9/L$ ]	
Median (IQR)	7.65 (6.30; 9.13)
Minimum/maximum	3.60/16.20
<b>Alkaline Phosphatase‡</b> [IU/L]	
Median (IQR)	311.5 (190.0; 551.5)
Minimum/maximum	49.0/3680.0

\* Age groups were defined according to Collins *et al.* (2013) [43]: *child* (male: 0-12 years; female: 0-11 years), *adolescent* (male: 13-17 years; female: 12-16 years) and *adult* (male: 18 or older; female: age 17 years or older).

\*\* The resected specimen was examined histologically to assess response to pre-operative chemotherapy [119]. *Good* histological response was defined as  $\geq 90\%$  necrosis in the tumour resected; 10% or more viable tumour after pre-operative chemotherapy was defined *poor* [119].

‡ Baseline White Blood Count (WBC) and Alkaline Phosphatase (ALP) levels refer to the measure performed before the beginning of cycle 1, i.e., at randomization.

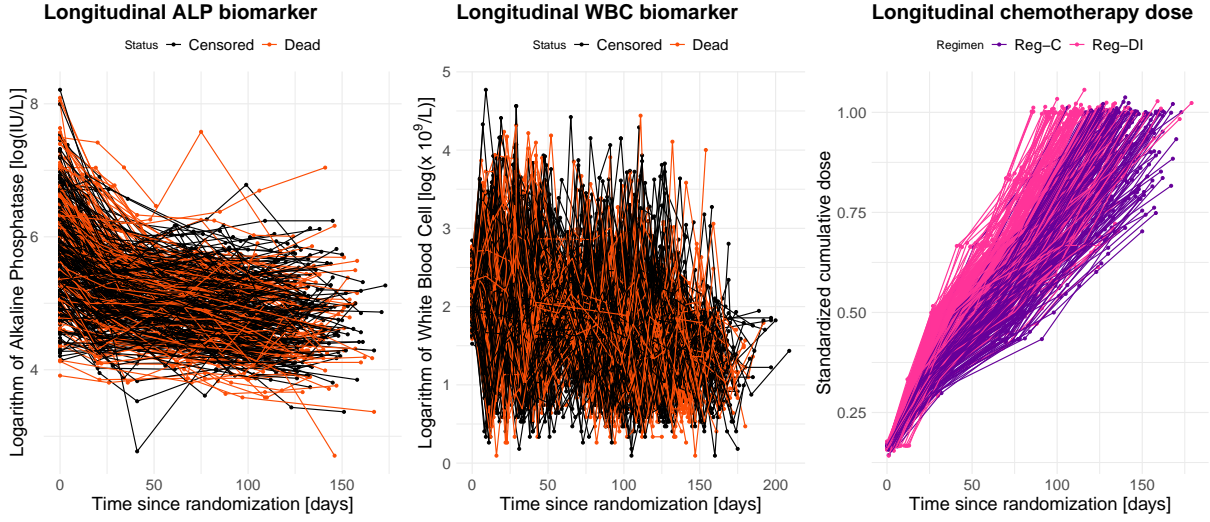
#### 4. Modelling time-varying covariates effect on survival via Functional Data Analysis

where  $t_{il}$  is the time of the  $l$ -th laboratory ALP or WBC test,  $z_i^{(ALP)}(t_{il}) = \log(ALP_{il} + 1)$  and  $z_i^{(WBC)}(t_{il}) = \log(WBC_{il} + 1)$  are the logarithmic values of ALP and WBC measurements at time  $t_{il}$ ,  $n_i^{(ALP)}$  and  $n_i^{(WBC)}$  are the number of different ALP and WBC laboratory tests, respectively. Left and central panels of Figure 4.4 show the longitudinal trajectories over time of  $z_i^{(ALP)}$  and  $z_i^{(WBC)}$  respectively. Each line represents the time-varying logarithmic biomarker values for a specific patient coloured by event status (black: *Censored*, red: *Dead*). Observed longitudinal data can be sparse and irregularly measured among patients and different biomarkers. ALP point-measurements  $z_i^{(ALP)}(t_{il})$  observed among all patients over time ranged from a minimum of 2.708 to a maximum of 8.211 (corresponding to ALP values of 14 and 3680 IU/L, respectively). WBC point-measurements  $z_i^{(WBC)}(t_{il})$  observed among all patients over time ranged from a minimum of 0.095 to a maximum of 4.771 (corresponding to WBC values of 0.1 and  $117.0 \times 10^9/L$ , respectively). The presence in both biomarkers of extremely high/low levels compared to normal ranges is due to the presence of conditions usually experienced by patients in childhood cancer therapies, such as bone growth, tumour necrosis, inflammatory states, infections or toxicity (see [218]).

The time-varying standardized cumulative dose of chemotherapy is now introduced. Let  $l \in \{1, \dots, 6\}$  be the cycle index and  $t_{il}$  the time of the  $l$ -th cycle for the  $i$ -th patient. The standardized cumulative dose of chemotherapy (DOX+CDDP) for the  $i$ -th patient at time  $t_{il}$  is defined as:

$$\begin{aligned} z_i^{(\delta)}(t_{il}) &= \frac{\text{Cumulative dose of DOX+CDDP until cycle } l \text{ [mg/m}^2\text{]}}{\text{Total target dose at the end of six cycles [mg/m}^2\text{]}} \\ &= \frac{1}{175 \text{ [mg/m}^2\text{]} \cdot 6} \cdot \sum_{c=1}^l \frac{DOX_{ic} + CDDP_{ic}}{\text{surface area}_{ic}} \left[ \frac{\text{mg}}{\text{m}^2} \right]. \end{aligned} \quad (4.7)$$

This can be interpreted as the regulated Received Dose Intensity (rRDI) introduced by Lancia *et al.* (2019) [110] evaluated over real time and not over cumulative time on treatment. For each patient  $i$ , the vector of longitudinal values of standardized cumulative dose of chemotherapy over time is defined as  $\mathbf{z}_i^{(\delta)} = \{z_i^{(\delta)}(t_{il}), l = 1, \dots, 6\}$ . The right panel of Figure 4.4 shows the longitudinal trajectories  $\mathbf{z}_i^{(\delta)}$  over time. Each line represents the individual time-varying standardized cumulative chemotherapy dose coloured by allocated regimen (pink: *Reg-DI*, purple: *Reg-C*). Patients – also within the same regimen – reported different values of standardized cumulative dose during time, depending on the delays and dose reductions required during chemotherapy due to toxicity. In particular, the lines form a tight bundle in the early phase of the treatment, but later they open up in a hand-fan shape because treatment adjustments are generally more frequent towards the end of the protocol. Median value of total standardized cumulative dose  $z_i^{(\delta)}(t_{i6})$  was 0.998 (IQR = [0.901; 1.000]), with minimum and maximum final values equal to 0.613 and 1.056, respectively. Median value of time from randomization to last cycle  $t_{i6}$  was 127 days (IQR = [114; 179]), with minimum and maximum periods of 85 and 179 days, respectively.



**Figure 4.4.** Time-varying covariates for each patient. Left panel: longitudinal logarithmic values of ALP biomarker over time coloured by event status (black: *Censored*, red: *Dead*). Central panel: longitudinal logarithmic values of WBC biomarker over time coloured by event status (black: *Censored*, red: *Dead*). Right panel: longitudinal values of standardized cumulative dose of chemotherapy coloured by allocated regimen (pink: *Reg-DI*, purple: *Reg-C*)

### 4.3.2. Time-Varying covariate Cox Model

To study the effect of time-varying biomarkers and doses on survival, a TVCM was fitted on the final cohort of 377 patients (see Figure 4.3). In particular, the hazard function in Equation (4.1) was adjusted for gender at randomization ( $\omega_i$ ) and stratified by age group  $g \in \{child, adolescent, adult\}$ , as follows:

$$h_{ig}(t|\omega_i, \mathbf{z}_i(t)) = h_{0g}(t) \exp \left\{ \theta_1 \cdot gender_i + \alpha_1 \cdot z_i^{(ALP)}(t) + \alpha_2 \cdot z_i^{(WBC)}(t) + \alpha_3 \cdot 100z_i^{(\delta)}(t) \right\} \quad (4.8)$$

where  $h_{0g}(t)$  is the baseline hazard function for the  $g$ -th age stratum,  $z_i^{(ALP)}(t)$ ,  $z_i^{(WBC)}(t)$  and  $z_i^{(\delta)}(t)$  are the time-varying covariates of ALP and WBC biomarkers and standardized cumulative dose (multiplied by 100 due to its different values scale), obtained applying LOCF method to longitudinal vectors  $\mathbf{z}_i^{(ALP)}$ ,  $\mathbf{z}_i^{(WBC)}$  and  $\mathbf{z}_i^{(\delta)}$  respectively.

In Table 4.2 hazard ratios along with their 95% confidence intervals are shown. Gender at randomization and time-varying WBC were associated to survival, whereas time-varying ALP biomarker and chemotherapy dose showed no effects on survival. Being a male was associated to a 1.5-times faster experience of the event. The higher the value of WBC at time  $t$ , the higher the risk of death. This model ignored the continuous nature of the processes underlying the data.

**Table 4.2.** Estimated hazard ratios (HR) along with 95% confidence intervals (CI) from the stratified time-varying covariate Cox model (TVCM) in Equation (4.8).

Covariates	HR	95% CI
<i>gender (male)</i>	1.539	[1.046; 2.263]
$z_i^{(ALP)}(t)$	0.991	[0.711; 1.383]
$z_i^{(WBC)}(t)$	0.647	[0.477; 0.877]
$z_i^{(\delta)}(t) \cdot 100$	1.005	[0.984; 1.027]

### 4.3.3. Functional covariate Cox Model

#### Functional representation of time-varying biomarkers and chemotherapy dose

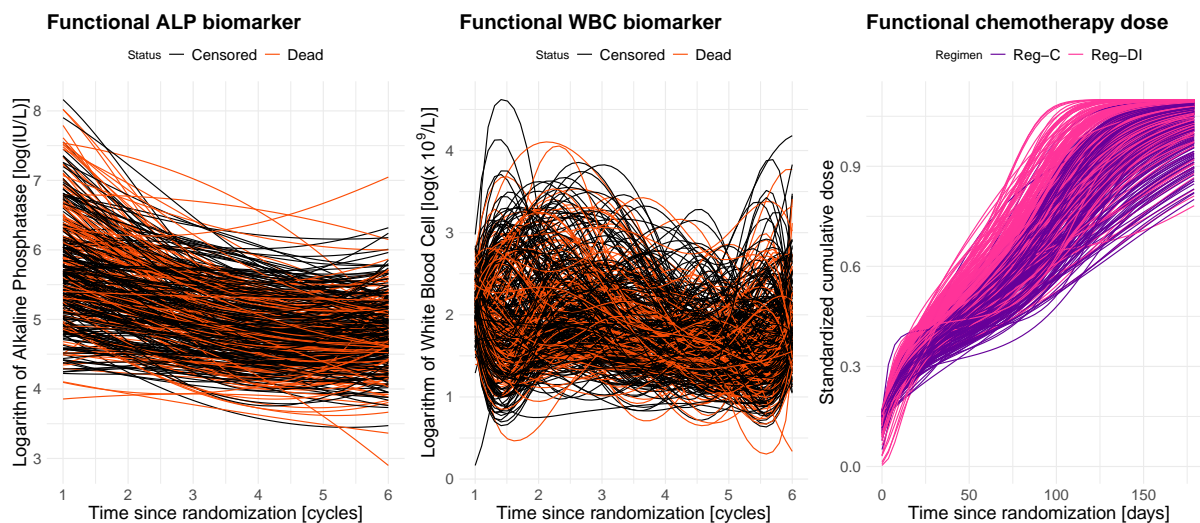
To convert the longitudinal values of ALP and WBC biomarkers registered during the *observation period*,  $\bar{z}_i^{(ALP)}$  and  $\bar{z}_i^{(WBC)}$ , into the functions  $\tilde{x}_i^{(ALP)}(t)$  and  $\tilde{x}_i^{(WBC)}(t)$ , measurements by cycles were used. This implies that all time-varying values were on the same temporal domain, i.e.,  $t \in S_{ALP} = S_{WBC} = [1, 6]$  cycles. For both ALP and WBC biomarkers ( $m = \{ALP, WBC\}$ ), B-spline basis functions  $\phi_i^{(m)}(t)$  (ALP: 2 or 3 basis of order 2 or 3; WBC: 6 or 7 basis of order 5, according to each patient  $i$ ) and a general functional form were used. Clinical bounds  $[L_m; U_m]$  (ALP: [0;9]; WBC: [0;5]) were employed in order to include the extremely high/low levels experienced by patients during treatment. Lower bounds equal to 0 were chosen to ensure the non-negativity of the functional values. A data driven approach was used to select the upper bounds defined as  $U_m = \left[ \max_{i,l} z_i^{(m)}(t_{il}) \right]$ . For each patient  $i$  the following functional ALP and WBC predictors were provided:

$$\tilde{x}_i^{(ALP)}(t) = \frac{9 \cdot \exp \left[ \phi_i^{(ALP)}(t)^T \hat{\mathbf{c}}_i^{(ALP)} \right]}{1 + \exp \left[ \phi_i^{(ALP)}(t)^T \hat{\mathbf{c}}_i^{(ALP)} \right]}, \quad (4.9)$$

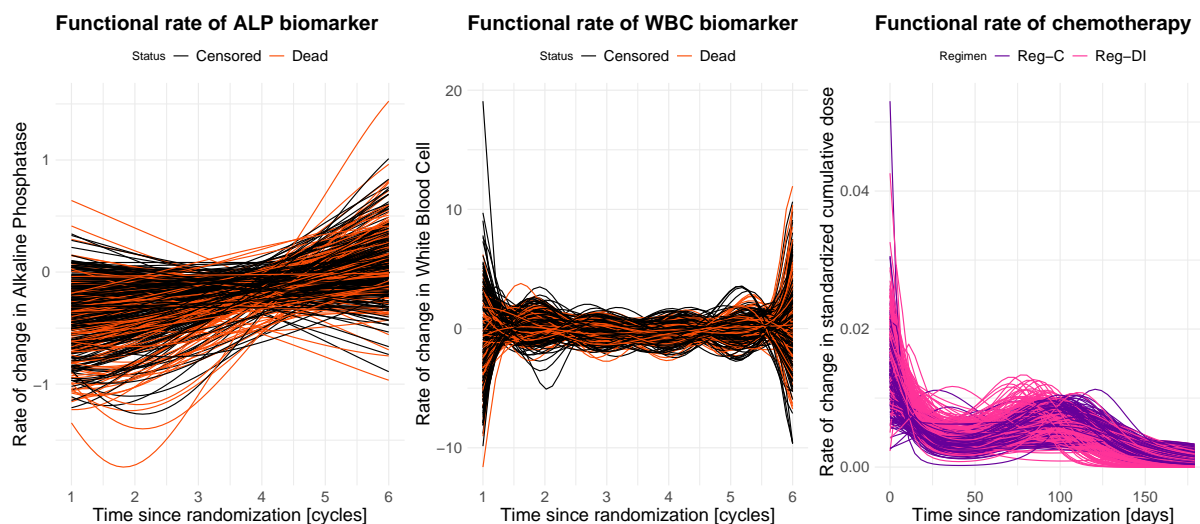
$$\tilde{x}_i^{(WBC)}(t) = \frac{5 \cdot \exp \left[ \phi_i^{(WBC)}(t)^T \hat{\mathbf{c}}_i^{(WBC)} \right]}{1 + \exp \left[ \phi_i^{(WBC)}(t)^T \hat{\mathbf{c}}_i^{(WBC)} \right]} \quad (4.10)$$

where  $\hat{\mathbf{c}}_i^{(m)}$  ( $m = \{ALP, WBC\}$ ) are the vectors of coefficients estimated by regression analysis using the transformation  $g(x) = \log \frac{x-L_m}{U_m-x}$ . Starting from the customized functional data in Equations (4.9) and (4.10), the derivatives  $\tilde{x}_i^{(dm)}(t)$  ( $m = \{ALP, WBC\}$ ), which represents the rate of change in the biomarkers values over time, were reconstructed. A graphical representation of functional biomarkers curves and their derivatives are shown in Figure 4.5 and 4.6, respectively (left panels: ALP biomarker; central panels: WBC biomarker). Each line represents the functional predictor for patient  $i$  coloured according to the death-event status.

To convert the longitudinal values of standardized cumulative chemotherapy dose  $z_i^{(\delta)}$  into the functional form  $\tilde{x}_i^{(\delta)}(t)$ , measurements in days were considered since different duration in treatment is a key-point in the chemotherapy protocol. Based on clinical motivations,



**Figure 4.5.** Left panel: functional representations of ALP biomarker over cycles coloured by status (black: *Censored*, red: *Dead*). Central panel: functional representations of WBC biomarker over cycles coloured by status (black: *Censored*, red: *Dead*). Right panel: functional representations of standardized cumulative dose of chemotherapy over time coloured by allocated regimen (pink: *Reg-DI*, purple: *Reg-C*). Each line is the graphical representation of the functional predictor of a patient.



**Figure 4.6.** Left panel: functional representations of the rate of change of ALP biomarker over cycles coloured by status (black: *Censored*, red: *Dead*). Central panel: functional representations of the rate of change of WBC biomarker over cycles coloured by status (black: *Censored*, red: *Dead*). Right panel: functional representations of the rate of change of standardized cumulative dose of chemotherapy over time coloured by allocated regimen (pink: *Reg-DI*, purple: *Reg-C*). Each line is the graphical representation of the functional predictor of a patient.

the interval  $S_\delta = [0, 180]$  days was selected, since all the patients completed the therapy within 180 days from randomization. B-spline basis functions  $\phi_i^{(\delta)}(t)$  (5 basis of order 5), a monotone functional form and clinical bounds  $L_\delta = 0$  and  $U_\delta = 1.1$  were used. For each patient  $i$  a functional predictor of standardized cumulative dose of chemotherapy was obtained:

$$\tilde{x}_i^{(\delta)}(t) = \frac{1.1 \cdot \exp\left(\widehat{\beta}_{0i} + \widehat{\beta}_{1i} \int_0^t \exp\left[\phi_i^{(\delta)}(u)^T \widehat{\mathbf{c}}_i^{(\delta)}\right] du\right)}{1 + \exp\left(\widehat{\beta}_{0i} + \widehat{\beta}_{1i} \int_0^t \exp\left[\phi_i^{(\delta)}(u)^T \widehat{\mathbf{c}}_i^{(\delta)}\right] du\right)} \quad (4.11)$$

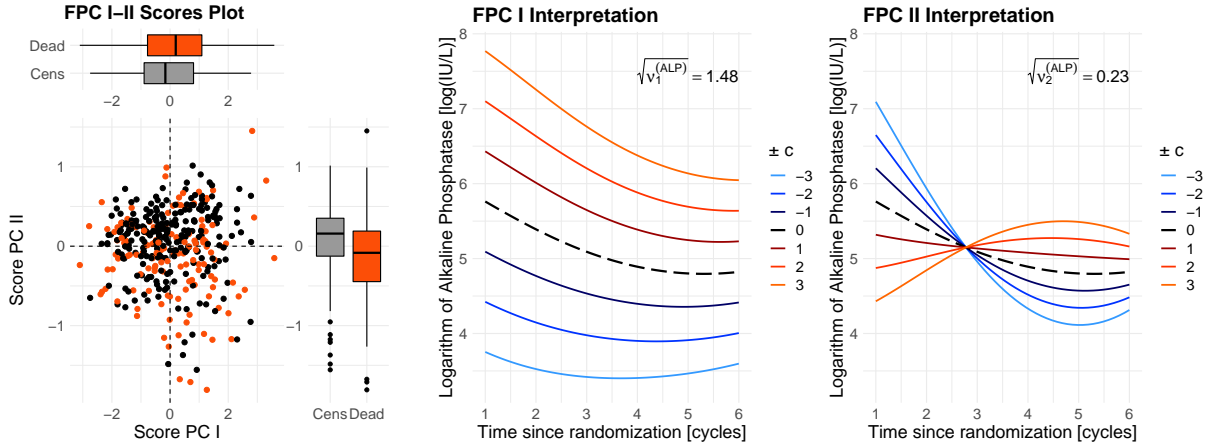
where  $\widehat{\mathbf{c}}_i^{(\delta)}$  is the vector of coefficients estimated by penalized regression analysis using the transformation  $g(x) = \log \frac{x-L_\delta}{U_\delta-x}$ . Finally, starting from the customized functional data in Equation (4.11), the derivatives  $\tilde{x}_i^{(d\delta)}(t)$ , which represents the rate of change of chemotherapy dose over time, were reconstructed. A graphical representation of functional standardized cumulative dose curves  $\tilde{x}_i^{(\delta)}(t)$  and their derivatives  $\tilde{x}_i^{(d\delta)}(t)$  are shown in right panels of Figure 4.5 and 4.6, respectively. Each line represents the functional predictor for patient  $i$  coloured according to the allocated regimen. Functional standardised cumulative dose curves  $\tilde{x}_i^{(\delta)}(t)$  (right panel in Figure 4.5) also provide information on treatment adjustments. Dose reductions are represented by final standardised cumulative dose smaller than 1. For patients with a similar final dose, the slope displays information on the duration of treatment: the lower the slope, the longer the duration of treatment, reflecting delays compared to protocol.

Figure 4.5 and 4.6 show that, taking into account the continuous nature of the processes underlying the data, a customized functional representation of the time-varying covariates and their derivatives highlights trends and variations in the shape of the processes over time.

### Functional principal component analysis for time-varying biomarkers and chemotherapy

The functional trajectories provided in Equations (4.9), (4.10) and (4.11) and their derivatives were summarised into a finite set of covariates by applying Functional Principal Component Analyses (FPCAs). Only results of FPCA on functional predictors  $\tilde{x}_i^{(ALP)}(t)$  and  $\tilde{x}_i^{(\delta)}(t)$  are presented. In both cases, two principal components were enough to account for at least 95% of the observed variability.

Results of FPCA on functional ALP predictors  $\tilde{x}_i^{(ALP)}(t)$  are provided in Figure 4.7. Left panel reports the FPC scores plot  $\left(f_{i1}^{(ALP)}, f_{i2}^{(ALP)}\right)$  with relative boxplots, which show the distributions of the estimated FPC score values among censored and dead patients. Each point represents a patient coloured by status (black: *Censored*, red: *Dead*). Central and right panels displays how to interpret the first two Principal Components  $\xi_k^{(ALP)}$ , showing the average ALP curve  $\mu^{(ALP)}(t) \pm c\sqrt{\nu_k^{(ALP)}} \cdot \xi_k^{(ALP)}(t)$  where  $\nu_k^{(ALP)}$  is the is eigenvalue related to the  $k$ -th component and  $c$  are constants chosen in order to let the scores values lie within one, two or three ( $\pm c = \pm 1, \pm 2, \pm 3$ ) standard deviations (i.e.,

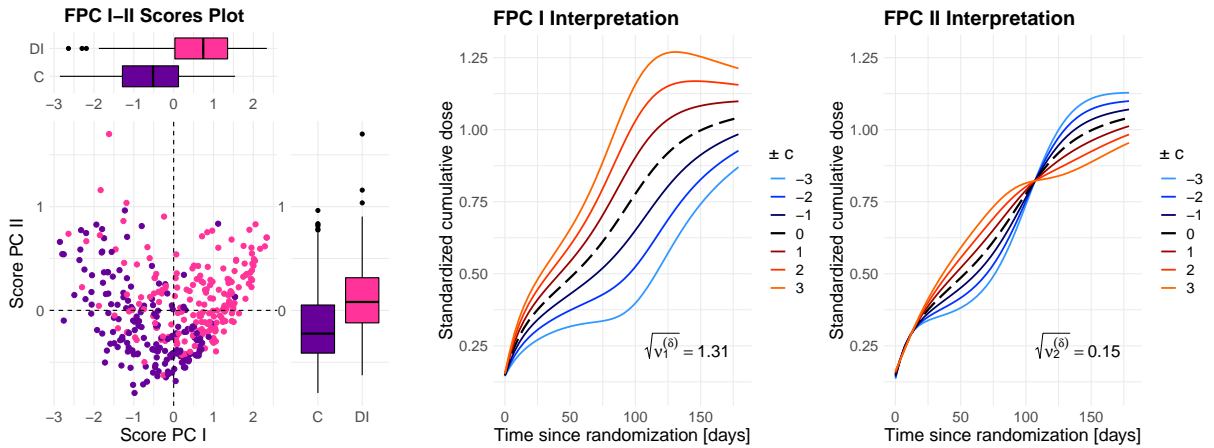


**Figure 4.7.** FPCA for functional Alkaline Phosphatase  $\tilde{x}_i^{(ALP)}(t)$ .

Left panel: Functional PC scores plot  $(f_{i1}^{(ALP)}, f_{i2}^{(ALP)})$  with boxplots (black: *Censored*, red: *Dead*).

Central panel: Interpretation of first FPC  $\xi_1^{(ALP)}$  – average standardized cumulative dose  $\mu^{(ALP)}(t) \pm c\sqrt{\nu_1^{(ALP)}} \cdot \xi_1^{(ALP)}(t)$ , with  $\sqrt{\nu_1^{(ALP)}} = 1.48$  and  $\pm c = \pm 1, \pm 2, \pm 3$ .

Right panel: Interpretation of second FPC  $\xi_2^{(ALP)}$  – average standardized cumulative dose  $\mu^{(ALP)}(t) \pm c\sqrt{\nu_2^{(ALP)}} \cdot \xi_2^{(ALP)}(t)$ , with  $\sqrt{\nu_2^{(ALP)}} = 0.23$  and  $\pm c = \pm 1, \pm 2, \pm 3$ .



**Figure 4.8.** FPCA for functional standardized cumulative dose  $\tilde{x}_i^{(\delta)}(t)$ .

Left panel: Functional PC scores plot  $(f_{i1}^{(\delta)}, f_{i2}^{(\delta)})$  with boxplots (pink: *Reg-DI*, purple: *Reg-C*).

Central panel: Interpretation of first FPC  $\xi_1^{(\delta)}$  – average standardized cumulative dose  $\mu^{(\delta)}(t) \pm c\sqrt{\nu_1^{(\delta)}} \cdot \xi_1^{(\delta)}(t)$ , with  $\sqrt{\nu_1^{(\delta)}} = 1.31$  and  $\pm c = \pm 1, \pm 2, \pm 3$ .

Right panel: Interpretation of second FPC  $\xi_2^{(\delta)}$  – average standardized cumulative dose  $\mu^{(\delta)}(t) \pm c\sqrt{\nu_2^{(\delta)}} \cdot \xi_2^{(\delta)}(t)$ , with  $\sqrt{\nu_2^{(\delta)}} = 0.15$  and  $\pm c = \pm 1, \pm 2, \pm 3$ .



#### 4. Modelling time-varying covariates effect on survival via Functional Data Analysis

square roots of  $\nu_k^{(ALP)}$ ). The first component  $\xi_1^{(ALP)}$  explained 83.8% of the variability and a positive (negative) score reflected higher (lower) values of ALP trajectories during treatment compared to the mean (left panel). The second component  $\xi_2^{(ALP)}$  explained 13.1% of the variability and positive scores reflected “flat” curves, whereas negative score reflected curves with highly negative slopes in the first cycles (right panel). The lower the score, the higher the ALP levels during the first two cycles of the treatment. FPC scores thus summarize the different patterns of the functional biomarker trajectories between patients during treatment, being a more informative representation than the baseline value or the last available measure used through LOCF.

Results of FPCA on functional standardized cumulative dose  $\tilde{x}_i^{(\delta)}(t)$  are shown in Figure 4.8. Left panel reports the FPC scores plot  $(f_{i1}^{(\delta)}, f_{i2}^{(\delta)})$  with relative boxplots, which show the distributions of the estimated FPC score values among the two regimens. Each point corresponds to a patient. Different colours represent the two regimens. Central and right panels displays how to interpret the first two Principal Components  $\xi_k^{(\delta)}$ , showing the average curve  $\mu^{(\delta)}(t) \pm c\sqrt{\nu_k^{(\delta)}} \cdot \xi_k^{(\delta)}(t)$  where  $\nu_k^{(\delta)}$  is the is eigenvalue related to the  $k$ -th component and  $c$  are constants chosen in order to let the scores values lie within one, two or three ( $\pm c = \pm 1, \pm 2, \pm 3$ ) standard deviations (i.e., square roots of  $\nu_k^{(\delta)}$ ). The first component  $\xi_1^{(\delta)}$  explained 86.9% of the variability and reflects information on treatment administration and adjustments with respect to protocol. Positive scores (i.e., curves above the average  $\mu^{(\delta)}(t)$  in the left panel) indicate patients without dose-reduction (i.e., their final standardized cumulative dose is greater or equal to 1) and with possible delays in treatment: the lower the positive score, the higher the time needed to end the treatment. Negative scores (i.e., curves below the average  $\mu^{(\delta)}(t)$ ) represent patients with both time-delays and dose-reduction: the lower the negative score, the higher the total dose-reduction. The second component  $\xi_2^{(\delta)}$  explained 9.8% of the variability and a positive score indicated a faster growth in the chemotherapy assumption in the first period compared to the second one, with respect to the mean (right panel). Every two patients reported different values of FPC scores, reflecting delays or dose reductions during chemotherapy. This representation illustrates different treatment dynamics, also among patients allocated to the same regimen. Summarizing differences in both trends and variations related to the shape of chemotherapy doses consumption processes over time, the use of FPC scores is more informative than an IIT analysis by different allocated regimens or a LOCF approach that considers only the last available value.

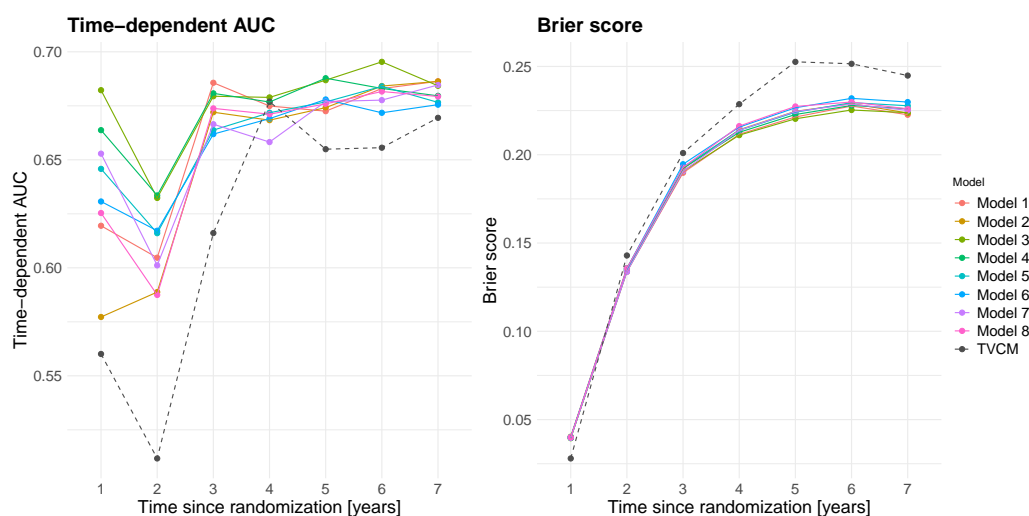
#### Multivariate functional linear Cox regression model

To study the effect of risk factors on survival, several MFLCRMs based on different sets of baseline and functional predictors (see Table 4.3) were estimated. Since functional trajectories and their relative derivatives are correlated, in each MFLCRM only one type was considered. Each model was adjusted for gender and stratified by age group at randomization  $g \in \{child, adolescent, adult\}$ . When functional rate of changes of ALP or WBC biomarkers were included in the models, the values of logarithmic ALP or WBC

**Table 4.3.** Selected truncation parameters  $K_m$  and integrated AUC (iAUC) for different sets of baseline and functional predictors. iAUC for stratified time-varying covariate Cox model (TVCM) in Equation (4.8).

Model	Baseline $\omega_i$	Truncation parameters $K_m$						iAUC
		ALP	dALP	WBC	dWBC	$\delta$	d $\delta$	
1	( $gender_i$ )	2	-	7	-	1	-	0.650
2	( $gender_i$ )	2	-	7	-	-	1	0.635
3	( $gender_i, wbc_i$ )	2	-	-	4	2	-	0.666
4	( $gender_i, wbc_i$ )	2	-	-	4	-	3	0.664
5	( $gender_i, alp_i, wbc_i$ )	-	2	-	4	2	-	0.650
6	( $gender_i, alp_i, wbc_i$ )	-	2	-	4	-	3	0.647
7	( $gender_i, alp_i$ )	-	1	7	-	1	-	0.645
8	( $gender_i, alp_i$ )	-	1	7	-	-	1	0.641
TVCM								0.592

levels at randomization were also considered as adjusting baseline covariates. Cross-validation with five folds was employed to select the truncation parameters  $K_m$  for each set of covariates (see Table 4.3). Time-dependent AUCs and Brier scores were estimated with R packages `tdROC` (function `tdROC`) by [123] and `ipred` (function `sbrier`) by [153], respectively. Figure 4.9 shows the cross-validated mean values of time-dependent AUC and Brier score over different time horizons for all estimated models (solid lines) and for TVCM in Equation (4.8) (dashed black lines). All functional models outperformed TVCM and showed similar Brier score measures over time, therefore time-dependent AUC was used to select the final model. Weighted averages of the several time-dependent AUCs over time, estimated through the integrated AUCs (iAUC) by [74], are reported in Table 4.3.



**Figure 4.9.** Left panel: time-dependent AUC over different time horizons (from 1 to 7 years after randomization) for Models 1-8 of Table 4.3 (solid coloured lines) and TVCM in Equation (4.8) (dashed black line). Right panel: Brier score over different time horizons.

#### 4. Modelling time-varying covariates effect on survival via Functional Data Analysis

According to the highest iAUC, the best MFLCRM was Model 3, defined as follows:

$$\begin{aligned}
 h_{ig} \left( t | \boldsymbol{\omega}_i, \tilde{x}_i^{(ALP)}(t), \tilde{x}_i^{(dWBC)}(t), \tilde{x}_i^{(\delta)}(t) \right) &= \\
 &= h_{0g}(t) \exp \left\{ \theta_1 \text{gender}_i + \theta_2 \text{wbc}_i + \sum_{k=1}^2 f_{ik}^{(ALP)} \alpha_k^{(ALP)} + \right. \\
 &\quad \left. \sum_{k=1}^4 f_{ik}^{(dWBC)} \alpha_k^{(dWBC)} + \sum_{k=1}^2 f_{ik}^{(\delta)} \alpha_k^{(\delta)} \right\} \quad (4.12)
 \end{aligned}$$

where  $h_{0g}(t)$  is the baseline hazard function for the  $g$ -th age stratum,  $\boldsymbol{\omega}_i = (\text{gender}_i, \text{wbc}_i)$  is the vector of baseline covariates;  $\tilde{x}_i^{(ALP)}(t)$ ,  $\tilde{x}_i^{(dWBC)}(t)$  and  $\tilde{x}_i^{(\delta)}(t)$  are the functional predictors of ALP biomarker, rate of change of WBC and standardized cumulative dose, respectively, with relative FPC scores  $f_{ik}^{(m)}$  ( $k = 1, \dots, K_m$ ;  $m \in \{ALP, dWBC, \delta\}$ ;  $K_{ALP} = 2$ ;  $K_{dWBC} = 4$ ;  $K_{\delta} = 2$ ).

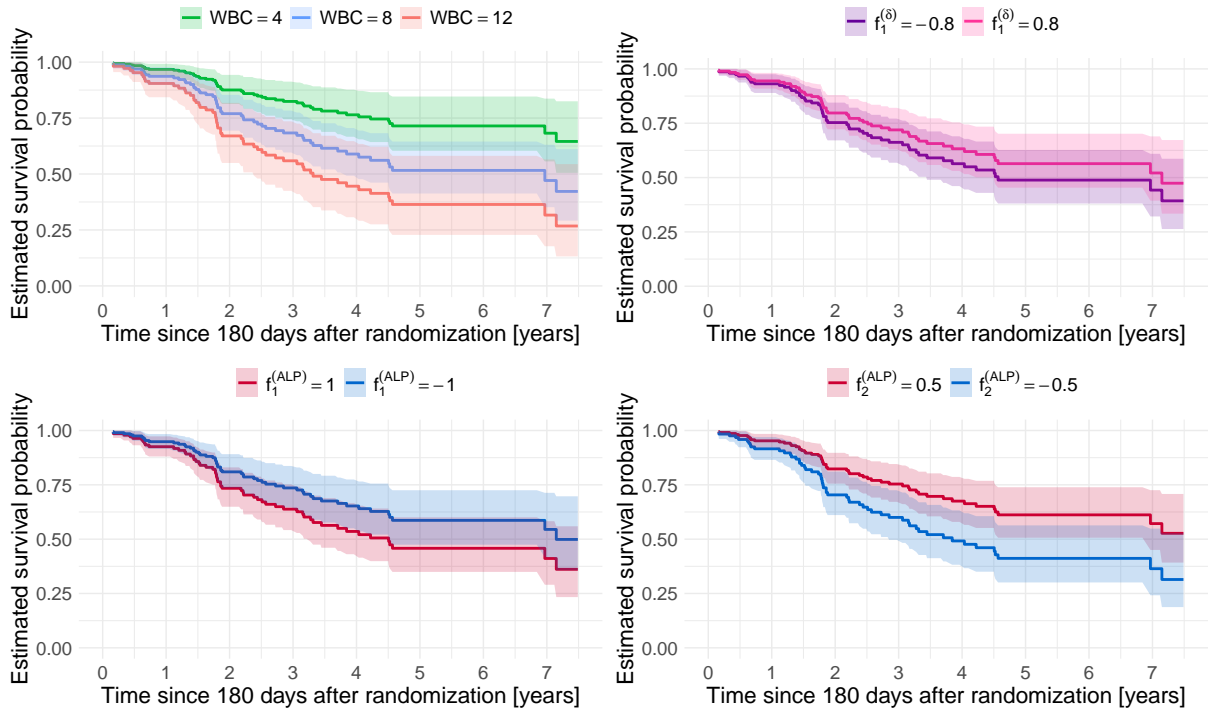
To estimate the effect of the selected functional predictors on survival, MFLCRM (4.12) was fitted on the FunCM cohort of 376 patients (see Figure 4.3). In Table 4.4 hazard ratios along with their 95% confidence interval are shown. Level of WBC at randomization and the FPC scores related to alkaline phosphatase  $f_{i1}^{(ALP)}$ ,  $f_{i2}^{(ALP)}$  were associate to survival. The higher the value of WBC at randomization the higher the risk of death, whereas no effects were observed due to the rate of change in WBC during the protocol observation period. Patients with high ALP trajectories had poor survival, especially in case of curves with highly negative slopes during the first cycles of chemotherapy protocol. FPC scores related to functional chemotherapy dose showed no effects on survival. Estimated survival probabilities are shown in Figure 4.10. High values of baseline WBC corresponded to poor survival (top-left panel). The score  $f_{i1}^{(\delta)}$  related to the first PC of functional chemotherapy indicated that there was no improvement on survival due to dose-intense profiles (top-right panel). The effect of functional ALP biomarker suggested that patients with high ALP trajectories over time (i.e., high value of  $f_{i1}^{(ALP)}$  – bottom-left panel), especially during the first cycles of the chemotherapy protocol (i.e., low value of  $f_{i2}^{(ALP)}$  – bottom-right panel), had poor survival.

#### 4.4. Final remarks

In this chapter, a novel approach based on FDA techniques to investigate the dynamics of time-varying processes over time and to include additional information that may be related to the survival into the time-to-event model was presented. Data from the MRC BO06/EORTC 80931 randomized clinical trial for osteosarcoma treatment were analysed. Biomarkers and chemotherapy dose were incorporated as time-varying covariates into time-to-event models using both a TVCM and a FunCM approach. The standard TVCM with LOCF approach ignored the continuous nature of the processes underlying the data. To overcome this issue, FunCM exploited FDA techniques to represent time-varying characteristics in terms of functions, enriching the information available for

**Table 4.4.** Estimated hazard ratios (HR) along with 95% confidence intervals (CI) from the multivariate functional linear Cox regression model.

Covariates	HR	95% CI
<i>gender (male)</i>	1.431	[0.964; 2.123]
<i>wbc</i>	3.169	[1.525; 6.585]
$f_1^{(ALP)}$	1.210	[1.018; 1.437]
$f_2^{(ALP)}$	0.554	[0.399; 0.768]
$f_1^{(\delta)}$	0.869	[0.719; 1.051]
$f_2^{(\delta)}$	0.885	[0.547; 1.432]
$f_1^{(dWBC)}$	0.990	[0.889; 1.102]
$f_2^{(dWBC)}$	0.916	[0.789; 1.062]
$f_3^{(dWBC)}$	1.161	[0.892; 1.512]
$f_4^{(dWBC)}$	1.219	[0.898; 1.655]

**Figure 4.10.** Estimated survival probability based on the multivariate functional linear Cox regression model (4.12). Time  $t_0 = 0$  corresponds to  $T_0^*$  in Fig. 4.1. Top-left panel: patients with different values of WBC [ $\times 10^9/L$ ] at randomization (green:  $WBC = 4$ ; blue:  $WBC = 8$ ; red:  $WBC = 12$ ). Top-right panel: patients with different values of the first PC score for functional chemotherapy (purple:  $f_1^{(\delta)} = -0.8$ ; pink:  $f_1^{(\delta)} = 0.8$ ). Bottom-left panel: patients with different values of the first PC score for functional ALP biomarker (red:  $f_1^{(ALP)} = 1$ ; blue:  $f_1^{(ALP)} = -1$ ). Bottom-right panel: patients with different values of the second PC score for functional ALP biomarker (red:  $f_2^{(ALP)} = 0.5$ ; blue:  $f_2^{(ALP)} = -0.5$ ). When not specified, the other risk factors are fixed to the most frequent class for categorical covariates, i.e., *adolescent males*, and to the median value for continuous ones, i.e.,  $WBC = 7.65 \times 10^9/L$  at randomization,  $f_1^{(\delta)} = 0.08$ ,  $f_2^{(\delta)} = -0.03$ ,  $f_1^{(ALP)} = -0.10$ ,  $f_2^{(ALP)} = 0.07$ ,  $f_1^{(dWBC)} = -0.12$ ,  $f_2^{(dWBC)} = -0.02$ ,  $f_3^{(dWBC)} = -0.07$  and  $f_4^{(dWBC)} = -0.08$ .

#### 4. Modelling time-varying covariates effect on survival via Functional Data Analysis

modelling survival with relevant time-varying features related to the evolution of the processes over time. These features were included into MFLCRMs by FPCA to study the effects of functional risk factors on patients' overall survival.

Differences in results for TVCM and MFLCRM were due to the different nature of the information incorporated in the two models. As a piecewise-constant approach, TVCM considered as constants the last biomarkers/dose levels over different time points (expressed in days). In practice, among the measurements recorded during the *observation period*, only the last value had any real impact on overall survival, as only one patient presented with a time-to-event of less than 180 days. This discarded both information about the continuous nature of the processes and the history of the actual levels measured. MFLCRM included information related to different levels variations and timing during the entire *observation period*, and functional biomarkers were defined over cycles. Thanks to the introduction of relevant dynamic features related to the continuous functional nature of the processes, MFLCRM resulted more informative than TVCM, outperforming it both in terms of calibration and discrimination over time. MFLCRM results suggested that osteosarcoma patients with high ALP trajectories during treatment, especially during the first cycles of the chemotherapy protocol, have poor overall survival. Dose-intense profiles were not associated with survival, even if functional chemotherapy representations were able to capture individual realisations of the intended treatment, detecting differences between patients randomised to the same regimen. This suggested that considering only the assumed dose as treatment proxy is not enough. Chemotherapy presents some particular aspects, such as latent accumulation of toxicity, which must be taken into account [112].

The proposed FunCM focused on the representation and the reconstruction of the functional trajectories related to the time-varying processes of interest. Such data are usually considered in a very simplistic way in cancer prediction models, where they act as fixed baseline or as time-dependent LOCF covariates. In this way the amount of information they may provide is not considered, as it is often the changing patterns of the functional trajectories rather than the baseline/last value that affects patients' survival. The strength and innovation of FunCM was the ability to capture the individual realisations of the process over time through a customized functional reconstruction. The developed techniques allowed (i) to account for the continuous time-varying nature of the processes underlying the data and their properties, such as nonlinearity, positivity, constraints, monotonicity, (ii) to move from sparse and irregular longitudinal data to functions defined over a common continuous domain, overcoming the issues of values missingness and different temporal grids, and (iii) to reconstruct and provide derivatives information in a tailored way. The use of derivatives is important both in extending the range of simple graphical exploratory methods and in the development of more detailed methodology [162]. In fact, interesting patterns are often much more apparent in derivatives than in the original curves. Furthermore, through a proper dimensionality reduction technique, this methodology allowed to extract additional information contained in the functions. This result is an effective exploratory and modelling technique to highlight trends and variations in the evolution of the processes over time.

In contrast to a TVCM approach, the use of FunCM requires that patients survived for a period at least equal to the length of the *observation period* used to compute the functional predictors. This might imply a loss of information in situations with high rate of mortality during the *observation period* (that is not the case under study as only one of the cohort patients who had completed the chemotherapy treatment protocol died during the first 6-months after randomization – see Figure 4.3). In those cases, a joint modelling approach can be used to overcome both LOCF and selection bias issues, since it allows the simultaneous modelling of longitudinal and time-to-event outcomes. However, joint models are computationally expensive in case of multiple longitudinal outcomes and require assumptions on the distributions of the processes that need to be carefully validated to avoid biased estimates.

This work opens doors to many further developments, both in the field of statistical methods and in cancer research. The dimensionality reduction via FPCA is just one way to work with these data in order to use them within inferential contexts. In fact, the reconstruction via FDA allows to properly use the functional data to address relevant clinical research questions, according to the needs of the analysis and the outcomes of interest. From a clinical point of view, it will be necessary to simultaneously consider chemotherapy modifications and the occurrence of adverse events. This aspect needs to be taken into account into the representation of the dynamic evolution of these processes. To model them simultaneously is not straightforward, as we will see in Chapter 7.

In conclusion, this study showed that working in this direction is a difficult but profitable approach, which could lead to new improvements for subject-specific survival predictions and personalised treatment. The complexity of the phenomenon asks for the developments of new methodologies able to deal with the peculiar aspects of chemotherapy treatment, such as the presence of multiple types of toxic side effects during chemotherapy cycles. In this sense, Chapters 5 and 6 will be devoted to the development of new methods, still lacking in the medical literature, capable of appropriately representing the overall toxicity burden experienced by patients during treatment.



# A novel longitudinal method for quantifying multiple overall toxicity

This chapter has been published in *BMJ Open*, 11:e053456, 2021 as M. Spreafico, F. Ieva, F. Arlati, *et al.* “Novel longitudinal Multiple Overall Toxicity (MOTox) score to quantify adverse events experienced by patients during chemotherapy treatment: a retrospective analysis of the MRC BO06 trial in osteosarcoma” [190].

In cancer trials the relationship between chemotherapy dose and clinical efficacy outcomes is problematic to analyse due to the presence of negative feedback between exposure to cytotoxic drugs and other aspects, such as latent accumulation of chemotherapy-induced toxicity. Toxic Adverse Events (AEs), developed by patients through a chemotherapy cycle, affect subsequent exposure by delaying the next cycle or reducing its dosage, representing one of the principal reasons for treatment discontinuation [186]. The introduction of the Common Terminology Criteria for Adverse Events (CTCAE) [208] multimodality grading system greatly facilitated the standardized reporting of AEs and the comparison of outcomes between trials and institutions [204, 226]. According to CTCAE, AEs range in severity from minor, asymptomatic changes to life-threatening injuries or death [204]. Characterisation of toxicity is of interest to patients and clinicians engaged in shared decision making about a treatment strategy [198]. Toxicities are at the same time risk factors for mortality and predictors of future exposure levels, representing time-dependent confounders for the effect of chemotherapy on patient’s status [112]. Incorporating time into analysis of toxicity is important for the comparison between different chemotherapy regimens or even multiple toxicities from the same regimen [199]. Therefore, it is crucial to provide an effective tool to assess the evolution of overall toxicity over chemotherapy treatment and to guide the therapy strategy.

Since patients might have different types and number of AEs, to summarize toxicity during treatment and investigate the true extent of toxic burden represent challenging problems in cancer research. Due to the complexity of longitudinal chemotherapy data, no standard method is available for summarising AEs data into a concise score of overall risk. Toxicity data are usually analysed in cancer prediction models by looking at the maximum toxicity over time (max-time) or maximum grade among events (max-grade) [204, 226, 199, 184, 140, 205]. Although both methods can summarise data over time, a lot of information are not used. The max-time method summarises longitudinal data



## 5. A novel longitudinal method for quantifying multiple overall toxicity

into a single AE profile by using the worst (maximum-severity) grade over treatment for each toxic event, without distinguish between isolated and repeated episodes. The max-grade method summarises all the toxic AEs through the maximum grade among all types of events, without discerning between single or multiple episodes. Other methods, i.e., weighted sums of individual toxic effects [205, 28, 172, 117, 35], have also been proposed to consider longer-lasting lower-grade chronic toxicities, which may have impact on patient's quality of life. However, these approaches do not provide information about AEs timing or severity at a given cycle during treatment. The inclusion of time-related information could provide intuitions on AEs and their evolution over time [198], giving new insights in cancer treatment.

In this framework, alternative methods of longitudinal toxic event evaluation have been proposed [198, 205, 114, 84, 200] but none of them is focused on analysing the evolution of high overall toxicity over treatment using a cycle-by-cycle approach. To quantify risk for each patient including a time-dimension, in this chapter a new longitudinal Multiple Overall Toxicity (MOTox) score is proposed. At each cycle, this score summarises multiple CTCAE-graded AEs, and describe the overall toxic status along with the most severe risk event. The evolution of high MOTox scores over cycles is then studied using logistic regression models to predict high overall toxicity at the end of the cycle using personalized characteristics, achieved chemotherapy dose, previous toxicities, biochemical and haematological factors over time. To illustrate the use of the longitudinal MOTox procedure to quantify how chemotherapy-induced toxicities may evolve in cancer patients, a retrospective analysis was conducted on MRC BO06/EORTC 80931 Randomized Controlled Trial (RCT) for the treatment of osteosarcoma [119]. As mentioned in Chapter 4, patients were treated with cisplatin (CDDP) and doxorubicin (DOX), two cytotoxic drugs commonly used in the treatment of various types of human cancers. Both DOX and CDDP are characterized by various toxic AEs: apart from nausea, specific renal and neurotoxicity [68, 9] for CDDP or cardiotoxicity [228, 227] for DOX. Longitudinal MOTox scores over therapy were computed considering non-haematological toxicity. Demographics, treatment-related and biochemical characteristics were used to examine high overall toxicity over cycles. Provided that longitudinal CTCAE-graded toxicity data are available, the novel MOTox scores can be applied to analyse data from any cancer treatment.

The rest of this chapter is organized as follows. In Section 5.1 non-haematological toxicity data in BO06 trial are described. Longitudinal MOTox scores and statistical methodologies are introduced in Section 5.2. Results are presented in Section 5.3. Section 5.4 ends with a discussion of strengths and limitations of the current approach, identifying some developments for future research.

### 5.1. BO06 data

Data from the MRC BO06/EORTC 80931 RCT for the treatment of osteosarcoma [119] were analysed. Specifically, we focused on the final cohort of 377 patients who completed all six cycles of chemotherapy within 180 days after randomization without abnormal

dosages (i.e., +25% higher than planned). This cohort is the same as the one analysed in Chapter 4, under TVCM analysis (see Figure 4.3 in Chapter 4). Patient characteristics at randomization have been reported in Table 4.1 in Chapter 4.

### 5.1.1. Treatment-related factors

As reported in Section 4.2.1 in Chapter 4, BO06 RCT protocol established that treatment-related factors (administered dose of chemotherapy, cycles delays, haematological and biochemical parameters, chemotherapy-induced toxicity and histological response to pre-operative chemotherapy) were collected prospectively at each cycle of chemotherapy [119].

Laboratory tests were performed over cycles in order to monitor patient's health status and the development of toxicities or adverse events. Specifically, levels of alkaline phosphatase, renal clearance, lactate dehydrogenase, calcium and magnesium were measured at the beginning of each cycle (i.e., before the drugs administration) according to local practice. Blood counts (white blood cells, neutrophils, platelets) were obtained before each cycle and at the expected nadir of the course (day 10 of the cycle in *Reg-C*, day 8 in *Reg-DI*). A summary of the biochemical and haematological values measured for the selected cohort over the entire dataset is shown in Table 5.1.

Delays or chemotherapy dose reductions during treatment were possible in case of toxicity. Specifically, non-haematological chemotherapy-induced toxicity related to nausea/vomiting (*naus*), infection (*inf*), oral mucositis (*oral*, i.e., inflammation of the mucosae of the gastrointestinal tract, especially the oral ones), cardiac toxicity (*car*, i.e., heart dysfunctions), ototoxicity (*oto*, i.e., hearing loss) and neurological toxicity (*neur*) were registered at each cycle and graded according to the Common Terminology Criteria for Adverse Events Version 3 (CTCAE v3.0) [208], with grades ranging from 0 (none) to 4 (life-threatening) (see Table 5.2). Grades of chemotherapy-induced non-haematological toxicity over cycles recorded for the selected cohort are reported in Figure 5.1. Nausea/vomiting was reported at least once over cycles in 97.3% of patients (367/377), with a percentage that decreased over cycles from 84.9% in cycle 1 to 52.5% in cycle 6. The percentages of patients who reported oral mucositis or infections were more stable over cycles:

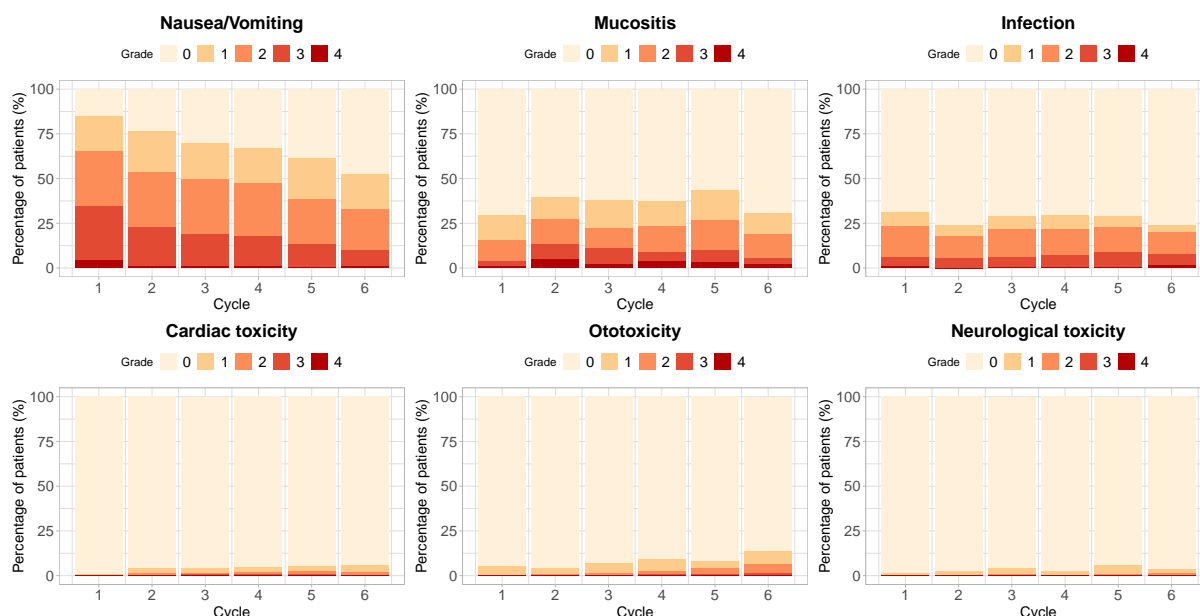
**Table 5.1.** Descriptive of biochemical and haematological values over the entire dataset.

Biomarkers*	Mean (s.d.)	Median (IQR)	Min/Max
<b>White Blood Count</b> [ $\times 10^9/L$ ]	7.36 (8.25)	5.00 (3.10; 8.20)	0.10/117
<b>Neutrophils</b> [ $\times 10^9/L$ ]	4.74 (6.93)	2.60 (1.12; 5.30)	0/83.38
<b>Platelets</b> [ $\times 10^9/L$ ]	219.8 (157.5)	190 (99; 311)	2/999
<b>Renal Clearance</b> [ $ml/min/1.73m^2$ ]	112.3 (34.9)	110 (90; 132)	8/396
<b>Alkaline Phosphatase</b> [ $IU/L$ ]	238.5 (279.1)	162.5 (98.0; 267.2)	14/3680
<b>Lactate Dehydrogenase</b> [ $IU/L$ ]	447.0 (264.2)	394.0 (298.8; 531.0)	4/4310
<b>Calcium</b> [ $mmol/l$ ]	2.34 (0.36)	2.35 (2.25; 2.45)	0.21/9.70
<b>Magnesium</b> [ $mmol/l$ ]	0.71 (0.24)	0.69 (0.57; 0.80)	0.07/3.06

\* Haematological and biochemical laboratory tests were usually performed before each cycle of chemotherapy; for blood count also at the expected nadir of the course, that is day 10 of the cycle in *Reg-C* or day 8 in *Reg-DI*.

## 5. A novel longitudinal method for quantifying multiple overall toxicity

30.5%–43.3% for mucositis, with 78% (294/377) reporting mucositis at least once, and 23.8%–31.3% for infection, with 69% (260/377) reporting an infection at least once. Ototoxicity was reported at least once in 21.5% (81/377), cardiac toxicity in 14.1% (53/377) and neurological toxicity in 11.7% (44/377).



**Figure 5.1.** Bar-plots of chemotherapy-induced toxicity CTCAE grades over cycles (wheat: 0; light-orange: 1; orange: 2; red: 3; dark-red: 4). Each panel refers to a different type of toxicity: nausea/vomiting [top-left], mucositis [top-centre], infection [top-right], cardiac toxicity [bottom-left], ototoxicity [bottom-centre] and neurological toxicity [bottom-right].

**Table 5.2.** Toxicity coding based on Common Terminology Criteria for Adverse Events (CTCAE) v3.0 by [208] for non-haematological chemotherapy-induced toxicity related to nausea/vomiting, infection, oral mucositis, cardiac toxicity, ototoxicity and neurological toxicity.

Toxicity	Grade 0	Grade 1	Grade 2	Grade 3	Grade 4
<b>Nausea or Vomiting</b>	None	Nausea	Transient vomiting	Continuative vomiting	Intractable vomiting
<b>Infection</b>	None	Minor infection	Moderate infection	Major infection	Major infection with hypotension
<b>Oral Mucositis</b>	No change	Soreness or erythema	Ulcers: can eat solid	Ulcers: liquid diet only	Alimentation not possible
<b>Cardiac toxicity</b>	No change	Sinus tachycardia	Unifocal PVC arrhythmia	Multifocal PVC	Ventricular tachycardia
<b>Ototoxicity</b>	No change	Slight hearing loss	Moderate hearing loss	Major hearing loss	Complete hearing loss
<b>Neurological toxicity</b>	None	Paraesthesia	Severe paraesthesia	Intolerable paraesthesia	Paralysis

PVC = premature ventricular contraction

## 5.2. Statistical Methodologies

### 5.2.1. Longitudinal Multiple Overall Toxicity (MOTox) scores and outcomes

The longitudinal chemotherapy-induced Multiple Overall Toxicity (MOTox) score is introduced. Let  $\mathcal{T}$  be the set of different toxicity categories. Let  $k$  be the cycle index (which takes value  $k \in \{1, \dots, 6\}$ ) and  $tox_{ij,k}$  be the  $j$ -th toxicity level for the  $i$ -th patient at the  $k$ -th cycle with value from 0 to 4. The chemotherapy-induced MOTox score for the  $i$ -th patient at cycle  $k$  is defined as:

$$\begin{aligned} MOTox_{i,k} &= \text{average toxic level}_{i,k} + \text{worst grade}_{i,k} \\ &= \frac{1}{|\mathcal{T}|} \sum_{j \in \mathcal{T}} tox_{ij,k} + \max_{j \in \mathcal{T}} (tox_{ij,k}) \in [0, 8] \end{aligned} \quad (5.1)$$

where the *average toxic level* is hence the arithmetic mean of the grades related to all the toxic AEs registered for the patient at cycle  $k$ , and the *worst grade* is the maximum CTCAE-grade among all the toxic AEs experienced by the patient at the cycle under analysis.

The MOTox score is a *cycle-dependent longitudinal mean-max index* that quantifies the multiple types of Adverse Events (AEs) experienced by patient  $i$  during cycle  $k$ . This choice was made to include the cycle-time component in the analysis and to take into account that (i) multiple lower-grade chronic toxicities may have impact on patient's quality of life and (ii) huge level in one specific toxicity can cause severe effects and permanent consequences for the patient. MOTox score can detect differences in health status among patients, providing more information with respect to traditional methods.

This novel score only requires that the different types of toxicity necessary for the computation, are recorded according to the CTCAE grading system. In this way, this definition can be applied to different groups of CTCAE-graded toxicities and applied to any cancer treatment.

The median value of MOTox scores over all the patients in all the cycles, computed as

$$\tau = \underset{i,k}{\text{median}} (MOTox_{i,k}),$$

is defined as *global median MOTox value*. It is used as a threshold to define a longitudinal binary score for *high* (or *low*) overall toxicity, named *longitudinal high-MOTox score* :

$$\text{high-MOTox}_{i,k} = \begin{cases} 1 & \text{if } MOTox_{i,k} > \tau \\ 0 & \text{otherwise.} \end{cases} \quad (5.2)$$

that indicates if patient experienced high MOTox with respect to the global median MOTox value  $\tau$  at cycle  $k$  ( $\text{high-MOTox}_{i,k} = 1$ ) or not ( $\text{high-MOTox}_{i,k} = 0$ ).

MOTox and high-MOTox scores represent new indices to measure patients' overall toxicity related to multiple types of AEs. Binary high-MOTox scores over cycles represent the clinical endpoints used as outcome measures for high overall toxicity over treatment.

### Interpretation of longitudinal MOTox scores

As a *mean-max* index of CTCAE-graded toxicity levels ranging from 0 to 4 each, the MOTox score  $MOTox_{i,k}$  in Equation (5.1) – as well as the global median MOTox value  $\tau$  – ranges from 0 to 8. A MOTox score equal to 0 reflects a patient  $i$  who did not experience any kind of toxicity for the cycle  $k$  under analysis, i.e., a patient with all the toxicities equal to CTCAE-grade 0 at cycle  $k$ . Conversely, a MOTox score equal to 8 represents a subject  $i$  who experienced the highest level of toxicity burden for each type of toxic AE for the cycle  $k$  under analysis, i.e., a subject  $i$  with all toxicities equal to CTCAE-grade 4 at cycle  $k$ .

The global median MOTox  $\tau$  represents the median value of MOTox scores computed over all the patients in all the cycles, i.e., the median overall toxicity related to multiple AEs experienced by all the patients over the entire chemotherapy treatment. If required by the needs of the study, different median MOTox values breakdown by arms/regimens represent an easily-applicable alternative to a global  $\tau$  in order to study other treatment regimen/cancer types where different arms/regimens are characterized by significantly different overall toxicity burden.

Binary variable *high* –  $MOTox_{i,k}$  in Equation (5.2) indicates if patient  $i$  experienced high MOTox with respect to the global median MOTox value  $\tau$  at cycle  $k$ , i.e., it distinguishes patients with *low* ( $high-MOTox_{i,k} = 0$ ) or *high* ( $high-MOTox_{i,k} = 1$ ) overall toxicity burden over treatment.

### 5.2.2. Statistical analysis

A retrospective analysis to examine prognostic factors for binary high-MOTox scores over cycles was conducted. Baseline and treatment-related characteristics were examined. In particular, chemotherapy dose given at cycle  $k$  was analysed as percentage of *achieved chemotherapy dose up to cycle  $k$*  for each patient  $i$ , defined as the percentage of the cumulative drugs administrated up to cycle  $k$  divided by the cumulative drugs planned up to  $k$ :

$$\begin{aligned} p\delta_{i,k} &= \frac{\text{cumulative drugs administrative up to cycle } k}{\text{cumulative drugs planned up to cycle } k} \cdot 100\% \\ &= \frac{\sum_{c=1}^k (DOX_{i,c} + CDDP_{i,c}) / sa_{i,c} [mg/m^2]}{175 [mg/m^2] \cdot k} \cdot 100\% \end{aligned} \quad (5.3)$$

where  $k \in \{1, \dots, 6\}$  is the cycle index,  $sa$  is patient's surface area in  $m^2$ ,  $DOX$  and  $CDDP$  are the administrated  $mg$  of doxorubicin and cisplatin, respectively. A two-sided significance level of 5% was adopted. R software [161] was used for the analyses.

Data on non-haematological toxicity were not available for 1.25% of measurements, which were considered as CTCAE 0-grade according to clinical indication. For treatment-related missing values (i.e., histologic response, biochemical and haematological markers), missing values were imputed using multiple imputations by chained equations algorithm [209].

At each cycle, the impact of factors on high overall toxicity (binary *high-MOTox*) was examined using multivariable logistic regression models and expressed by odds ratios (OR) [137]. An  $OR > 1.0$  indicates a greater risk of achieving a high overall toxicity in case of a 1-unit increase for numerical characteristics or compared to the baseline category for categorical ones. Covariates with more than 15% of missing values in the original data were not included in the multivariable models. A stepwise backward selection procedure was applied to select the best set of covariates at each cycle based on Akaike Information Criterion (AIC). Variance Inflation Factor (VIF) was also used to remove non-significant and highly collinear covariates. Predictive capacities of models were assessed by sensitivity and specificity metrics and Area Under the receiver operating characteristic Curve (AUC) [55].

## 5.3. Results

### 5.3.1. Non-haematological longitudinal Overall Toxicity scores

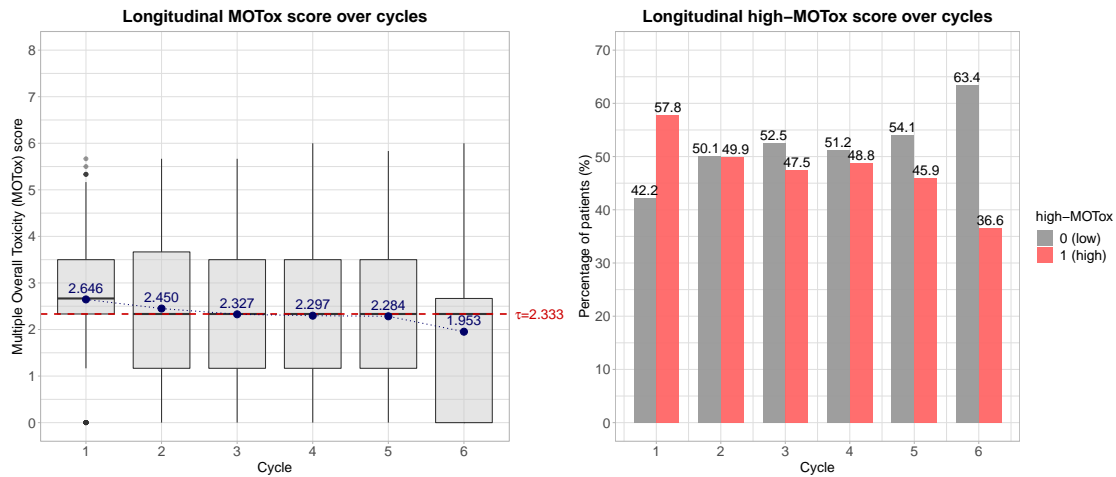
For each patient, non-haematological chemotherapy-induced toxicity related to nausea, mucositis, infection, neurological toxicity, cardiac toxicity, and ototoxicity, i.e., set  $\mathcal{T} = \{naus, oral, inf, car, oto, neur\}$ , were considered to compute the longitudinal MOTox scores over cycles for each patient, according to Equation (5.1) and Equation (5.2). MOTox scores (Figure 5.2 – left panel) ranged between 0 and 6 and the mean values (blue points) decreased over cycles from 2.626 (cycle 1) to 1.953 (cycle 6). The global median MOTox value  $\tau$ , i.e., the median value of overall toxicity over all the patients in all the cycles, was 2.333 (dashed red line). An example of longitudinal MOTox scores over cycles for five random patients from the study cohort is shown in Figure 5.3. The global mean MOTox value  $\tau$  is reported as solid black line. Different evolution patterns of longitudinal MOTox score over cycles are presented: increasing pattern (*orange*: patient A), decreasing pattern (*light blue*: patient B), isolated severe status (*violet*: patient C), low-values (*blue*: patient D) and high-values (*red*: patient E) over cycles.

To evaluate which regimens is characterized by high toxicity over cycles, Table 5.3 reports the means of MOTox scores at each cycle for patients allocated in *Reg-DI* and *Reg-C*, and respectively. In cycles 2-3, mean overall toxicity for patients in *Reg-DI* was higher than for those in *Reg-C* ( $p < 0.05$ ), whereas from cycle 4 the difference was not statistically significant. Figure 5.4 shows the mean values of each non-haematological toxicity along with 95% Bonferroni's confidence intervals over cycles, stratified by regimens. Each panel refers to a different type of toxicity: nausea/vomiting, mucositis, infection, cardiac toxicity, ototoxicity and neurological toxicity. The biggest contribution to the difference in the mean MOTox scores by regimes was given by mucositis, significantly higher in *Reg-DI* than in *Reg-C* at cycles 2 and 3.

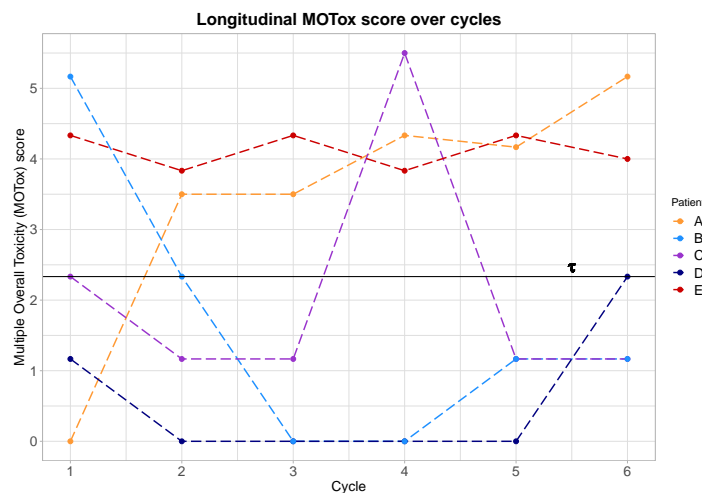
Median MOTox values by arms (*Reg-DI* or *Reg-C*)  $\tau_{DI}$  and  $\tau_C$  were both equal to 2.333. Therefore, the global MOTox median value  $\tau = 2.333$  was then used to compute the longitudinal dichotomous high-MOTox scores over cycles. Right panel in Figure 5.2 shows

## 5. A novel longitudinal method for quantifying multiple overall toxicity

the percentages of patients with high-MOTox, which decrease from 57.8% (218/377) at cycle 1 to 36.6% (138/377) at cycle 6. Association between chemotherapy regimens and high overall toxicity at cycles 2–3 ( $p < 0.05$ ) was found, supporting results shown in Table 5.3. At each cycle, high overall toxicity was strongly associated with *low/high* MOTox at previous cycles.



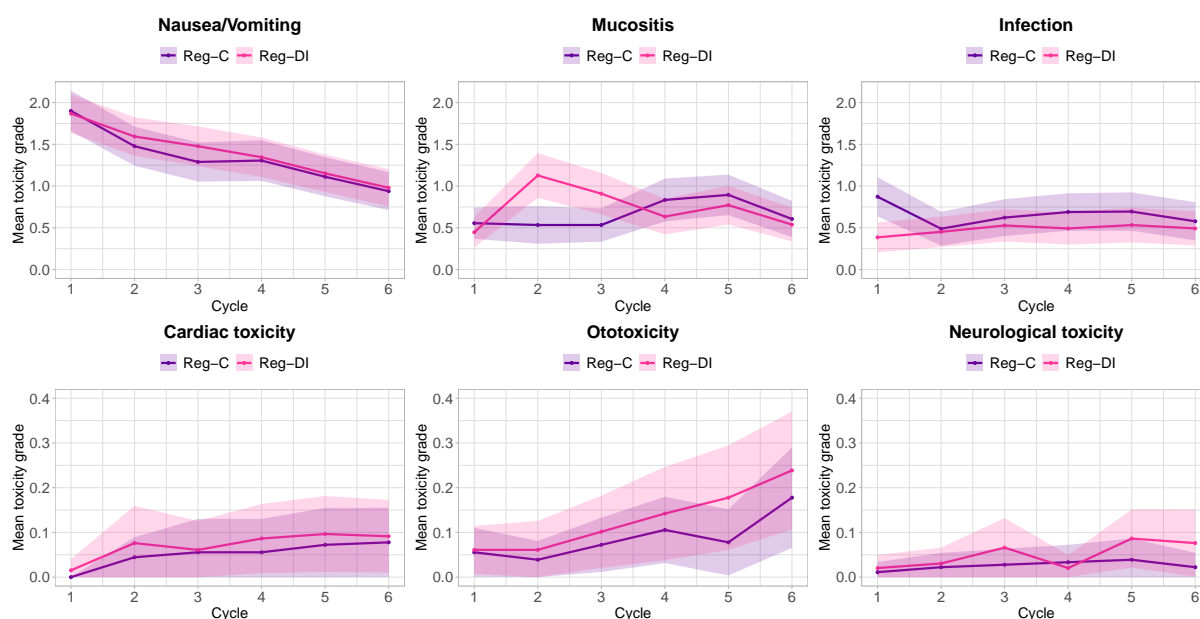
**Figure 5.2.** Left panel: Boxplots of longitudinal MOTox scores over cycles. Blue points refers to the mean MOTox values per cycle. Dashed red line refers to the global median MOTox value  $\tau = 2.333$ . Right panel: Bar-plots of longitudinal high-MOTox scores over cycles (grey: 0 or low; magenta: 1 or high).



**Figure 5.3.** Example of evolution of longitudinal Multiple Overall Toxicity (MOTox) scores over cycles for five patients from the study cohort. Solid black line refers to the global median MOTox value  $\tau = 2.333$ .

**Table 5.3.** Overall toxicity differences between Dose-Intense (DI) and Conventional (C) regimens.  $\overline{MOTox}_{DI}^k$  and  $\overline{MOTox}_C^k$  are the means of MOTox scores at cycle  $k$  for patients allocated in *Reg-DI* and *Reg-C*, respectively.

	Cycle 1	Cycle 2	Cycle 3	Cycle 4	Cycle 5	Cycle 6
$\overline{MOTox}_{DI}^k$	2.552	2.653	2.488	2.240	2.261	1.920
$\overline{MOTox}_C^k$	2.782	2.229	2.150	2.359	2.309	1.989
p-value of test	0.045	0.003	0.018	0.437	0.737	0.657



**Figure 5.4.** Mean value of chemotherapy-induced toxicity during cycles along with 95% Bonferroni confidence intervals, stratified by the regimens (purple: *Reg-C*; pink: *Reg-DI*). Each panel refers to a different type of toxicity: nausea/vomiting [top-left], mucositis [top-centre], infection [top-right], cardiac toxicity [bottom-left], ototoxicity [bottom-centre] and neurological toxicity [bottom-right].

### 5.3.2. Multivariable logistic regression models for high overall toxicity over cycles

The evolution of longitudinal binary *high-MOTox* score over cycles defined was analysed through multivariable logistic regression models, using a cycle-by-cycle approach. Starting from the second cycle, each logistic regression modelled the binary dependent variable *high-MOTox* at the of the cycle in terms of patient's characteristics and previous toxicity levels. Baseline and treatment-related information with less than 15% of missing values in the original dataset were considered as possible prognostic factors for toxicity. In particular, among haematological and biochemical factors, measurements of white blood count (WBC), neutrophils (N), platelets (PLT), alkaline phosphatase (ALP) and calcium (Ca) were considered before the beginning of each cycle (i.e., the administration of the course). Only WBC values were considered at the planned nadir of each cycle, due to the high percentage of missing values (>15%) for other blood counts. Due to the skewed nature of biomarkers distributions, haematological and biochemical factors were included in the models as difference between the logarithmic measure and the logarithmic value measured at randomization. Neutrophils-Platelets Score (NPS), a three-level systemic inflammation-based score (*good*:  $N \leq 7.5 \times 10^9/L$  and  $PLT \leq 400 \times 10^9/L$ ; *intermediate*:  $N > 7.5 \times 10^9/L$  or  $PLT > 400 \times 10^9/L$ ; *poor*:  $N > 7.5 \times 10^9/L$  and  $PLT > 400 \times 10^9/L$ ) [127], and Neutrophils-White blood count Ratio (NWR, i.e., the neutrophils count divided by the white blood cell count) were also considered. For each model, multicollinear variables with VIF greater than 5 were removed. Then, stepwise backward procedures were used to select covariates according to AIC. The selected models were fitted on the whole dataset.

Table 5.4 shows estimated Odds Ratios (ORs) along with 95% Confidence Intervals (CIs)



## 5. A novel longitudinal method for quantifying multiple overall toxicity

and overall performances (i.e., specificity, sensitivity and AUC) of each logistic regression model. All the models have similar overall performances: sensitivity and specificity values ranged between 0.66 and 0.77; AUCs were between 0.72 and 0.79. No sex effect was found. In cycle 2 and 3, higher percentage of achieved chemotherapy dose is associated to the risk of high toxicity, especially for patient in *Reg-DI* (cycle 2). Haematological factors were selected in each model. Both PLT before the administration of the course and WBC at nadir had a protective role on the risk of having high overall toxicity ( $OR < 1$ ). In particular, an increase in the dynamic difference between the logarithmic levels decreased the risk of high toxicity. Patients with previous *high-MOTox* had higher risk to experience again high overall toxicity with respect to patients with previous *high-MOTox* ( $OR > 1$ ), showing that *high-MOTox* conditions during previous cycles were risk factors for the occurrence of *high-MOTox* at the current cycle. In particular, toxicity information related to different previous cycles were selected and statistically significant in the final models, meaning that patient's global history – and not only the last condition – had impact on his/her current *low/high* overall toxicity burden.

The performed analyses were finally used to develop a demo webapp available at <http://osteowebapp.prod.s3-website.eu-central-1.amazonaws.com/>. The demo shows how the multivariable models developed to predict high overall toxicity index at each cycle could be used as a support tool for clinical decision making. The webapp is presented in Appendix B.1.

### 5.4. Final remarks

Due to the presence of multiple types of Adverse Events (AEs) with different levels of toxicity burden, to study the overall toxicity progression during chemotherapy is a difficult problem in cancer research. The development of statistical methods able to deal with the complexity of longitudinal chemotherapy data and to provide a methodology to use the information of AEs data into a score of overall risk is necessary and of clinical relevance.

This chapter explored the evolution of chemotherapy-induced toxicity over treatment in patients with osteosarcoma. First, a novel approach to analyse longitudinal chemotherapy data was discussed, the cycle-dependent longitudinal mean-max Multiple Overall Toxicity (MOTox) score over therapy. Starting from recorded CTCAE grades, the MOTox score summarised the occurrence of repeated AEs allowing to (i) describe the overall toxicity burden, (ii) consider the most severe collateral effect, and (iii) incorporate the time-component of treatment cycles. Results showed that the inclusion of worst-graded events, multiple lower-grade chronic toxicities, and time-dimension related to chemotherapy cycles allowed to consider different evolutions of overall toxic levels over treatment. This approach investigates in more details the effect of AEs on patients' life compared to traditional methods (i.e., max-grade or max-time). The cycle-by-cycle longitudinal evolution of high overall toxicity was analysed using multivariable logistic regression models to predict binary high-MOTox at the end of the cycle in terms of previous toxicity levels

**Table 5.4.** Multivariable logistic regression model for each cycle  $k \in \{2, 3, 4, 5, 6\}$  with estimated Odds Ratios (ORs) along with their 95% Confidence Intervals (CIs).

Covariates	Cycle 2		Cycle 3		Cycle 4		Cycle 5		Cycle 6		
	ORs	95% CIs	ORs	95% CIs	ORs	95% CIs	ORs	95% CIs	ORs	95% CIs	
<b>Baseline</b>											
<i>Sex (male)</i>											
<i>Regimen (Reg-DI)</i>	2.379	[1.455; 3.889]					1.458	[0.912; 2.33]	1.548	[0.967; 2.478]	
<b>Treatment-related factors</b>											
Achieved dose (%)	1.112	[1.042; 1.187]	1.056	[1.008; 1.106]							
<i>WBC</i>									1.664	[0.978; 2.831]	
<i>WBC at nadir</i>	0.778	[0.615; 0.983]			0.701	[0.569; 0.864]			0.637	[0.499; 0.813]	
<i>PLT</i>			0.535	[0.375; 0.765]	0.625	[0.392; 0.996]				0.629	[0.424; 0.932]
<i>NWR</i>					0.367	[0.094; 1.436]					
<b>Previous toxicities</b>											
<i>High MOTox (k = 1)</i>	4.439	[2.788; 7.070]	1.561	[0.968; 2.516]	1.522	[0.953; 2.430]					
<i>High MOTox (k = 2)</i>			4.429	[2.666; 6.772]	1.569	[0.972; 2.532]			2.639	[1.664; 4.186]	
<i>High MOTox (k = 3)</i>					2.701	[1.696; 4.304]			3.718	[2.346; 5.893]	
<i>High MOTox (k = 4)</i>										2.542	[1.523; 4.244]
<i>High MOTox (k = 5)</i>										3.341	[2.001; 5.580]
<b>Sensitivity</b>		0.681		0.704		0.674		0.699		0.717	
<b>Specificity</b>		0.667		0.661		0.715		0.701		0.766	
<b>AUC [95% CI]</b>		0.733 [0.683; 0.784]		0.743 [0.694; 0.793]		0.728 [0.677; 0.780]		0.756 [0.707; 0.805]		0.787 [0.737; 0.837]	

WBC = white blood count; PLT = platelets; NWR = neutrophils-white blood count ratio; MOTox = multiple overall toxicity.

When not specified, haematological factors were computed before the administration of the course.

WBC and PLT are included in the models as difference between the current logarithmic measure and the logarithmic value measured at randomization.

## 5. A novel longitudinal method for quantifying multiple overall toxicity

and patient's characteristics. At each cycle, previous toxicity levels were selected: high-MOTox during previous cycles were risk factors for the occurrence of high-MOTox at the next cycle. The highest impact on the risk was observed for the last available toxic condition. Patient's history of toxicity played a fundamental role in the risk of high overall toxicity burden during cycles and, consequently, on patient's health status during the therapy. This analysis also suggested that the Conventional Regimen might be preferred to the Dose-Intense in terms of life conditions during the first half of the therapy (i.e., up to the third cycle): mean MOTox values in *Reg-DI* were statistically higher than in *Reg-C* during cycles 2-3 and *Reg-DI* was a risk factor for the occurrence of high-MOTox at the end of the second cycle. However, in terms of survival, a beneficial effect of low level (grade 1-2) platelet and nausea/vomiting toxicity and more severe (grade 3-4) mucositis on survival in osteosarcomas was previously shown [172]. Appraisal of the experienced toxicity against survival encourages the genetic exploration of the individual sensitivity to both adverse effects as well as the sensitivity of the tumour to chemotherapy.

Different statistical and machine learning methods for high/low binary classification were considered, among others support vector machines or ensemble methods (e.g., random forests or XGBoost). More complex methods showed no significant improvements in terms of predictive performances with respect to logistic regression models. Therefore, the choice was driven by the clinical interpretability offered by the cycle-by-cycle logistic regression approach.

The presented MOTox and binary high-MOTox scores can be used to (i) describe patient's response to therapy over cycles, (ii) predict the upcoming overall toxicity level given patient's history and (iii) support clinical decisions, trying to reduce the impact of therapies in terms of toxic AEs. Provided that longitudinal CTCAE-graded toxicity data are available from drug administrations, the new approach is a flexible procedure that can be adapted and applied to other cancer studies. The possible generalization to many different settings, added to a cooperation with medical staff, could lead to improvements in the definition of useful tools for health care assessment and treatment planning. As shown in the demo webapp presented in Appendix B.1, once validated, the multivariable models could be used to set up a support tool to predict high overall toxicity at the end of each cycle. This would allow to monitor patient's toxic burden during treatment and to inform dose reductions or dose delays to make treatment more tolerable.

This retrospective exploratory analysis comes with some challenges and limitations. Although the toxicity data were recorded using the standardised CTCAE grading system, heterogeneity in assessing non-haematological toxicity is present in the data, especially considering that MRC BO06 RCT is limited to a young population with a rare tumour. The analysis was performed on a single RCT in osteosarcoma, where only non-haematological toxicities were recorded according to CTCAE. Other factors such as nephrotoxicity, lymphocytes count, tumour size, CTCAE-graded haematological toxicities or quality-of-life were not collected. Although over the last twenty years the main chemotherapy protocol has been used, some aspects of osteosarcoma treatment and supportive care have changed from current measures [140], such as the prophylaxis of nausea and vomiting. Such changes are not always easily identifiable and are difficult to account

for in retrospective analyses [140]. Finally, this work focused on the quantification and evolution of overall toxicity in patients who completed all 6 cycles of chemotherapy treatment. This choice was due to a specific research question. However, this may lead to bias selection due to the exclusion of patients who may have had high toxicity levels as the reason for treatment discontinuation. Since the definition of the MOTox score is general, it can be computed also for those excluded patients, but alternative statistical methods to multivariable logistic models must be developed to also take into account therapy discontinuation. In fact, subsequent analyses should include patients who have discontinued treatment to better understand if MOTox is a potential measure of treatment tolerability and if may be associated with treatment discontinuation.

External validation is needed to evaluate the application of the novel score in order to guide prospective treatment decisions in clinical practice, both for osteosarcoma and for other types of cancer. On one hand, integration with data from other osteosarcoma studies could help in further investigating the performance of the models and in examining whether the analysis should be integrated with more information on toxicity or other potential predictors. On the other hand, to apply the developed procedure to the clinical decision-making process in different treatment regimen/cancer types, the multivariate methods need to be tailored according to each specific study.

This work opens doors to many further developments, both in the field of statistical methodology and in cancer research. From a clinical point of view, the interest may lie in identifying patients with extremely high or extremely low overall toxicity with respect to intermediate toxic conditions. As consequences multiple MOTox categories related to different levels of overall toxicity (e.g., extremely-high/high/intermediate/low/extremely-low MOTox) are defined. Thresholds to establish the MOTox ranges for the different categories needs to be created. This is not a trivial task which requires a proper external validation. Furthermore, the comparison between the MOTox score and Quality-Of-Life (QOL) represents a challenging area of investigation in clinical research. MRC BO06 trial did not collect QOL data, but it would be of interest to evaluate MOTox in the context of rigorously collected Health-related QOL (HrQOL) or Patient-Reported Outcome (PRO) data to investigate the role of the developed tool in better understanding treatment tolerability. Therefore, future analyses must focus on data where QOL is properly measured and reported. From a statistical point of view, (i) the CTCAE-grades of toxicity could be analysed in greater depth through an appropriate longitudinal approach to categorical data, and (ii) an adequate modelling of the intricate mechanism between toxicity, chemotherapy dose and survival, which is still lacking in the medical literature, represents a major challenge of clinical relevance. Due to the complexity of the problem, both aspects are not straightforward and ask for the developments of new methodologies, as we will see in Chapters 6 and Chapter 7, respectively.

In summary, this chapter introduced a novel longitudinal method to explore and quantify AEs experienced by patients during cancer treatment. Preliminary results from the retrospective analysis of MRC BO06 RCT showed that longitudinal methods should be considered in future analyses of cancer trials, since they could lead to new insights into chemotherapy-induced toxicity compared to traditional approaches. For this reason, in

## 5. *A novel longitudinal method for quantifying multiple overall toxicity*

the next chapter we develop a new taxonomy based on latent Markov models [22] and compositional data [6] to model the evolution of latent overall toxicity burden on the basis of nominal CTCAE-grades observed over cycles.

## B. Appendix to Chapter 5

### B.1. Demo OsteoWebApp

The demo *OsteoWebApp* displays how the novel MOTox approach can become a useful tool for health care assessment and cancer treatment planning. In particular, it shows how the multivariable models to predict high overall toxicity at the end of each cycle developed in Chapter 5 could be used as a support tool for clinical decision making. It is available at: <http://osteowebapp.prod.s3-website.eu-central-1.amazonaws.com/>.

The application is implemented through Amazon Web Services (<https://aws.amazon.com/it/>) tools and executes the R [161] code related to the models in Table 5.4. Thanks to the intuitive interface, the webapp is easy to use and complete in the information it provides.

An example of the user interface, showing the inputs and results for model related to cycle 2, is reported in Figure 5.5. The top bar shows the cycle of chemotherapy of interest. The main form asks a series of information, depending on the variables selected for each cycle. The “*Predict Toxicity Index*” button in blue allows to get the results of the prediction, which are provided in terms of probability of develop a high overall toxicity level. Sensitivity and specificity of each model are also reported. Example in Figure 5.5 shows that a patient in *Reg-DI* with *high-MOTox* at cycle 1, a cumulative administrated dose of  $350 \text{ mg/m}^2$  (which corresponds to a 100% of achieved dose), WBC values of  $7.65 [\times 10^9/L]$  at randomization and of  $3.9 [\times 10^9/L]$  at nadir has 73.5% probability to be in *high-MOTox* status at the end of cycle 2.

Figure 5.5. Example of user interface for *OsteoWebApp*.



## CHAPTER 6

---

# Modelling longitudinal profiles of latent probability and relative risk via latent Markov models and compositional data

This chapter has been extracted and extended from M. Spreafico, F. Ieva and M. Fiocco “Longitudinal Latent Overall Toxicity (LOTox) profiles in osteosarcoma: a new taxonomy based on latent Markov models” in *arXiv*, 2107.12863, 2021 [191]. *[Submitted]*

Since patients may have multiple toxic Adverse Events (AEs) with different levels of severity, identifying the actual extent of toxic burden and investigating the evolution of patient’s overall toxicity status during treatment represent challenging problems in cancer research, as explained in Chapter 5. No standard method is available for analysing AEs due to the complexity of longitudinal chemotherapy data. Toxicity data are usually recorded as nominal grades of AEs severity [204] according to the Common Terminology Criteria for Adverse Events (CTCAE) [208], and analysed as summary indexes over the whole treatment period [28, 172, 205, 117, 140, 184, 199, 198, 226, 35], discarding substantial amount of information. As neglecting the time component may give an inaccurate depiction of toxicity, alternative methods for a longitudinal analysis of AEs have been proposed [205, 198, 200, 84, 190], such as the longitudinal MOTox procedure introduced in Chapter 5. These approaches are not suitable for the nominal CTCAE grades still they provide more insights into treatment-related toxicity, suggesting that longitudinal methods should become routine in future analyses of cancer trials. Models to deal with both longitudinal and categorical aspects of toxicity levels progression are then necessary, still not well developed.

Longitudinal data are often of interest in a wide range of research fields, such as social, economic and behavioural sciences, education or public health. In many applications involving longitudinal data, the interest lies in analysing the evolution of a latent characteristic of a group of individuals over time, rather than in studying their observed attributes [23]. The phenomenon which affects the distribution of the response variables that are relevant for the problem under consideration may not be directly observable. In a clinical context, this latent characteristic may reflect patients’ quality-of-life and could contain valuable information related to patient’s health status and disease progression.



In the statistical literature many models have been proposed for the analysis of longitudinal data; for a concise review see [57]. For longitudinal categorical data, where the interest is in describing individual changes with respect to a latent status, Latent Markov (LM) models can be used [216, 22]. These models study the evolution of an individual characteristic of interest, when it is not directly observable. The idea behind a LM model is that the latent process fully explains the observable behaviour of a subject, assuming that the response variables are conditionally independent given the latent process. The latent process follows a Markov chain with a finite number of states, which represent different conditions of the latent characteristic of interest. LM models can also account for the effect of observable covariates, serial dependence between observations, measurement errors, or unobservable heterogeneity. For a detailed overview on LM models see [22, 23].

Motivated by the need to improve methods for summarising and quantifying the overall toxicity level and its evolution during treatment, in this chapter a novel procedure based on LM models for longitudinal toxicity data is proposed. The latent status of interest is the Latent Overall Toxicity (LOTox) condition of a patient, which affects the distribution of the observed categorical toxic grades measured over treatment. The proposed approach aims at identifying different latent states of overall toxicity burden (*LOTox states*) and investigating how patients move between states during chemotherapy treatment.

A LM model for longitudinal toxicity data assumes that at each time occasion for each patient a vector of probabilities of being in the various LOTox states is given. Since the probability elements of each vector are non-negative coordinates whose sum is one, these vectors are naturally confined to a suitably dimensioned simplex, thus being *Compositional Data* (CoDa) or *compositions*. In statistics, CoDa are quantitative descriptions of the parts of some whole, carrying relative information. In this context, Aitchison (1986) [6] developed a methodology based on log-ratio transformations of CoDa, which nowadays represent the mainstream approach in the analysis of compositions formed by probabilities or percentages. Among the developed transformations, the *additive log-ratios* consider a specific reference part in contrast with all the others. In this chapter, this approach is exploited to compare over time a reference “good” overall toxicity condition (i.e., the LOTox state characterized by the lowest toxicity burden) in contrast with all the other LOTox states, characterized by worsening overall toxicity. In this way, the dynamic risk of experiencing “worse” overall toxicity statuses relative to a “good” toxic condition over time is investigated.

Three are the main novelties presented in this work: (i) the introduction of a new method based on LM models to summarize and quantify multiple AEs and their evolution during treatment, where both longitudinal and categorical aspects of the observed toxic levels are included in the model; (ii) the identification of groups of patients with a common distribution for the observed toxic categories, and thus a similar overall toxicity burden; (iii) the reconstruction of personalized *longitudinal LOTox profiles*, which represent the probability over time of being in a specific LOTox state or the relative risk with respect to a reference “good” toxic condition, to study the individual overall toxic risk evolution during treatment for each subject. The proposed approach is applied to osteosarcoma treatment

to provide novel techniques which could support clinicians in planning new protocols and guidelines for childhood cancer therapy. Provided that longitudinal CTCAE-graded toxicity data are available, the developed procedure is a flexible approach that can be adapted and applied to other cancer studies.

The rest of the chapter is organized as follows. Statistical methods are introduced in Section 6.1. Results for MRC BO06/EORTC 80931 Randomized Controlled Trial data are presented in Section 6.2. Section 6.3 ends with a discussion of strengths and limitations of the proposed approach, identifying some possible developments for future research.

## 6.1. Statistical Methods

Motivations for a latent Markov approach for treating the longitudinal toxicity data are discussed in Section 6.1.1. Mathematical details are provided in Section 6.1.2. Model selection procedure and longitudinal profiles are presented in Sections 6.1.3 and 6.1.4, respectively.

### 6.1.1. Motivations for latent Markov models for longitudinal toxicity data

LM models are statistical methods employed for the analysis of longitudinal (categorical) data specifically designed to study the evolution of an individual characteristic of interest, when it is not directly observable [216, 22]. A LM approach for longitudinal toxicity data assumes the existence of a latent process representing the “true” LOTox status, which affects the distribution of the response variables, in our case the observed toxicities. Two main motivations justify the use of LM models to quantify the toxic risk in cancer studies: (i) account for *measurement errors* in the observed toxicity variables, and (ii) identify different *LOTox sub-populations* (i.e., the latent states) in the global population (i.e., the patients’ cohort) and their changes over time.

Since therapy protocol is adapted at each cycle depending on patient’s reaction to treatment, it is reasonable to assume that the latent variables follow a first-order Markov chain, so that the “true” level of overall toxicity at a given cycle is influenced only by the previous level. Non-haematological toxicities (see Section 5.1.1) do not depend directly on each other as they relate to different systems and functions of the human body (i.e., nausea/vomiting is part of the stomach-gastrointestinal system, infections of the immune system, oral mucositis of the mouth-gastrointestinal system, cardiotoxicity of the cardiovascular system, ototoxicity of the auditory-sensory system and neurotoxicity of the nervous system). Therefore, the response toxicity variables can be assumed conditionally independent, as each observed response is expected to depend only on the corresponding “true” LOTox level.

In this context, a LM model may be seen as an extension of the latent class model [42], where patients are allowed to move between latent states during the observation period. LM models for longitudinal toxicity data are characterized by several parameters: the initial probability of each LOTox state, the transition probabilities among different states over chemotherapy cycles, and the conditional response probabilities given the latent variable. Individual covariates (if available) can be included in the latent model and may affect the initial and transition probabilities of the Markov chain [24], as explained in Section 6.1.2.

A LM approach is appropriate to both identify the actual overall toxicity burden and investigate its evolution during treatment for each patient. On one hand, patients that at a specific time result in the same sub-population are characterized by a common distribution for the observed toxic categories, and by a similar overall toxicity burden. On the other hand, individual dynamic changes among latent states allow to evaluate the LOTox evolution during treatment for each subject.

### 6.1.2. Latent Markov model with covariates

Let  $\mathcal{J}$  be the set of  $J = |\mathcal{J}|$  categorical response variables measured at each time  $t = 1, \dots, T$ . Denote by  $Y_{ij}^{(t)}$  the response variable  $j \in \{1, \dots, J\}$  for subject  $i \in \{1, \dots, n\}$  at time  $t$ , with set of categories  $\mathcal{C}_j$  coded from 0 to  $c_j - 1$ . Let  $\mathbf{Y}_i^{(t)} = (Y_{i1}^{(t)}, \dots, Y_{iJ}^{(t)})$  denote the observed multivariate response vector at time  $t$  for patient  $i$  and  $\tilde{\mathbf{Y}}_i = (\mathbf{Y}_i^{(1)}, \dots, \mathbf{Y}_i^{(T)})$  be the corresponding complete response vector. Denote by  $\tilde{\mathbf{X}}_i = (\mathbf{X}_i^{(1)}, \dots, \mathbf{X}_i^{(T)})$  the complete vector of individual covariates, where elements  $\mathbf{X}_i^{(t)} = (\mathbf{S}_i, \mathbf{Z}_i^{(t)})$  are the vectors of time-fixed  $\mathbf{S}_i$  and time-varying  $\mathbf{Z}_i^{(t)}$  covariates for subject  $i$  at occasion  $t$ . The general LM model assumes the existence of a latent process  $\mathbf{U}_i = (U_i^{(1)}, \dots, U_i^{(T)})$  which affects the distribution of the response variables  $\tilde{\mathbf{Y}}_i$ . The latent process follows a first-order Markov chain with state space  $\{1, \dots, k\}$ , where  $k$  is the total number of *latent states*. LM models usually assume that the response vectors  $\mathbf{Y}_i^{(1)}, \dots, \mathbf{Y}_i^{(T)}$  are conditionally independent given the latent process  $\mathbf{U}_i$  (*local independence of the response vectors*) and that the elements  $Y_{ij}^{(t)}$  are conditionally independent given  $U_i^{(t)}$  (*conditional independence of elements*). The motivation of these assumptions is that the latent process fully explains the observable behaviour of a subject, as explained in Section 6.1.1.

LM models are made by two components: the *measurement model* concerns the conditional distribution of the response variables given the latent process, and the *latent model* is related to the distribution of the latent process (i.e., initial and transition probabilities). The latent process represents an individual characteristic of interest that is not directly observable that may evolve over time, also depending on observable covariates. The main research interest hence lies in modelling the latent process and the effect of covariates on its dynamic. LM models where both the initial and the transition probabilities of the

latent process may depend on covariates is considered. Three different sets of probabilities (i.e., parameters) can be defined.

- *Conditional response probability* (or item-response probability)  $\phi_{jy|u}^{(t)}$  is the probability of observing a response  $y$  for variable  $j$  at time  $t$ , given the latent status  $u \in \{1, \dots, k\}$ :

$$\mathrm{P}\left(Y_{ij}^{(t)} = y | U_i^{(t)} = u\right) = \phi_{jy|u}^{(t)} \quad j = 1, \dots, J \quad y = 0, \dots, c_j - 1.$$

To ensure that the interpretation of the latent states remains constant over time, conditional response probabilities are assumed time-homogeneous, i.e.,  $\phi_{jy|u}^{(t)} = \phi_{jy|u}$   $\forall t = 1, \dots, T$ . Given the estimated  $\hat{\phi}_{jy|u}$ , the latent states can be characterized in terms of observed response categories.

- *Initial latent states prevalence*  $\delta_{u|\mathbf{x}_i^{(1)}}$  is the probability of membership in latent state  $u \in \{1, \dots, k\}$  at time  $t = 1$ , given the vector of covariates  $\mathbf{x}_i^{(1)}$  for individual  $i$ :

$$\mathrm{P}\left(U_i^{(1)} = u | \mathbf{X}_i^{(1)} = \mathbf{x}_i^{(1)}\right) = \delta_{u|\mathbf{x}_i^{(1)}}.$$

The estimated  $\hat{\delta}_{u|\mathbf{x}_i^{(1)}}$  may be interpreted as quantities proportional to the size of each latent state at the first time-occasion, given the covariates. A natural way to allow the initial probabilities of the LM chain to depend on individual covariates is a multinomial logit parametrization:

$$\log \frac{\mathrm{P}\left(U_i^{(1)} = u | \mathbf{X}_i^{(1)} = \mathbf{x}_i^{(1)}\right)}{\mathrm{P}\left(U_i^{(1)} = 1 | \mathbf{X}_i^{(1)} = \mathbf{x}_i^{(1)}\right)} = \log \frac{\delta_{u|\mathbf{x}_i^{(1)}}}{\delta_{1|\mathbf{x}_i^{(1)}}} = \beta_{0u} + \mathbf{x}_i^{(1)\top} \boldsymbol{\beta}_{1u} \quad (6.1)$$

where  $u = 2, \dots, k$  and  $\boldsymbol{\beta}_u = (\beta_{0u}, \boldsymbol{\beta}_{1u}^\top)^\top$  are the parameters vectors to be estimated.

- *Transition probability*  $\tau_{u|\bar{u}\mathbf{x}_i^{(t)}}^{(t)}$  is the probability of a transition to latent state  $u$  at time  $t$ , conditional on membership in latent state  $\bar{u}$  at time  $t-1$ , given the individual vector of covariates  $\mathbf{x}_i^{(t)}$  (if available):

$$\mathrm{P}\left(U_i^{(t)} = u | U_i^{(t-1)} = \bar{u}, \mathbf{X}_i^{(t)} = \mathbf{x}_i^{(t)}\right) = \tau_{u|\bar{u}\mathbf{x}_i^{(t)}}^{(t)}$$

where  $t = 2, \dots, T$  and  $u, \bar{u} = 1, \dots, k$ . The estimated  $\hat{\tau}_{u|\bar{u}\mathbf{x}_i^{(t)}}^{(t)}$  reflect changes or persistence in the various states over time, given the individual covariates whose effects can be modelled through a multinomial logit parametrization:

$$\log \frac{\mathrm{P}\left(U_i^{(t)} = u | U_i^{(t-1)} = \bar{u}, \mathbf{X}_i^{(t)} = \mathbf{x}_i^{(t)}\right)}{\mathrm{P}\left(U_i^{(t)} = \bar{u} | U_i^{(t-1)} = \bar{u}, \mathbf{X}_i^{(t)} = \mathbf{x}_i^{(t)}\right)} = \log \frac{\tau_{u|\bar{u}\mathbf{x}_i^{(t)}}^{(t)}}{\tau_{\bar{u}|\bar{u}\mathbf{x}_i^{(t)}}^{(t)}} = \gamma_{0\bar{u}u} + \mathbf{x}_i^{(t)\top} \boldsymbol{\gamma}_{1\bar{u}u} \quad (6.2)$$

for  $t = 2, \dots, T$  and  $\bar{u}, u = 1, \dots, k$  with  $\bar{u} \neq u$ .  $\boldsymbol{\gamma}_{\bar{u}u} = (\gamma_{0\bar{u}u}, \boldsymbol{\gamma}_{1\bar{u}u}^\top)^\top$  are the parameters vectors to be estimated.

Under the assumptions of *local* and *conditional independence*, the *manifest distribution* of the response variables (i.e., the conditional distribution of  $\widetilde{\mathbf{Y}}_i$  given  $\widetilde{\mathbf{X}}_i$ ) is given by:

$$\begin{aligned}
P(\widetilde{\mathbf{y}}_i \mid \widetilde{\mathbf{x}}_i) &= P\left(\widetilde{\mathbf{Y}}_i = \widetilde{\mathbf{y}}_i \mid \widetilde{\mathbf{X}}_i = \widetilde{\mathbf{x}}_i\right) = \\
&= \sum_{\mathbf{u}} P\left(\widetilde{\mathbf{Y}}_i = \widetilde{\mathbf{y}}_i \mid \widetilde{\mathbf{X}}_i = \widetilde{\mathbf{x}}_i, \mathbf{U}_i = \mathbf{u}\right) \times P\left(\mathbf{U}_i = \mathbf{u} \mid \widetilde{\mathbf{X}}_i = \widetilde{\mathbf{x}}_i\right) = \\
&= \sum_{\mathbf{u}} P\left(\mathbf{U}_i = \mathbf{u} \mid \widetilde{\mathbf{X}}_i = \widetilde{\mathbf{x}}_i\right) \times P\left(\widetilde{\mathbf{Y}}_i = \widetilde{\mathbf{y}}_i \mid \mathbf{U}_i = \mathbf{u}\right) = \\
&= \sum_{\mathbf{u}} \delta_{\mathbf{u}^{(1)} \mid \mathbf{x}_i^{(1)}} \prod_{t=2}^T \tau_{\mathbf{u}^{(t)} \mid \mathbf{u}^{(t-1)} \mathbf{x}_i^{(t)}} \times \prod_{t=1}^T \prod_{j=1}^J \phi_{j y_{ij}^{(t)} \mid \mathbf{u}^{(t)}}
\end{aligned} \tag{6.3}$$

where  $\mathbf{u} = (u^{(1)}, \dots, u^{(T)})$ . The vector  $\widetilde{\mathbf{y}}_i = (\mathbf{y}_i^{(1)}, \dots, \mathbf{y}_i^{(T)})$  is a realization of  $\widetilde{\mathbf{Y}}_i$ , where  $\mathbf{y}_i^{(t)}$  is an observation of  $\mathbf{Y}_i^{(t)}$  with elements  $y_{ij}^{(t)}$ . The vector  $\widetilde{\mathbf{x}}_i = (\mathbf{x}_i^{(1)}, \dots, \mathbf{x}_i^{(T)})$  is a realization of  $\widetilde{\mathbf{X}}_i$ , where  $\mathbf{x}_i^{(t)} = (\mathbf{s}_i, \mathbf{z}_i^{(t)})$  is an observation of  $\mathbf{X}_i^{(t)} = (\mathbf{S}_i, \mathbf{Z}_i^{(t)})$ .

Parameters estimation is performed maximizing the log-likelihood for a sample of  $n$  independent units, i.e.,  $\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \log P(\widetilde{\mathbf{y}}_i \mid \widetilde{\mathbf{x}}_i)$ , using an Expectation-Maximization algorithm ([22, 23, 25]). Deterministic and random initializations are implemented to reach the global maximum of  $\ell(\boldsymbol{\theta})$  and prevent identifiability issue related to the multimodality of the likelihood function.

### 6.1.3. Model selection

The choice of the final LM model for the application consists of two steps: (i) identification of the number of latent states  $k$ , and (ii) selection of the covariates to be included in the final model. When the number of latent states  $k$  can not be a priori defined based on clinical indications, it can be selected according different measures. Akaike information criterion (AIC) by [7] or the Bayesian information criterion (BIC) by [183], defined as

$$\text{AIC} = -2\hat{\ell} + 2g \quad \text{and} \quad \text{BIC} = -2\hat{\ell} + \log(n)g,$$

where  $\hat{\ell}$  is the maximum of the log-likelihood of the model of interest and  $g$  denotes the number of free parameters, are used. In particular, the smaller the values of the above criteria, the better the model represents the optimum compromise between goodness-of-fit and complexity. If the two criteria lead to selecting a different number of states, BIC is usually preferred [20, 26].

Basic LM models (i.e., LM models with time-heterogeneous transitions and no covariates –named M1) were fitted increasing the value of  $k$  from 1 to 10, and the number of latent states  $k$  was selected according to the minimum BIC. Once  $k$  was determined, a forward strategy was adopted to identify the covariates to be included in the final model. In particular, the smallest basic LM model with  $k$  latent states and time-homogeneous transitions (i.e., the LM model restricted to the case in which initial and transition probabilities are

parametrized by multinomial logit without covariates – named M2) was initially fitted and then the effect of each covariate on initial and/or transition probabilities (models M3-M12) was added. Only the covariates whose effect reduces the value of the BIC index were included in the final LM model.

#### 6.1.4. Longitudinal profiles: latent probability and relative risk

In LM models literature, once the model has been estimated, a decoding procedure is usually implemented to obtain a path prediction for each subject, i.e., finding the most likely sequence of latent states on the basis patient-specific observed data [22, 23]. However, this sequence represents a summary of how the entire latent process evolves over time, as it only provides information about the most-likely condition without giving details about other states (see Appendix C.1). To obtain more insights into the entire latent process and its evolution, longitudinal information related to each latent state can be reconstructed for each subject.

For each patient-specific observed data  $(\tilde{\mathbf{x}}_i, \tilde{\mathbf{y}}_i)$ , the Expectation-Maximization algorithm provides the *posterior* probabilities of variables  $U_i^{(t)}$

$$p_{iu}^{(t)} = \text{P} \left( U_i^{(t)} = u \mid \tilde{\mathbf{Y}}_i = \tilde{\mathbf{y}}_i, \tilde{\mathbf{X}}_i = \tilde{\mathbf{x}}_i \right) \quad t = 1, \dots, T \quad u \in \{1, \dots, k\}, \quad (6.4)$$

which can be estimated using recursions and involving the *manifest* distribution in Equation (6.3). For each latent state  $u \in \{1, \dots, k\}$ , probabilities in (6.4) can be used to reconstruct the *longitudinal latent probability profile* of the  $i$ -th subject, as follows:

$$\mathbf{p}_{iu} = \left\{ p_{iu}^{(t)} = \text{P} \left( U_i^{(t)} = u \mid \tilde{\mathbf{Y}}_i = \tilde{\mathbf{y}}_i, \tilde{\mathbf{X}}_i = \tilde{\mathbf{x}}_i \right), \quad t = 1, \dots, T \right\}. \quad (6.5)$$

Each profile  $\mathbf{p}_{iu}$  represents the probability over time  $t$  of being in latent state  $u$  for individual  $i$ , given the observed complete response  $\tilde{\mathbf{y}}_i$  and covariates  $\tilde{\mathbf{x}}_i$  (if available). Applying this procedure,  $k$  *longitudinal latent probability profiles* (one for each latent state) are obtained for each subject  $i$ , which can be expressed as a  $k \times T$  matrix

$$\mathbf{P}_i = \begin{bmatrix} \mathbf{p}_{i1} \\ \dots \\ \dots \\ \mathbf{p}_{ik} \end{bmatrix} = \begin{bmatrix} p_{i1}^{(1)} & p_{i1}^{(2)} & \dots & p_{i1}^{(T)} \\ \dots & & & \dots \\ \dots & & & \dots \\ p_{ik}^{(1)} & p_{ik}^{(2)} & \dots & p_{ik}^{(T)} \end{bmatrix} = \left[ \mathbf{p}_i^{(1)} \quad \mathbf{p}_i^{(2)} \quad \dots \quad \mathbf{p}_i^{(T)} \right]$$

with longitudinal latent probability profiles  $\mathbf{p}_{iu}$  as row-components. Columns of  $\mathbf{P}_i$  represent the vectors  $\mathbf{p}_i^{(t)}$  of posterior probabilities over time  $t = 1, \dots, T$  and can be seen as Compositional Data (CoDa) vectors belonging to the  $k$ -part Aitchison-Simplex  $\mathcal{S}^k$  [6], i.e.,

$$\mathbf{p}_i^{(t)} \in \mathcal{S}^k = \left\{ \mathbf{p} = [p_1, \dots, p_k] \in \mathbb{R}^k \mid p_u > 0, u = 1, \dots, k; \sum_{u=1}^k p_u = 1 \right\}. \quad (6.6)$$

Due to the sum constraint in Equation (6.6), elements  $p_{iu}^{(t)}$  of the composition  $\mathbf{p}_i^{(t)}$  are mutually dependent features which only carry relative information. In this context, Aitchison

(1986) [6] introduced a methodology based on log-ratio transformations of CoDa, which are required to remove constraints and eventually to map the composition to a real space, allowing standard statistical techniques to be applied to the transformed data. In most practical settings, the choice of transformation will depend on the preferred interpretation.

In the current framework, rather than considering the absolute individual elements  $p_{iu}^{(t)}$ , it is interesting to study the relative risk over time of being in a reference latent state  $u = R$  compared to all the other latent states. Among the transformations introduced by Aitchison (1986) [6], this can be done considering the *additive log-ratios* of each CoDa vector  $\mathbf{p}_i^{(t)}$ , as follows:

$$\begin{aligned} \text{alr}\left(\mathbf{p}_i^{(t)}\right) &= \left[ \log \frac{p_{i1}^{(t)}}{p_{iR}^{(t)}} \cdots \log \frac{p_{iR-1}^{(t)}}{p_{iR}^{(t)}} \log \frac{p_{iR+1}^{(t)}}{p_{iR}^{(t)}} \cdots \log \frac{p_{ik}^{(t)}}{p_{iR}^{(t)}} \right]^T \\ &= \left[ r_{i1}^{(t)} \cdots r_{iR-1}^{(t)} r_{iR+1}^{(t)} \cdots r_{ik}^{(t)} \right]^T \\ &= \mathbf{r}_i^{(t)} \in \mathbb{R}^{k-1} \end{aligned} \tag{6.7}$$

where  $R$  is the reference latent state which can be chosen arbitrary among  $\{1, \dots, k\}$ . Note that this transformation maps each bounded sample into a real space (alr:  $\mathcal{S}^k \rightarrow \mathbb{R}^{k-1}$ ) and if one of the  $p_{iu}^{(t)}$  elements is exactly zero, a zero-handling procedure is needed before applying the transformation. In that case, an easily applicable possibility would be to replace each zero with a small appropriate value, modifying the non-zero values of the relative composition in a multiplicative way in order to satisfy the sum constraint requirement. For further details see [135]. Applying this procedure to each compositions,  $k - 1$  *longitudinal relative risk profiles* (one for each non-reference state) are obtained for each subject  $i$ , given as a  $(k - 1) \times T$  matrix

$$\mathbf{R}_i = \begin{bmatrix} \mathbf{r}_i^{(1)} & \mathbf{r}_i^{(2)} & \cdots & \mathbf{r}_i^{(T)} \end{bmatrix} = \begin{bmatrix} \mathbf{r}_{i1} \\ \cdots \\ \mathbf{r}_{iR-1} \\ \mathbf{r}_{iR+1} \\ \cdots \\ \mathbf{r}_{ik} \end{bmatrix}$$

where column-element  $\mathbf{r}_i^{(t)}$  are given by Equation (6.7) and row-element  $\mathbf{r}_{iu}$  with  $u \neq R$  are the *longitudinal relative risk profile* of state  $u$  for subject  $i$

$$\mathbf{r}_{iu} = \left\{ r_{iu}^{(t)} = \log \frac{p_{iu}^{(t)}}{p_{iR}^{(t)}}, \quad u \neq R, t = 1, \dots, T \right\}. \tag{6.8}$$

Each profile  $\mathbf{r}_{iu}$  represents the relative risk (in logarithmic scale) over time  $t$  of being in latent state  $u \neq R$  with respect to the reference state  $R$  for individual  $i$ . Since this procedure is a transformation-based analysis, transformed elements  $r_{iu}^{(t)}$  must then be interpreted with respect to the chosen reference. A positive (negative) value  $r_{iu}^{(t)}$  at time  $t$  means that the risk for subject  $i$  of being in latent state  $u \neq R$  is  $\exp\left\{r_{iu}^{(t)}\right\}$  times higher (lower) than being in reference state  $R$ .

For the application discussed in this work, the *LOTox states* summarize different levels of overall toxicity burden, representing a proxy for patient’s quality of life. Therefore, for each patient  $i$ , longitudinal latent probability profile (6.5) represents the probability over time of being in the LOTox state  $u$  (i.e., the probability over time of developing an overall toxic burden quantified by state  $u$ ) given patient’s history: observed toxicity categories  $\tilde{\mathbf{y}}_i$  and personal characteristics  $\tilde{\mathbf{x}}_i$  over treatment. Once the LOTox states have been identified, it is reasonable to analyse and interpret the different results in relation to the state characterized by the lowest overall toxicity burden (i.e., “good” toxic condition), which is chosen as the reference  $R$ . In this way, the longitudinal relative risk profile (6.8) represents the risk of being in LOTox condition  $u \neq R$  compared to the lowest toxic status.

By reconstructing the *longitudinal LOTox profiles*, it is possible to (i) describe patient’s response to therapy over cycles, (ii) quantify the overall toxicity burden evolution over treatment cycles given patient’s history and (iii) investigate the individual dynamic changes among latent states, detecting differences in health status and quality of life among patients.

## 6.2. Data application

In childhood cancer research, the development of new evidence-based guidelines to support clinical decisions in tailored interventions for an effective management of adverse symptoms and treatments is still a key issue. In this section, the results obtained from the application of the proposed LM model to the MRC BO06/EORTC 80931 randomized clinical trial are reported. Analysing the evolution of toxicities in patients who have completed the treatment could lead to new insights into the progression and tolerance of toxic AEs during therapy. For these reasons, we focused on the same cohort analysed in Chapters 4 and 5 concerning the 377 patients who completed the entire chemotherapy protocol, finishing the sixth cycle within 180 days after randomisation without abnormal dosages (see Figure 4.3 in Chapter 4). Patient characteristics at randomization are shown in Table 4.1 in Chapter 4. Statistical analyses were performed in the R-software environment [161], using LMest package by [26]. R code for the current study is available at <https://github.com/mspreafico/B006-LOTox>.

### 6.2.1. Longitudinal toxicity data: item-response categories

During the trial treatment, case report forms were used to document across cycles all the information required by the MRC BO06/EORTC 80931 trial protocol for each patient (see Section 4.2.1 in Chapter 4). Non-haematological chemotherapy-induced toxicity for nausea/vomiting (*naus*), infection (*inf*), oral mucositis (*oral*), cardiac toxicity (*car*), ototoxicity (*oto*) and neurological toxicity (*neur*) were graded according to the CTCAE v3.0 [208], with grades ranging from 0 (none) to 4 (life-threatening) (see Table 5.2 in Chapter 5). Nausea/vomiting, infection and oral mucositis were classified as *generic*



toxicities since they represent common adverse events for chemotherapeutic treatments in general. Cardiac toxicity, ototoxicity and neurological toxicity, which could also cause irreversible conditions, were classified as *drug-specific* toxicities since they are related to the use of cisplatin or doxorubicin [8, 51].

Grades of chemotherapy-induced non-haematological toxicity over cycles recorded for the selected cohort have been reported in Figure 5.1 in Chapter 5. At each cycle, CTCAE-grade 4 for *generic* toxicities and CTCAE-grades  $\geq 2$  for *drug-specific* toxicities were reported in less than 5% of patients. Low-frequency classes were merged and toxic categories were represented according to the degree of severity or as present or not, depending on the type of toxicity as follows:

- the severity of the toxic event for *generic* toxicities: *none* (CTCAE-grade 0), *mild* (CTCAE-grade 1), *moderate* (CTCAE-grade 2), *severe* (CTCAE-grades 3 or 4);
- the absence or the presence of toxic event for *drug-specific* toxicities: *no* (CTACE-grade 0) and *yes* (CTACE-grades  $\geq 1$ ).

These categories identified for each toxicity constitute the item-response elements selected to model the latent process representing the “true” overall toxic status. Table 6.1 shows the observed frequencies (and percentages) of the selected categories for each toxicity over cycles for the final cohort. The observed responses for each patient are then given by the longitudinal toxic categories measured along the cycles, which are then used to evaluate the LOTox condition during treatment.

### 6.2.2. Latent Markov model for longitudinal toxicity data

For each cycle  $t = 1, \dots, 6$ , let  $\mathcal{J} = \{naus, inf, oral, car, oto, neur\}$  be the set of non-haematological toxicities, representing response variables  $Y_{ij}^{(t)}$ . The relative sets of response categories identified in the previous section were coded from 0 to  $c_j - 1$ , as follows:

$$\begin{aligned} \mathcal{C}_j &= \{0 : none, 1 : mild, 2 : moderate, 3 : severe\} \quad \text{for } generic \text{ toxicities } (j = 1, 2, 3), \\ \mathcal{C}_j &= \{0 : no, 1 : yes\} \quad \text{for } drug-specific \text{ toxicities } (j = 4, 5, 6). \end{aligned}$$

The procedure described in Section 6.1.3 was applied to first identify the number of latent states  $k$  and then select the covariates to be included in the final model. Age, gender and allocated regimen at randomization were considered as time-fixed covariates, while percentage of achieved chemotherapy dose up to cycle  $t$  (see Equation 5.3), white blood cell, neutrophils and platelets counts measured at each cycle were considered as time-varying ones. Results are shown in Table 6.2. The unrestricted LM model without covariates (M1) with the minimum BIC (16728.90) was obtained for  $k = 4$ , identifying a latent process with four *LOTox states*. Moreover, the basic model M2 with initial and transition probabilities parametrized by multinomial logit was preferable (BIC = 16512.16) to the unrestricted model M1 with the same number of latent states. Several

**Table 6.1.** Frequencies of toxic categories over the six cycles. For nausea, infection and mucositis ( $j = 1, 2, 3$ ), the set of toxic categories indicating the severity of the toxic event is defined as  $\mathcal{C}_j = \{none; mild; moderate; severe\}$ . For cardiotoxicity, ototoxicity and neurological toxicity ( $j = 4, 5, 6$ ), the set of toxic categories indicating the presence or the absence of the toxic event is defined as  $\mathcal{C}_j = \{no; yes\}$ .

Toxicity	Cycle 1	Cycle 2	Cycle 3	Cycle 4	Cycle 5	Cycle 6
<b>Nausea</b>						
<i>none</i>	57 (15.1%)	88 (23.3%)	115 (30.5%)	126 (33.4%)	146 (38.7%)	179 (47.5%)
<i>mild</i>	74 (19.6%)	87 (23.1%)	76 (20.2%)	72 (19.1%)	86 (22.8%)	74 (19.6%)
<i>moderate</i>	117 (31.1%)	117 (31.1%)	114 (30.2%)	113 (30.0%)	96 (25.5%)	87 (23.1%)
<i>severe</i>	129 (34.2%)	85 (22.5%)	72 (19.1%)	66 (17.5%)	49 (13.0%)	37 (9.8%)
<b>Infection</b>						
<i>none</i>	259 (68.7%)	287 (76.1%)	268 (71.1%)	265 (70.3%)	268 (71.1%)	286 (75.9%)
<i>mild</i>	30 (7.9%)	24 (6.4%)	26 (6.9%)	31 (8.2%)	23 (6.1%)	16 (4.3%)
<i>moderate</i>	64 (17.0%)	45 (11.9%)	61 (16.2%)	54 (14.3%)	52 (13.8%)	45 (11.9%)
<i>severe</i>	24 (6.4%)	21 (5.6%)	22 (5.8%)	27 (7.2%)	34 (9.0%)	30 (8.0%)
<b>Mucositis</b>						
<i>none</i>	265 (70.3%)	228 (60.5%)	234 (62.1%)	237 (62.9%)	214 (56.8%)	262 (69.5%)
<i>mild</i>	54 (14.3%)	46 (12.2%)	59 (15.6%)	52 (13.8%)	62 (16.4%)	44 (11.7%)
<i>moderate</i>	44 (11.7%)	54 (14.3%)	43 (11.4%)	55 (14.6%)	63 (16.7%)	50 (13.2%)
<i>severe</i>	14 (3.7%)	49 (13.0%)	41 (10.9%)	33 (8.7%)	38 (10.1%)	21 (5.6%)
<b>Cardiotoxicity</b>						
<i>no</i>	374 (99.2%)	361 (95.8%)	362 (96.0%)	359 (95.2%)	357 (94.7%)	355 (94.2%)
<i>yes</i>	3 (0.8%)	16 (4.2%)	15 (4.0%)	18 (4.8%)	20 (5.3%)	22 (5.8%)
<b>Ototoxicity</b>						
<i>no</i>	357 (94.7%)	361 (95.8%)	350 (92.8%)	342 (90.7%)	346 (91.8%)	326 (86.5%)
<i>yes</i>	20 (5.3%)	16 (4.2%)	27 (7.2%)	35 (9.3%)	31 (8.2%)	51 (13.5%)
<b>Neurological toxicity</b>						
<i>no</i>	371 (98.4%)	367 (97.3%)	362 (96.0%)	367 (97.3%)	356 (94.4%)	363 (96.3%)
<i>yes</i>	6 (1.6%)	10 (2.7%)	15 (4.0%)	10 (2.7%)	21 (5.6%)	14 (3.7%)

models (M3-M12) with four latent states, obtained from M2 adding covariates effect to initial and/or transition probabilities, were fitted. By comparing models M3-M12 with M2, age (centred with respect to the mean) at randomization was the only covariate leading to a significant improvement in terms of both BIC and AIC (M5). Model M5, whose path diagram for a given subject  $i$  is shown in Figure 6.1, was then selected as final model:

- initial probabilities were associated with patient's *age* at randomization and Equations (6.1) for a patient  $i$  became

$$\log \frac{\delta_{u|age_i}}{\delta_{1|age_i}} = \beta_{0u} + \beta_{1u} \cdot (age_i - 15) \quad u = 2, 3, 4; \quad (6.9)$$

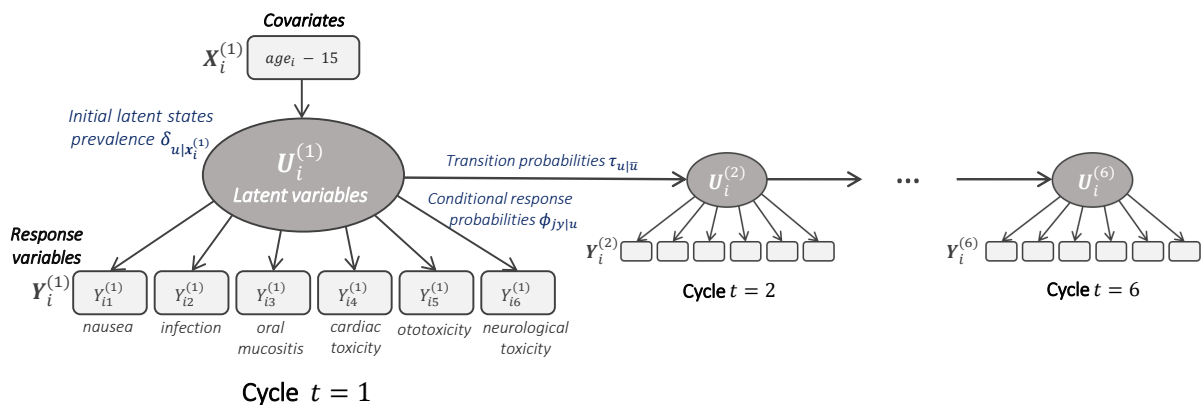
- transition probabilities were assumed time-homogeneous and Equations (6.2) became

$$\log \frac{\tau_{u|\bar{u}}}{\tau_{\bar{u}|\bar{u}}} = \gamma_{0\bar{u}u} \quad \bar{u}, u = 1, 2, 3, 4 \text{ with } \bar{u} \neq u. \quad (6.10)$$

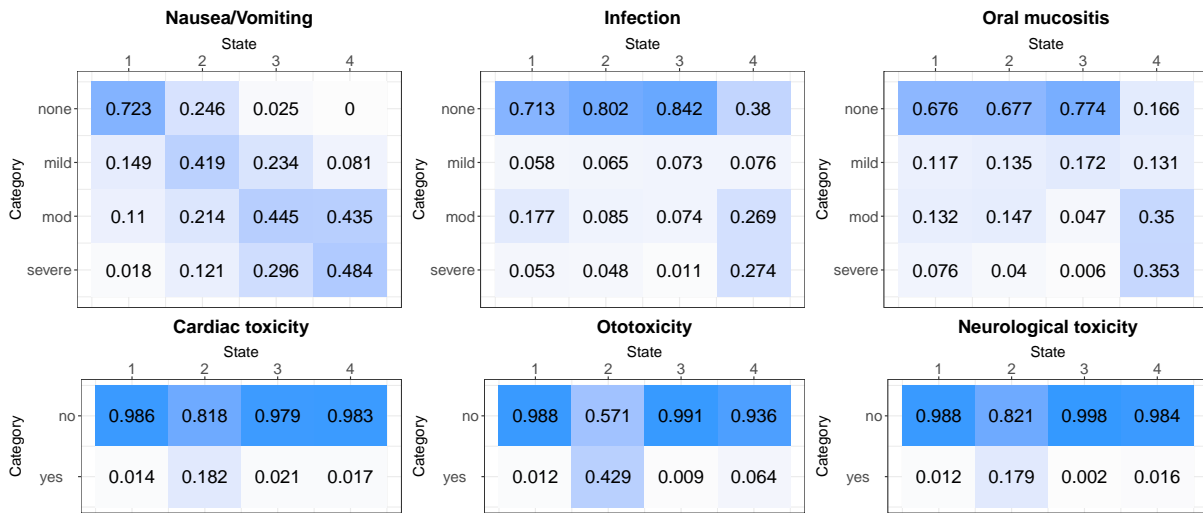
Figure 6.2 shows the estimated conditional response probabilities  $\hat{\phi}_{jy|u}$  for each type of non-haematological toxicity under the selected model M5, which can be used for interpreting the latent states. In each toxicity-panel, each column refers to a different latent

**Table 6.2.** Results for Latent Markov (LM) model selection for longitudinal toxicity data with different values of latent states  $k$  and different restrictions. The maximum log-likelihood of each model is denoted by  $\hat{\ell}$  and  $g$  is the number of free parameters. WBC, PLT and NEUT in models M10-12 refers to white blood cell, platelets and neutrophils counts, respectively.

Latent Markov (LM) model	$k$	$g$	$\hat{\ell}$	AIC	BIC
M1: Unrestricted LM model without covariates	1	18	-8794.91	17625.81	17696.59
	2	35	-8420.19	16910.38	17048.01
	3	68	-8216.99	16569.98	16837.37
	4	111	-8035.21	16292.42	16728.90
	5	164	-7902.59	16133.18	16778.07
	6	227	-7793.14	16040.29	16932.91
	7	300	-7688.12	15976.24	17155.91
	8	383	-7603.30	15972.61	17478.66
	9	476	-7530.49	16012.98	17884.73
	10	579	-7462.34	16082.68	18359.45
M2: Multinomial logit LM model without covariates	4	63	-8069.21	16264.43	16512.16
M3: M2 + regimen effect on initial prob.	4	66	-8065.49	16262.97	16522.50
M4: M2 + gender effect on initial prob.	4	66	-8061.73	16255.45	16514.98
M5: M2 + age effect on initial prob.	4	66	-8055.35	16242.69	16502.22
M6: M2 + regimen effect on transition prob.	4	75	-8063.37	16276.74	16571.66
M7: M2 + gender effect on transition prob.	4	75	-8060.33	16270.66	16565.58
M8: M2 + age effect on transition prob.	4	75	-8061.07	16272.14	16567.06
M9: M2 + time-var chemotherapy dose on both prob.	4	78	-8045.55	16247.10	16553.82
M10: M2 + time-var WBC count on both prob.	4	78	-8062.53	16281.07	16587.78
M11: M2 + time-var PLT count on both prob.	4	78	-8047.15	16250.30	16557.02
M12: M2 + time-var NEUT count on both prob.	4	78	-8062.67	16281.34	16588.05



**Figure 6.1.** Path diagram for a given subject  $i$  under the latent Markov model M5 with non-haematological toxicities as response variables, time-homogeneous transitions and  $age$  at randomization as covariate affecting the initial probabilities of the latent variables.



**Figure 6.2.** Estimated conditional response probabilities  $\hat{\phi}_{jy|u}$  for the final LM model in Figure 6.1. Each panel refers to a different toxicity  $j \in \mathcal{J} = \{1 : \text{naus}, 2 : \text{inf}, 3 : \text{oral}, 4 : \text{car}, 5 : \text{oto}, 6 : \text{neur}\}$ . Each row refers to a response categories  $y \in \{\text{none}; \text{mild}; \text{moderate}; \text{severe}\}$  for  $j = 1, 2, 3$  (*generic* toxicities) and  $y \in \{\text{no}; \text{yes}\}$  for  $j = 4, 5, 6$  (*drug-specific* toxicities). Each column refers to a latent states  $u \in \{1, 2, 3, 4\}$ .

state  $u \in \{1, 2, 3, 4\}$ . People in good conditions are allocated in state 1, since for all non-haematological toxicities the most probable category was the absence of the adverse event. State 2 seems to correspond to patients with non-severe nausea and it was the only state where *drug-specific* toxicities occurred with a relevant probability, especially for ototoxicity where  $\hat{\phi}_{51|2} = 0.429$ . State 3 seems to be characterized by patients undergoing only nausea or vomiting, mostly moderate or severe. In State 4 people with multiple *generic* toxicities - mostly severe or moderate - with the certainty of having nausea ( $\hat{\phi}_{10|4} = 0$ ) are present. Based on these results, the following LOTox states labelling were derived:

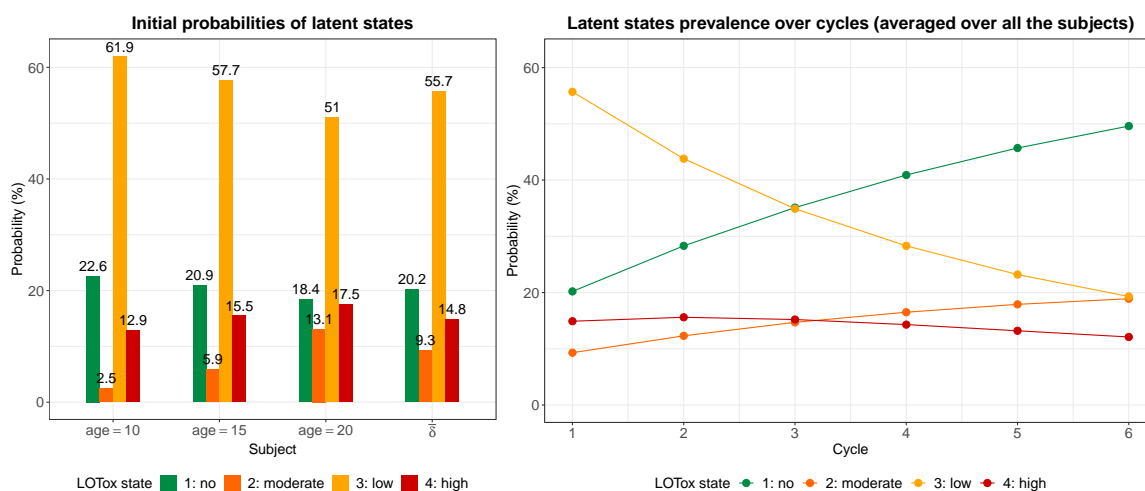
- State 1: quite good conditions (non-toxic)  $\rightarrow$  *no LOTox*
- State 2: non-severe nausea with possible *drug-specific* AEs  $\rightarrow$  *moderate LOTox*
- State 3: moderate/severe nausea/vomiting only  $\rightarrow$  *low LOTox* (limited to nausea)
- State 4: multiple severe/moderate *generic* toxicities  $\rightarrow$  *high LOTox*.

Note that the states numbering (from 1 to 4) does not correspond with the progressive severity of overall toxicity burden (from *no* to *high*).

Table 6.3 displays the estimated regression parameters  $\hat{\beta}_u = (\hat{\beta}_{0u}, \hat{\beta}_{1u})$  for the initial probabilities in Equation (6.9) and the estimated transition probabilities  $\hat{\tau}_{u|\bar{u}}$  in Equation (6.10). The estimated intercepts indicates that for 15-year patients the most prevalent state at cycle 1 was *low LOTox* state 3 (limited to nausea), followed by *no LOTox* state 1, *high LOTox* state 4 and *moderate LOTox* state 2. The estimates for *age* were all positive, indicating that older individuals reported a higher overall severity at the first cycle compared to younger patients. The estimated transition probabilities  $\hat{\tau}_{u|\bar{u}}$  shows a quite high persistence in the same state, especially for non-toxic state 1 and moderate

**Table 6.3.** Estimated regression parameters affecting the distribution of the initial probabilities in Equation (6.9) and estimated transition probabilities in Equation (6.10).

Regression parameters for initial probabilities				
	$u$	2	3	4
Intercept	$\hat{\beta}_{0u}$	-1.2679	1.0138	-0.3031
Age	$\hat{\beta}_{1u}$	0.1858	0.0014	0.0512
Transition probabilities from $\bar{u}$ to $u$ ( $\hat{\tau}_{u \bar{u}}$ )				
$\bar{u} \setminus u$	1	2	3	4
1	0.9674	0.0167	0.0032	0.0127
2	0.0525	0.9214	0.0245	0.0016
3	0.1070	0.0526	0.7581	0.0824
4	0.1555	0.0356	0.0868	0.7221

**Figure 6.3.** Left panel: estimated initial probabilities of latent states for patients aged 10, 15 and 20 years old and average  $\bar{\delta}$  of the initial probabilities over all the 377 subjects in the sample.Right panel: latent states prevalences over cycles  $t = 1, \dots, 6$  averaged over all the subjects.Different colours refer to different Latent Overall Toxicity (LOTox) state (green: *no LOTox state 1*; yellow: *low LOTox state 3*; orange: *moderate LOTox state 2*; red: *high LOTox state 4*).

state 2, where *drug-specific AEs* may also lead to permanent conditions (see Table 5.2 in Chapter 5). The highest transition probability was 15.6% and was observed from the *high LOTox state 4*, where the effects of *generic AEs* are reversible and temporary, to the first non-toxic state. Other transitions were observed from *high LOTox state 4* to nausea/vomiting only in state 3 (8.7%) and from *low LOTox state 3* (limited to nausea) to *no LOTox state 1* (10.7%) or *high LOTox state 4* (8.2%). The remaining transition probabilities were always lower than 8%.

Starting from these parameter estimates, Figure 6.3 (left panel) displays the estimated vectors of initial probabilities  $\hat{\delta}_i = (\hat{\delta}_{1|age_i}, \hat{\delta}_{2|age_i}, \hat{\delta}_{3|age_i}, \hat{\delta}_{4|age_i})$  for patients aged 10, 15 and 20 years old and the vector  $\bar{\delta} = (\bar{\delta}_1, \bar{\delta}_2, \bar{\delta}_3, \bar{\delta}_4) = (0.202, 0.093, 0.557, 0.148)$  obtained as average of vectors  $\hat{\delta}_i$  over all the 377 subjects in the sample. On average, at cycle 1 *low LOTox state 3* of subjects with nausea/vomiting only had the largest dimension (55.7%), followed by 20.2% of individuals for *no LOTox state 1*. *No* and *low LOTox states* together, representing the states with the lowest overall toxic severity, accounted

for more than 75% of the patients, whereas less than 25% belonged to the latent states corresponding to the worst toxic conditions (*moderate* and *high LOTox* states 2 and 4).

Right panel in Figure 6.3 shows the estimated average probability of each latent state at each time-occasion, i.e., the latent states prevalences averaged over all the subjects at each cycle. On average, the presence of *low* overall severity limited to nausea (state 3) decreased over cycles from 55.7% to 19.3% ( $t = 6$ ), whereas *no* and *moderate* overall toxicity (state 1 and 2, respectively) increased from 20.2% to 49.7% and from 9.2% to 18.9%. The presence in *high* overall severity (state 4) was rather stable over cycles ranging in 10.1%-15.6%, with peaks at cycles 2 and 3.

### 6.2.3. Longitudinal profiles of Latent Overall Toxicity

Once the parameters were estimated for the final LM model, the longitudinal latent probability profiles  $\mathbf{p}_{iu}$  were reconstructed for each patient  $i$  and latent state  $u$ , as explained in Section 6.1.4. In case of longitudinal toxicity data, profiles  $\mathbf{p}_{iu}$  in (6.5) are defined as *longitudinal Probability profiles of LOTox (P-LOTox)* since they represent the probability over cycles  $t = 1, 2, \dots, 6$  of being in the LOTox state  $u \in \{1, 2, 3, 4\}$  for each patient  $i$ , given the observed toxic categories over treatment and individual characteristics (i.e., the age at randomization).

Figure 6.4 shows the longitudinal P-LOTox profiles  $\mathbf{p}_{iu}$  for four patients  $i = \{A, B, C, D\}$  aged 15 years old and with different observed toxic categories over cycles, as reported in Table 6.4. Each panel refers to a different patient and displays the individual realisations of the latent process over cycles. Different patterns of overall toxicity evolution during treatment can be observed between subjects, based on patient-specific observed toxicity data. For example, right panel shows that at cycle 1 patient  $D$  had probabilities 79.6% of being in *low* LOTox state, 15.5% of having a non-toxic condition, 4.5% and 0.4% of *high* and *moderate* LOTox, respectively. The probabilities evolved over the cycles, as shown by the four profiles, ending with a 99.7% probability of being in quite good conditions at the end of treatment.

The lowest toxic burden is represented by the non-toxic state 1 of patients in good conditions, chosen as reference state ( $R = 1$ : *no* LOTox) to reconstruct the longitudinal latent relative risk profiles  $\mathbf{r}_{iu}$  for each patient  $i$  and latent state  $u \in \{2, 3, 4\}$ . In case of longitudinal toxicity data, profiles  $\mathbf{r}_{iu}$  in Equation (6.8) can be also called *longitudinal Relative Risk profiles of LOTox (RR-LOTox)*. They represent for each patient  $i$  the relative risk (in logarithmic scale) over cycles  $t = 1, 2, \dots, 6$  of being in the LOTox state  $u \in \{2, 3, 4\}$  rather than in the non-toxic state  $R = 1$ , given the observed toxic categories over treatment and individual characteristics.

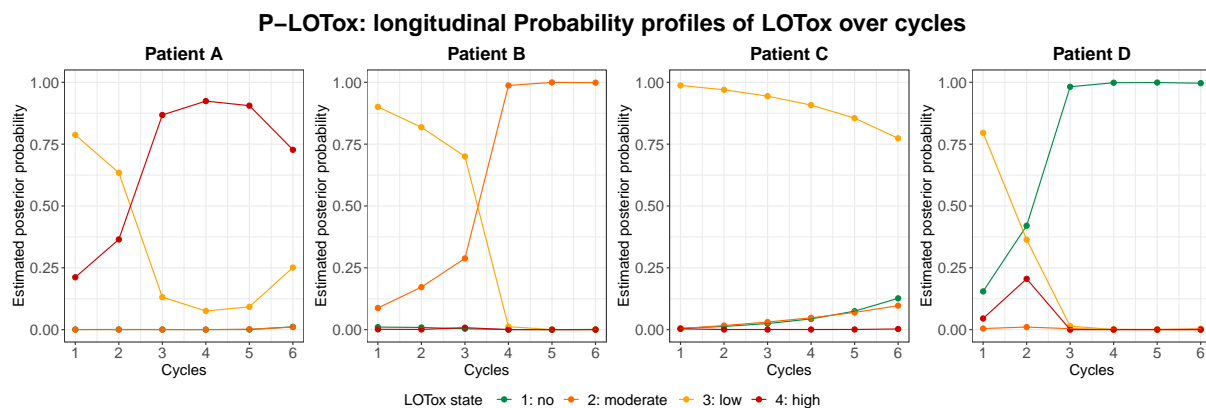
Figure 6.5 shows the longitudinal RR-LOTox profiles  $\mathbf{r}_{iu}$  for patients  $i = \{A, B, C, D\}$ . Different toxic risk progressions during treatment can be observed among patients, depending on their observed toxicity data. For example, right panel shows that at first cycle patient  $D$ 's risk of being in *low* LOTox state was 5.14 times higher the risk of having a

**Table 6.4.** Observed toxicity categories over cycles  $t = 1, \dots, 6$  for four random patients  $i \in \{A, B, C, D\}$  aged 15 years old. Categories for *generic* toxicities (nausea, infection and oral mucositis) are  $\{0 : \text{none}, 1 : \text{mild}, 2 : \text{moderate}, 3 : \text{severe}\}$  ( $j = 1, 2, 3$ ). Categories for *drug-specific* toxicities (cardiac toxicity, ototoxicity and neurological toxicity) are  $\{0 : \text{no}, 1 : \text{yes}\}$  ( $j = 4, 5, 6$ ). For each patient  $i$  the complete response vector is  $\tilde{\mathbf{y}}_i = (\mathbf{y}_i^{(1)}, \dots, \mathbf{y}_i^{(6)})$  where  $\mathbf{y}_i^{(t)} = (y_{i1}^{(t)}, \dots, y_{i6}^{(t)})$ .

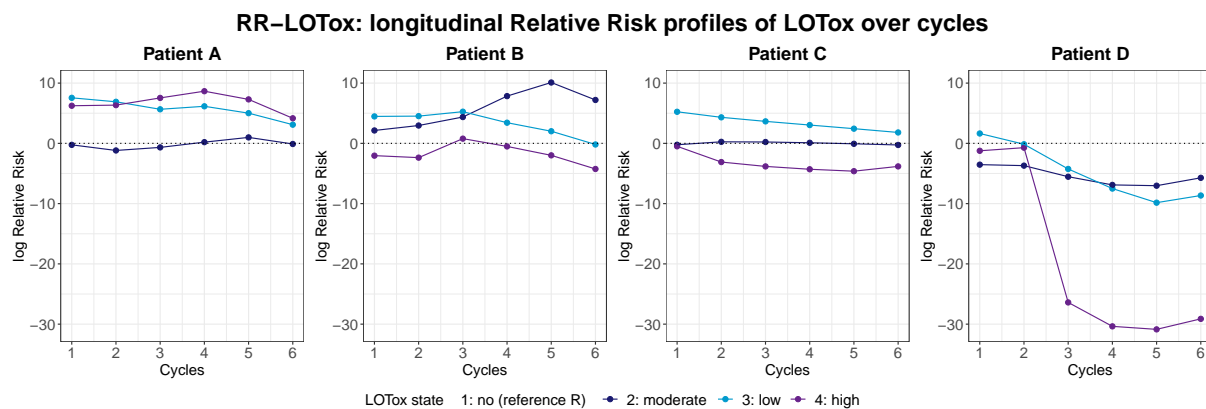
Patient $i$	Cycle $t$	$age_i$	Naus $y_{i1}^{(t)}$	Inf $y_{i2}^{(t)}$	Oral $y_{i3}^{(t)}$	Car $y_{i4}^{(t)}$	Oto $y_{i5}^{(t)}$	Neur $y_{i6}^{(t)}$
A	1	15	3	0	1	0	0	0
	2		3	1	0	0	0	0
	3		3	3	0	0	0	0
	4		3	2	1	0	0	0
	5		3	0	2	0	0	0
	6		3	0	1	0	0	0
B	1	15	1	0	0	0	0	0
	2		1	0	0	0	0	0
	3		3	0	0	0	0	0
	4		1	0	0	0	1	0
	5		1	0	0	0	1	0
	6		1	0	0	0	1	0
C	1	15	2	0	0	0	0	0
	2		1	0	0	0	0	0
	3		1	0	0	0	0	0
	4		1	0	0	0	0	0
	5		1	0	0	0	0	0
	6		1	0	0	0	0	0
D	1	15	2	0	0	0	0	0
	2		2	0	2	0	0	0
	3		0	0	0	0	0	0
	4		0	0	0	0	0	0
	5		0	0	0	0	0	0
	6		0	0	0	0	0	0

non-toxic condition, whereas risks of *high* and *moderate* LOTox were 0.29 and 0.03 times lower, respectively. Then, RR-LOTox profiles evolved over the cycles, as shown by the four trajectories, ending up with negligible relative risks ( $< 0.01$ ) for *low/moderate/high* LOTox conditions compared with a non-toxic condition at the end of treatment.

Both longitudinal P-LOTox and RR-LOTox profiles summarize and quantify the overall toxic risk over time for each patient based on observed individual characteristics, capturing differences in the overall history of toxicity across patients. P-LOTox profiles reflect the absolute size of the probabilities over time for each latent state, whereas RR-LOTox profiles focus on the relative risk with respect to the clinically desirable condition, i.e., the non-toxic one.



**Figure 6.4.** Longitudinal Probability profiles of Latent Overall Toxicity (P-LOTox)  $p_{iu}$ . Each panel refers to a different patient  $i = \{A, B, C, D\}$  in Table 6.4. Different colours refer to different latent states  $u \in \{1, 2, 3, 4\}$  (green: *no LOTox* state 1; yellow: *low LOTox* state 3; orange: *moderate LOTox* state 2; red: *high LOTox* state 4).



**Figure 6.5.** Longitudinal Relative Risk profiles of Latent Overall Toxicity (RR-LOTox)  $r_{iu}$ . Each panel refers to a different patient  $i = \{A, B, C, D\}$  in Table 6.4. Reference LOTox state is *no LOTox* state  $R = 1$ . Different colours refer to different non-reference latent states  $u \in \{2, 3, 4\}$  (light-blue: *low LOTox* state 3 vs *no LOTox*; blue: *moderate LOTox* state 2 vs *no LOTox*; purple: *high LOTox* state 4 vs *no LOTox*).

## 6.3. Final remarks

Due to the presence of multiple types of Adverse Events (AEs) with different levels of severity, identifying the actual extent of toxic burden and investigating the evolution of patient's overall toxicity represent challenging problems in cancer research. AEs are one of the main factors determining clinical decisions in medical interventions and treatment planning, playing a fundamental role in health assessment and patient monitoring. The development of statistical methods able to summarize multiple AEs and to deal with the complexity of chemotherapy data, considering both the longitudinal and categorical aspects of toxicity levels progression, is necessary and clinically relevant.

This chapter proposed a new taxonomy based on LM models with covariates and CoDa methods to provide novel techniques for investigating the evolution of the latent overall toxicity condition for each patient over chemotherapy treatment. This is important for the development of new tools to support clinical decisions in tailored interventions for effec-



tive management of adverse symptoms and treatments. The novel approach was applied to longitudinal chemotherapy data for osteosarcoma patients from MRC BO06/EORTC 80931 Randomized Controlled Trial.

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 common distribution of toxic categories, and by a similar overall toxicity burden. Four LOTox states were found, which represent different levels of multiple AEs severity: (i) people in quite good conditions (*no LOTox state 1*), (ii) patients undergoing only nausea or vomiting - mostly moderate or severe - (*low LOTox state 3*), (iii) subjects with non-severe nausea and the possibility to develop *drug-specific* AEs (*moderate LOTox state 2*), or (iv) people with multiple severe/moderate *generic* toxicities (*high LOTox state 4*). The LM approach estimated the initial prevalence of each state and the probability of individual changes over time. This allowed to reconstruct the patient-specific longitudinal LOTox profiles to assess the dynamic evolution of overall toxicity burden during treatment for each subject.

Both longitudinal P-LOTox and RR-LOTox profiles captured the individual realisations of the latent process over cycles, showing different patterns of overall toxicity evolution during treatment among patients. P-LOTox profiles illustrated the latent process using absolute terms, giving insights into the actual probabilities of being in the various LOTox states over cycles. RR-LOTox profiles – obtained by additive log-ratios transformation – reported relative risk measures to emphasize the difference between low/moderate/high LOTox states and the clinically desirable non-toxic condition. These aspects can not be investigated using a simple path prediction (see Appendix C.1). Together, absolute probabilities and relative risks provide a full picture of the individual LOTox dynamics during treatment, which may be considered as a proxy for patient’s quality of life and used to describe patient’s response to therapy over cycles in terms of toxic AEs.

This retrospective exploratory analysis has some limitations. The procedure used to select the final model may miss the best available one, since not all possible models have been fitted. However, it is computationally efficient and follows a standard stepwise forward selection approach. The analysis was performed on a single trial in osteosarcoma, considering only non-haematological toxicities. Other factors of potential interest were not routinely recorded during the trial, including among others nephrotoxicity, lymphocytes count or tumour size. To get more information about the robustness of the model developed in this study, it should be applied to other osteosarcoma data provided that the toxicity are longitudinally recorded. Nevertheless, this work opens doors to further researches, both in the field of statistical methodology development as well as in cancer research. The additive log-ratios transformation allowed to remove non-negative and sum-to-one constraints of the CoDa vectors, mapping the compositions to a real space. Standard statistical techniques could then be applied to the transformed data, opening doors for further research. Based on their different LOTox dynamics, patients could be stratified in different risk groups to be used during treatment. The relationship between AEs, treatment modifications and time-to-event outcomes may be investigated to provide new insights in the treatment effect during the evolution of the disease. To model

them simultaneously is not a trivial task since a suitable characterisation of the intricate mechanism between toxicity, chemotherapy dose and survival requires both statistical and clinical expertises, as we will see in Chapter 7.

In summary, in this chapter we proposed a novel approach to summarise and quantify patient's overall toxic risk and its evolution during treatment. Provided that toxicities are recorded according to the CTCAE scale or an analogous grading system, the LM approach represents a general and flexible method to quantify the personal evolution of overall toxic risk during chemotherapy. In cooperation with medical staff, this novel methodology might provide insights for the definition of new guidelines to reduce the impact of chemotherapy treatment in terms of toxicity burden.

## C. Appendix to Chapter 6

### C.1. Path prediction for latent Markov models

In latent Markov models literature, once the model has been estimated, a decoding procedure is usually implemented to obtain a path prediction  $\mathbf{u}_i^* = \left( u_i^{*(1)}, \dots, u_i^{*(T)} \right)$  of the most likely latent states over time for each subject  $i$ , on the basis patient-specific observed data.

Among the developed procedures, *local decoding* finds the most likely state occupied by a subject at any time point  $t$ : elements of  $\mathbf{u}_i^*$  can be obtained by maximizing the *posterior probabilities* at each time  $t$  in Equation 6.4, as follows

$$u_i^{*(t)} = \max_{u \in \{1, \dots, k\}} p_{iu}^{(t)} = \max_{u \in \{1, \dots, k\}} \text{P} \left( U_i^{(t)} = u \mid \tilde{\mathbf{Y}}_i = \tilde{\mathbf{y}}_i, \tilde{\mathbf{X}}_i = \tilde{\mathbf{x}}_i \right) \quad \text{for all } t = 1, \dots, T.$$

As an alternative, *global decoding* finds the most likely sequence of latent states for a given subject on the basis of the responses he/she provided. It is based on an adaptation of the Viterbi algorithm [211, 96] which maximises the *joint conditional probability* for each subject  $i$ , i.e.,

$$\mathbf{u}_i^* = \arg \max_{\mathbf{u}} \text{P} \left( \mathbf{U}_i = \mathbf{u} \mid \tilde{\mathbf{Y}}_i = \tilde{\mathbf{y}}_i, \tilde{\mathbf{X}}_i = \tilde{\mathbf{x}}_i \right),$$

through a forward-backward recursion. For further details see [22, 23].

#### C.1.1. B006 data application: LOTox sequences

In case of longitudinal toxicity data, path prediction  $\mathbf{u}_i^*$  represents the sequence of LOTox states over time for subject  $i$ . Let us consider the four patients aged 15 years old with different observed toxic categories over cycles reported in Table 6.4. The *LOTox sequences* for patients  $i = \{A, B, C, D\}$  can be then obtained as

- (i) the sequences of the most probable LOTox states at each cycle  $t$  (i.e., *local decoding*)

$$\mathbf{u}_A^* = (3, 3, 4, 4, 4, 4), \quad \mathbf{u}_B^* = (3, 3, 3, 2, 2, 2), \quad \mathbf{u}_C^* = (3, 3, 3, 3, 3, 3), \quad \mathbf{u}_D^* = (3, 1, 1, 1, 1, 1);$$

- (ii) or the sequences of the most likely LOTox states across cycles (i.e., *global decoding*)

$$\mathbf{u}_A^* = (3, 3, 4, 4, 4, 4), \quad \mathbf{u}_B^* = (3, 3, 3, 2, 2, 2), \quad \mathbf{u}_C^* = (3, 3, 3, 3, 3, 3), \quad \mathbf{u}_D^* = (3, 3, 1, 1, 1, 1).$$

Differences between (i) and (ii) are due to the different types of probabilities that are maximized, respectively posterior and joint conditional probabilities. The individual *LOTox sequence* allows to predict the LOTox state to which every patient belongs at a given cycle. However, it represents a summary of how the entire latent process evolves during treatment for a patient, as it only provides information about the most-likely condition without giving details about other states.

## CHAPTER 7

---

# Investigating the causal effects of joint-exposure on survival outcome in presence of time-varying confounders

The content of this chapter is based on the work by M. Spreafico, C. Spitoni, C. Lancia, F. Ieva and M. Fiocco “Causal effects of chemotherapy regimen intensity on survival outcome in osteosarcoma patients through Marginal Structural Cox Models” 2022.

Although multidisciplinary management of chemotherapy has improved clinical outcomes in patients with osteosarcoma, over the past 40 years there have been no further improvements in survival [15]. The strongest prognostic factor of both event-free and overall survival known so far in osteosarcoma is Histological Response (HRe) [31], i.e., improvement in the appearance of microscopic tissue specimens in a patient after pre-operative chemotherapy, whereas the impact of chemotherapy dose modification on patients’ survival is still unclear [111].

As mentioned in the previous chapters, in cancer trials the relationship between chemotherapy regimen intensity and survival is problematic to analyse due to the presence of negative feedback between exposure to cytotoxic drugs and consequent toxic side effects. Chemotherapy is usually modelled by different allocated regimens, i.e., by Intention-To-Treat (ITT) analysis [70]. Since ITT ignores anything that happens after randomization, such as protocol deviations or changes in drug intake over time [110], the Received Dose Intensity (RDI) [86] indicator has been introduced to analyse how close the *actual treatment* delivered is to the planned treatment, marking a significant departure from ITT in the direction of a closer description of the actual clinical practice. Lancia *et al.* (2019) [111] showed that there is mismatch between target and achieved chemotherapy-RDI in osteosarcoma due to toxic side effects developed by patients through therapy. Toxicities affect subsequent exposure by delaying the next cycle or reducing chemotherapy doses [112], representing one of the principal reasons for treatment discontinuation [186]. Being at the same time risk factors for mortality and predictors of future exposure levels, toxicities hence represent *time-dependent confounders* for the effect of chemotherapy on patient’s survival. In the presence of confounders, classical survival approaches, such

## 7. Causal effects of joint-exposure on survival in presence of time-varying confounders

as the Cox model [46], have limitations in causally interpreting the hazard ratio of the treatment variable, even if the treatment assignment is randomized [2].

Time-dependent confounding of the exposure-outcome association represents a specific challenge for estimating the effect of a treatment on an outcome of interest. Standard analyses fail to give consistent estimators in the presence of time-varying confounders if those confounders are themselves affected by the treatment [48]. For this reason, different statistical methods to control for exposure-affected time-varying confounding have been proposed, including, among others, g-computation formula [169], g-estimation of structural nested models [170] or Marginal Structural Models (MSMs) estimated using Inverse Probability of Treatment Weighting (IPTW) [171]. In case of time-to-event outcomes, Clare *et al.* (2019) [39] found that the *Cox-type Marginal Structural Model* (Cox MSM, or marginal structural Cox model) approach is by far the most commonly used method in practice.

Cox MSMs were introduced by Hernán *et al.* (2000) [78] as a class of methods for estimating the causal effect of therapy modifications on survival in presence of time-dependent confounders through IPTW. Making use of marginal (population average) rather than conditional hazard models [105], Cox MSMs target *counterfactual* (or *potential*) time-to-event variables, i.e., variables indicating when an event would have been observed if the patient had been administered a specific exposure level. IPTW is a propensity score-based method that creates a pseudo-population by weighting each subject with the inverse probability of observing a certain treatment allocation given past exposure and confounders. In such a new pseudo-population, confounders no longer predict exposure and the causal effects of treatment modifications on survival can be just obtained by a crude analysis. IPTW construction requires a thoughtful process that includes the determination of an adequate set of confounding covariates which enter into the decision-making process of allocating a treatment modification and on which the four main assumptions of causal inference with MSMs (i.e., no unmeasured confounding, consistency, positivity, no model misspecification) [77] can be tolerated [41]. Compared to a standard propensity score matching, IPTW has the advantages of retaining all eligible patients in the analysis, which may be preferred if there are limitations in terms of sample size, as well as the ability to include more than two treatment comparisons simultaneously [10].

Motivated by a clinical question concerning the effect of changes in therapy intensity on survival for osteosarcoma patients, in this chapter treatment-administration data are used to assess the causal effect on Event-Free Survival (EFS) of chemotherapy-exposure seen in terms of both (i) improvement in the appearance of microscopic tissue specimens in a patient after pre-operative treatment, i.e., by HRe, and (ii) reductions in actual versus anticipated/planned dose intensity, i.e., by RDI reductions. Data from the control arms of two clinical trials of chemotherapy in osteosarcoma, namely, European Osteosarcoma Intergroup studies MRC BO03/EORTC 80861 [120] and MRC BO06/EORTC 80931 [119] are analysed. These data are complex because the drug administration is longitudinal while only the most severe side-effects are recorded. The analysis of such mixed longitudinal/non-longitudinal data requires both an original analytical strategy and an unconventional model formulation. Moreover, since adjustments in treatment allocation

are determined by the overall toxic burden of each patient, the different types and number of side effects must be adequately summarized and quantified [190]. Suitable IPTW-based techniques and Cox MSMs are hence designed to mimic a randomized trial where joint-exposure intensity is no longer confounded by toxicities or other confounders, and a crude analysis suffices to estimate the causal effect of exposure modifications. This requires (i) a proper (time-dependent) definition of the joint-exposure, (ii) a tailor-made identification of all possible (time-dependent) confounders, and (iii) a suitable characterisation of the causal structure of the chemotherapy data. In particular, two alternative definitions of joint-exposure, based on *time-fixed final RDI* or *time-dependent pre/post-operative RDI* [120] combined with HRe, are proposed along with their relative confounding factors and Direct Acyclic Graphs (DAGs) [67, 77] to characterize the causal exposure-confounders-outcome structure. To the best of our knowledge, this is the first application of IPTW-based techniques to survival data from randomized trials of chemotherapy in order to eliminate the *toxicity-treatment-adjustment* bias.

The aim of this chapter is hence presenting an all-round RDI-based analysis of complex chemotherapy data, with tutorial-like explanations of the difficulties encountered and the problem-solving strategies deployed. Data from BO03 and BO06 trials are presented in Section 7.1. The process of building proper causal models based on joint-exposure (difficult due to lack of longitudinal confounders) using two alternative strategies is shown in detail in Section 7.2. Sections 7.3 and 7.4 are devoted to discussing the Cox MSMs results, in contrasts with their standard Cox analogues fitted on the unweighted original population, and drawing final conclusions, respectively.

## 7.1. Data description

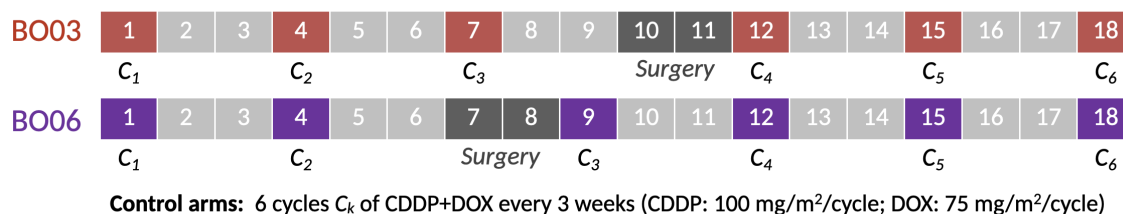
Data from control arms (i.e., conventional regimen *Reg-C*) of the Randomized Controlled Trials (RCTs) MRC BO03/EORTC 80861 and MRC BO06/EORTC 80931 (*International Standard Randomised Controlled Trial Number: ISRCTN 11824145* and *ISRCTN 86294690* respectively, <https://www.isrctn.com>) were analysed. Both RCTs were funded by the Medical Research Council (MRC) (<https://www.ukri.org/councils/mrc/>) and the European Organisation for Research and Treatment of Cancer (EORTC) (<https://www.eortc.org>). In both trials, control arms were characterized by the standard European Osteosarcoma Intergroup (EOI) treatment structured in 6 cycles of 3-weekly Cisplatin (CDDP) ( $100 \text{ mg/m}^2$ ) plus Doxorubicin (DOX) ( $75 \text{ mg/m}^2$ ), and compared to a different therapy regimen (i.e., variant of Rosen's T10 regimen [178] in BO03 and a 2-weekly dose-intensified version of CDDP+DOX [119] in BO06). Results of the primary analyses on BO03 and BO06 data can be found in Lewis *et al.* (2000; 2007) [120, 119].

In Section 7.1.1 the selected cohort of patients from BO03 and BO06 trials is illustrated. Longitudinal chemotherapy data and patient characteristics are presented in Section 7.1.2.

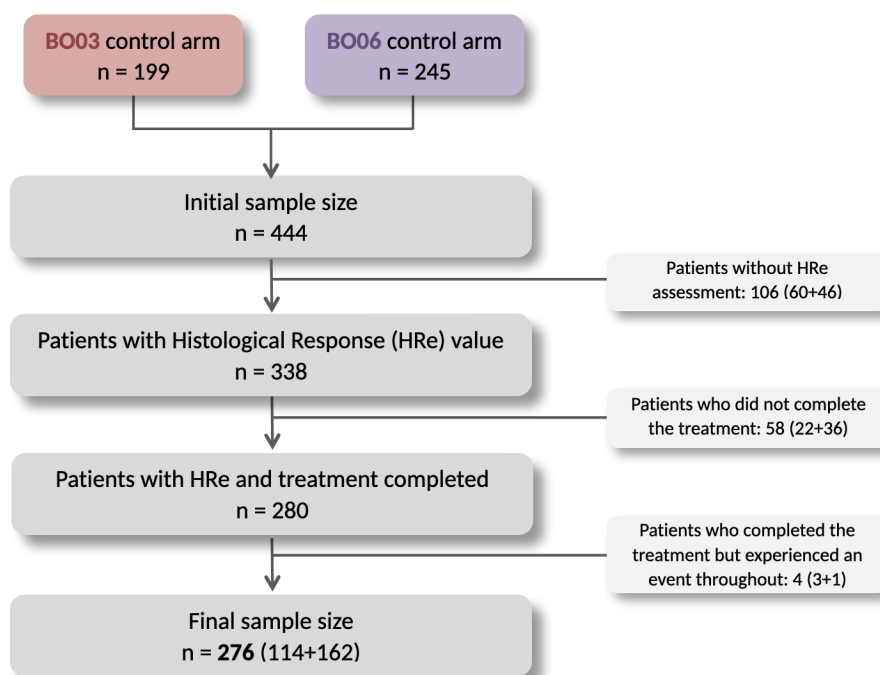
### 7.1.1. Control arms protocol and Cohort selection

As the control arms design in Figure 7.1 shows, in both RCTs chemotherapy was administered before and after surgical removal of the primary osteosarcoma. At the end of the pre-operative treatment, with a nominal duration of 3 cycles in BO03 and 2 in BO06, the tumour was surgically resected, and the levels of tumour necrosis and HRe evaluated. Variations to the planned surgery-schedule happened quite often due to administrative reason (delayed surgery) or disease progression (premature surgery), in a limited number of cases surgery was delayed due to haematological toxicity (low platelets count). Post-operative chemotherapy was intended to resume 2 weeks after surgery.

Originally, 444 patients were enrolled in the control arms of BO03 (199) and BO06 (245). In this sample, 106 (23.9%) patients were excluded due to missing HRe. Of the remaining 338 patients, 58 terminated the chemotherapy treatment prematurely or without surgery, while 4 completed the treatment but experienced an event throughout. The final cohort of 276 patients (114 from BO03 and 162 from BO06, respectively) included in the analyses (62.2% of the initial sample) is shown in the consort diagram in Figure 7.2.



**Figure 7.1.** Control arms design for BO03 and BO06 randomised clinical trials, characterized by the standard European Osteosarcoma Intergroup treatment structured in 6 cycles of 3-weekly Cisplatin (CDDP) (100 mg/m<sup>2</sup>) plus Doxorubicin (DOX) (75 mg/m<sup>2</sup>).



**Figure 7.2.** Flowchart of cohort selection.

### 7.1.2. Complexity of chemotherapy data

In cancer trials, therapy administration is usually complicated by the dynamical adjustment of the treatment on patients' clinical picture. Exposure to chemotherapy is likely to produce multi-systemic side effects, e.g. organ toxicity or myelosuppression. These side effects are a threat to patient's life and must be controlled by allocating either dose reductions/discontinuations or delays in the administration of the next course [112].

In BO03 and BO06 trials, case report forms were used to document across cycles all the information required by protocols for each patient. Patients baseline characteristics (age, gender, allocated chemotherapy regimen, site and location of the tumour) were registered at randomization. Therapy starting day was usually on the day of randomization or the day after, but could be postponed in case of administrative or clinical reasons. Treatment-related factors (administered dose of chemotherapy, cycles delays, haematological parameters, chemotherapy-induced toxicity and histological response to pre-operative chemotherapy) were collected prospectively during therapy.

A summary of baseline and trial characteristics over the entire dataset and by trial is shown in Table 7.1. Among 276 patients, 167 (60.5%) were males. Median age was 15.1 years (IQR [11.7; 18.2]). Therapy started *on time* in 71.0% of patients and surgery was performed *on time* since the start of the first cycle in 29.0% of patients.

In both studies, toxic side effects were recorded using the Common Terminology Criteria for Adverse Events Version 3 (CTCAE v3.0) [208], with grades ranging from 0 (none) to 4 (life-threatening) (see Table 7.2). Toxicity were collected longitudinally in BO06 trial, whereas in BO03 only the highest CTCAE grade (i.e., the most severe) was recorded for each toxicity in both the pre-operative and post-operative periods. According to protocols, the following side effects were linked to specific dose reduction or delay rules: *leucopenia* (i.e., a decrease in the number of white blood cells), *thrombocytopenia* (i.e., a decrease in the number of neutrophils), *oral mucositis*, *ototoxicity*, *cardiotoxicity* and *neurotoxicity*. If different *rule-specific* conditions co-existed and more than one dose reduction (or cumulative delays) applied, the lowest dose (or the highest delays) calculated was employed. According to expert knowledge, although not directly related to a specific adjustment rule, the patient's *generic* conditions of *nausea/vomiting* and *infections* was also taken into account during therapy. Treatment adjustments were hence determined as a combination of overall toxic burden related to both rule-specific and generic conditions, representing the confounding mechanisms due to toxicities.

To let pre- and post-operative toxicities be properly considered as confounding covariates and included in the analyses, individual side effects had to be appropriately summarized in order to quantify the overall toxic burden. For this purpose, the longitudinal Multiple Overall Toxicity (MOTox) score [190] introduced in Chapter 5 can be exploited. Since toxicity data over cycles were not recorded for BO03, MOTox computation was based on pre-operative and post-operative periods, considering the highest CTCAE grade recorded for each toxicity during pre/post-operative cycles.



Table 7.1. Patients baseline and trial characteristics.

	All	BO03	BO06
<b>Patients</b>	276	114 (41.3%)	162 (58.7%)
<b>Age [years]</b>			
<i>child</i> *	76 (27.5%)	26 (22.8%)	50 (30.9%)
<i>adolescent</i> *	117 (42.4%)	49 (43.0%)	68 (42.0%)
<i>adult</i> *	83 (30.1%)	39 (34.2%)	44 (27.1%)
Median [IQR]	15.1 [11.7;18.2]	16.0 [12.8;19.0]	14.6 [11.3;17.7]
Min/Max	3.6/37.5	4.7/32.6	3.6/37.5
<b>Gender</b>			
<i>Female</i>	109 (39.5%)	43 (37.7%)	66 (40.7%)
<i>Male</i>	167 (60.5%)	71 (62.3%)	96 (59.3%)
<b>Starting day**</b>			
<i>on time</i> (day 0-1)	196 (71.0%)	63 (55.3%)	133 (82.1%)
<i>low-delay</i> (day 2-3)	43 (15.6%)	23 (20.2%)	20 (12.3%)
<i>delay</i> (day $\geq$ 4)	37 (13.4%)	28 (24.5%)	9 (5.6%)
Median [IQR]	1 [0;2]	1 [0;3]	0 [0,1]
Min/Max	0/15	0/15	0/7
<b>Surgery time‡</b>			
<i>on time</i>	80 (29.0%)	29 (25.4%)	51 (31.5%)
<i>delayed</i>	196 (71.0%)	85 (74.6%)	111 (68.5%)
Median [IQR]	11 [4;22]	14 [5.25;22]	10 [4;21]
Min/Max	-39/132	-39/103	-3/132

\* Age groups were defined according to Collins *et al.* (2013) [43]: *child* (male: 0–12 years; female: 0–11 years), *adolescent* (male: 13–17 years; female: 12–16 years) and *adult* (male: 18 or older; female: age 17 years or older).

\*\* Starting day since randomization date. P-value of two-sided Mann-Whitney U test for starting day in BO03 vs BO06: 7.571e-08; p-value of chi-squared test among starting day category and trial: 1.096e-06.

‡ Surgery time (i.e., days since start of the first cycle) with respect to schedule is considered *on time* if performed from at most at the end of the scheduled week (BO03: week 10 – day 63 since start of first cycle; BO06: week 7 – day 42 since start of first cycle), or *delayed* if performed 7 or more days after scheduled date. P-value of two-sided Mann-Whitney U test for surgery time wrt schedule in BO03 vs BO06: 0.0899; p-value of chi-squared test among surgery time category and trial: 0.3397.

**Multiple Overall Toxicity score.** Let  $\mathcal{T}$  and  $k$  denote the set of different toxicities and the time-period index, respectively. Let  $tox_{ij,k}$  (with value from 0 to 4) be the most severe CTCAE grade of the  $j$ -th toxicity (with  $j = 1, \dots, |\mathcal{T}|$ ) measured during period  $k$  for the  $i$ -th patient. The Multiple Overall Toxicity (MOTox) score for the  $i$ -th patient during period  $k$  is defined as:

$$MOTox_{i,k} = \frac{1}{|\mathcal{T}|} \sum_{j \in \mathcal{T}} tox_{ij,k} + \max_{j \in \mathcal{T}} (tox_{ij,k}).$$

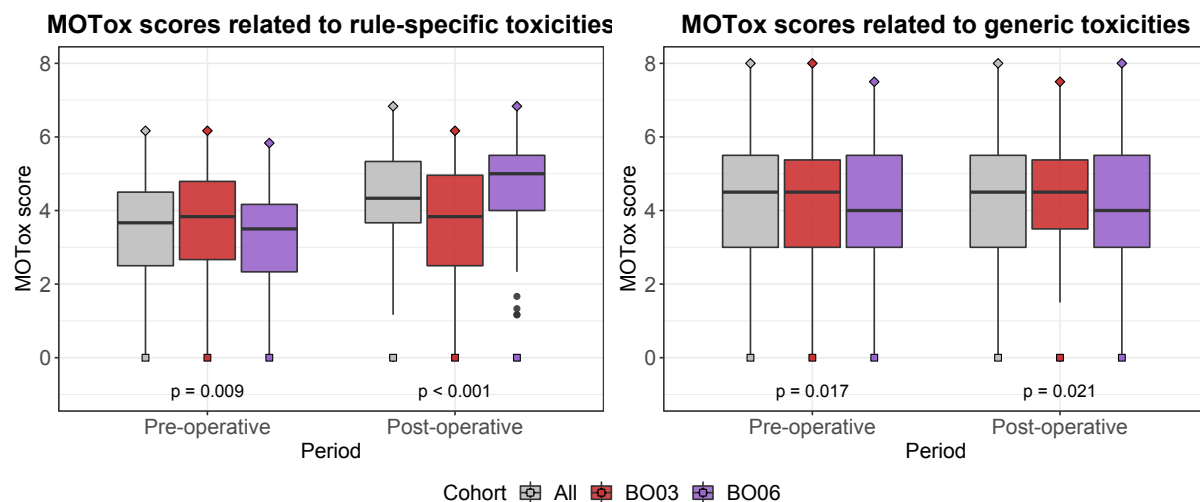
In particular, for each patient  $i$  four different MOTox scores could be computed considering as time-period index the pre-operative and post-operative periods, i.e.,  $k \in \{pre, post\}$ , and two disjoint sets of toxicities related to *rule-specific* and *generic* conditions, i.e.,  $\mathcal{T}^{(rule)} = \{leucopenia, thrombocytopenia, oral mucositis, ototoxicity, cardiotoxicity, neurotoxicity\}$  and  $\mathcal{T}^{(gen)} = \{nausea, infection\}$ .

**Table 7.2.** Toxicity coding based on Common Terminology Criteria for Adverse Events (CTCAE) v3.0 by [208] for *rule-specific* (i.e., leucopenia, thrombocytopenia, oral mucositis, ototoxicity, cardiotoxicity and neurotoxicity) and *generic* (i.e., nausea/vomiting and infections) toxicities.

Toxicity	Grade 0	Grade 1	Grade 2	Grade 3	Grade 4
<b>Rule-specific</b>					
<i>Leucopenia</i>					
White Blood Cells	$\geq 4.0 \times 10^9/L$	$[3.0 - 4.0) \times 10^9/L$	$[2.0 - 3.0) \times 10^9/L$	$[1.0 - 2.0) \times 10^9/L$	$< 1.0 \times 10^9/L$
<i>Thrombocytopenia</i>					
Neutrophils	$\geq 100 \times 10^9/L$	$[75 - 100) \times 10^9/L$	$[50 - 75) \times 10^9/L$	$[25 - 50) \times 10^9/L$	$< 25 \times 10^9/L$
<i>Oral Mucositis</i>	No change	Soreness or erythema	Ulcers: can eat solid	Ulcers: liquid diet only	Alimentation not possible
<i>Cardiac toxicity</i>	No change	Sinus tachycardia	Unifocal PVC arrhythmia	Multifocal PVC	Ventricular tachycardia
<i>Ototoxicity</i>	No change	Slight hearing loss	Moderate hearing loss	Major hearing loss	Complete hearing loss
<i>Neurological toxicity</i>	None	Paraesthesia	Severe paraesthesia	Intolerable paraesthesia	Paralysis
<b>Generic</b>					
<i>Nausea/Vomiting</i>	None	Nausea	Transient vomiting	Continuative vomiting	Intractable vomiting
<i>Infection</i>	None	Minor infection	Moderate infection	Major infection	Major infection with hypotension

PVC = Premature Ventricular Contraction

## 7. Causal effects of joint-exposure on survival in presence of time-varying confounders



**Figure 7.3.** Left panel: boxplots of pre-operative and post-operative MOTox scores related to *rule-specific* toxicities, i.e.,  $MOTox_{i,k}^{(rule)}$  with  $k \in \{pre, post\}$  and  $\mathcal{T}^{(rule)} = \{leucopenia, thrombocytopenia, oral mucositis, ototoxicity, cardiotoxicity, neurotoxicity\}$ . Right panel: boxplots of pre-operative and post-operative MOTox scores related to *generic* toxicities, i.e.,  $MOTox_{i,k}^{(gen)}$  with  $k \in \{pre, post\}$  and  $\mathcal{T}^{(gen)} = \{nausea, infection\}$ .

Boxplots are grouped by cohorts (gray: *All*; red: *BO03*; purple: *BO06*). Squares and diamonds represent minimum and maximum values, respectively. P-values  $p$  refer to Mann-Whitney U tests for the distribution of MOTox scores in *BO03* vs *BO06* cohorts.

Figure 7.3 displays a summary of pre/post-operative MOTox characteristics for both *rule-specific* (left panel) and *generic* (right panel) conditions. Overall (gray boxes), *generic* MOTox scores were high: pre/post-operative median MOTox values were equal to 4.5 meaning that in median patients experienced at least one generic side effect of CTCAE-grade 3, i.e., severe or medically significant. This is not surprising because nausea is the most common chemotherapy-induced adverse event. *Rule-specific* MOTox resulted higher in the post-operative period than in the pre-surgery one. This indicates that toxicity levels have accumulated over time resulting in more severe overall toxic burden in the second phase of treatment.

### 7.1.3. Chemotherapy exposure characteristics

Data on chemotherapy administration (administered dose of chemotherapy, cycle starting dates, delays) were collected prospectively at each treatment cycle in both trials. After pre-operative treatment cycles, surgery was performed and data about HRe were measured. Chemotherapy exposure can hence be evaluated in terms of both (i) reductions in the actual dose intensity with respect to anticipated/planned one (i.e., by RDI reduction) and (ii) improvement in the appearance of microscopic tissue specimens in a patient after pre-operative treatment (i.e., by HRe).

As mentioned in Section 7.1.1, control arm patients in both *BO03* and *BO06* underwent the standard EOI treatment structured in 6 cycles of 3-weekly CDDP plus DOX. Reductions of CDDP and/or DOX dosage at each cycle may be assessed considering the *cycle-standardized dose*, defined as follows:

**Cycle-standardized dose.** The cycle-standardized dose of drug  $d$  for patient  $i$  at cycle  $j$  is

$$\delta_{ij}^d = \frac{\text{actual dose of drug } d \text{ assumed at cycle } j \text{ by patient } i \text{ [mg/m}^2\text{]}}{\text{anticipated dose of drug } d \text{ [mg/m}^2\text{]}} \quad (7.1)$$

where  $d$  is the type of drug (CDDP or DOX). As established by trial protocols (see Figure 7.1), anticipated doses of CDDP and DOX are  $100 \text{ mg/m}^2$  and  $75 \text{ mg/m}^2$ , respectively.

Figure 7.4 shows the longitudinal nature of drug-dosage data and how treatment modifications were differently deployed in the two studies. Reductions were usually allocated in the last cycles. This is in line with the common understanding that toxicity levels are more severe towards the end of the treatment and tend to cumulate over time.

To evaluate both dose reductions/discontinuations, time-delays, and their impact in reducing the intensity of the whole therapy, the so-called Received Dose Intensity [86] approach can be adopted. RDI method is able to summarize information on treatment adjustments during the whole therapy, considering both *standardized dose* and *standardized time*.

**Standardized dose.** The standardized dose for patient  $i$  at the end of the treatment is

$$\Delta_i = \frac{1}{2} (\Delta_i^{CDDP} + \Delta_i^{DOX}) = \frac{1}{12} \left( \sum_{j=1}^6 \delta_{ij}^{CDDP} + \sum_{j=1}^6 \delta_{ij}^{DOX} \right). \quad (7.2)$$

$\Delta_i < 1$  indicates dose-reduced therapies, whereas  $\Delta_i > 1$  corresponds to dose-augmented therapies.

**Standardized time.** The standardized time for patient  $i$  at the end of the treatment is

$$\Gamma_i = \frac{\text{actual treatment time}}{\text{anticipated treatment time}} \quad (7.3)$$

where

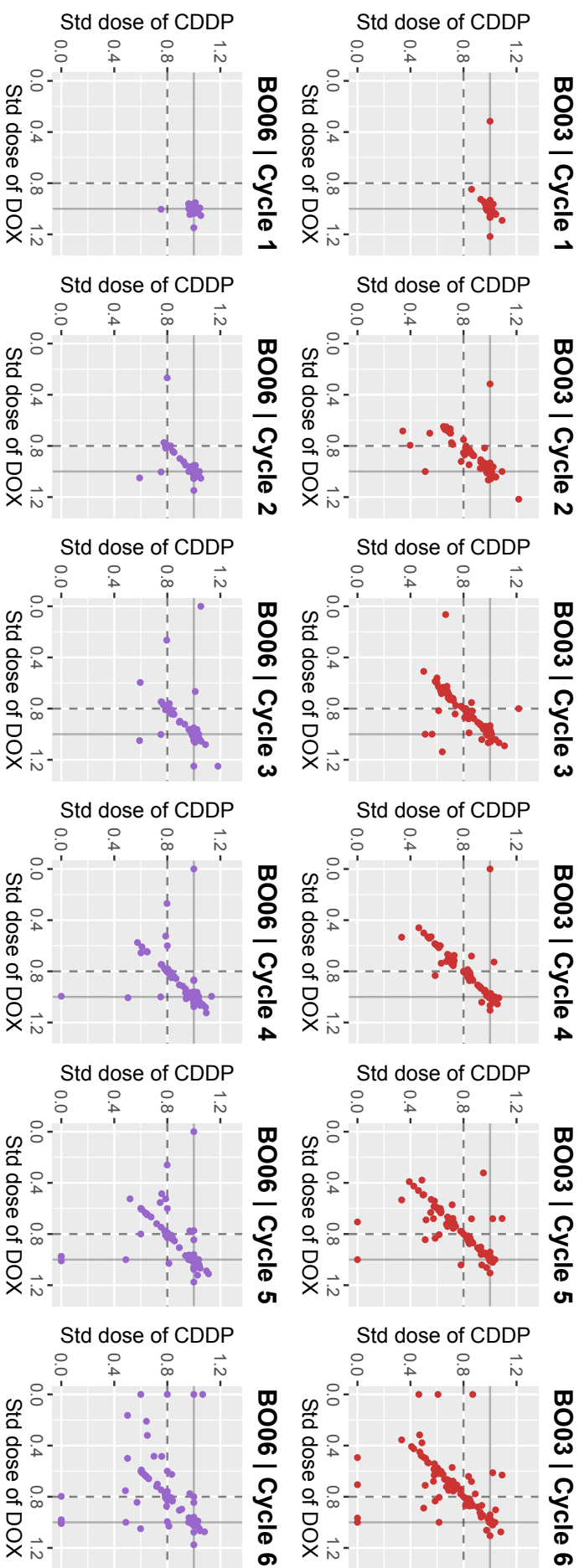
- *actual treatment time* is the difference in days between the starting date of cycle 1 and the 3rd day after the start of cycle 6,
- *anticipated treatment time* is  $21 \times 5 + 14 + 3 = 122$  days, i.e., 5 cycles lasting 21 days each, 14 days of surgery and 3 days after the start of cycle 6.

$\Gamma_i > 1$  indicates delayed therapies, whereas  $\Gamma_i < 1$  corresponds to accelerated treatments.

**Received Dose Intensity.** The Received Dose Intensity at the end of the treatment (i.e., *final* RDI) for patient  $i$  is defined as the ratio between standardized dose  $\Delta_i$  and standardized time  $\Gamma_i$ , as follows

$$RDI_i = \frac{\Delta_i}{\Gamma_i}. \quad (7.4)$$

7. Causal effects of joint-exposure on survival in presence of time-varying confounders



**Figure 7.4.** Scatter plots of cycle-standardised CDDP-dose ( $\delta_{ij}^{CDDP}$ ) vs. cycle-standardised DOX-dose ( $\delta_{ij}^{DOX}$ ) over cycles  $j = 1, \dots, 6$  (see Equation (7.1)) conditional on cycle and trial (*BO03*: top panels; *BO06*: bottom panels). Solid lines mark the planned dose, while dotted lines mark reductions of 20%. In order to improve the readability of the plot, very large values are trimmed off to 1.25.

In general,  $\Delta_i \leq 1$  and  $\Gamma_i \geq 1$  due to dose reductions and delays, respectively, and so  $RDI_i \leq 1$ .

Instead of considering the whole treatment from cycle 1 to 6, standardized dose, time and RDI could be computed for *pre-operative* and *post-operative* periods separately. For each patient, the pre-operative period is made up of cycles performed before surgery, while the post-operative period of cycles performed after surgery. Appendix D.1 reports how definitions in Equations (7.2), (7.3) and (7.4) can be adapted to consider *pre-operative* and *post-operative* periods separately, i.e.,  $\Delta_{i,k}$ ,  $\Gamma_{i,k}$  and  $RDI_{i,k}$  with  $k \in \{pre, post\}$ . Note that  $RDI_i \neq RDI_{i,pre} + RDI_{i,post}$ .

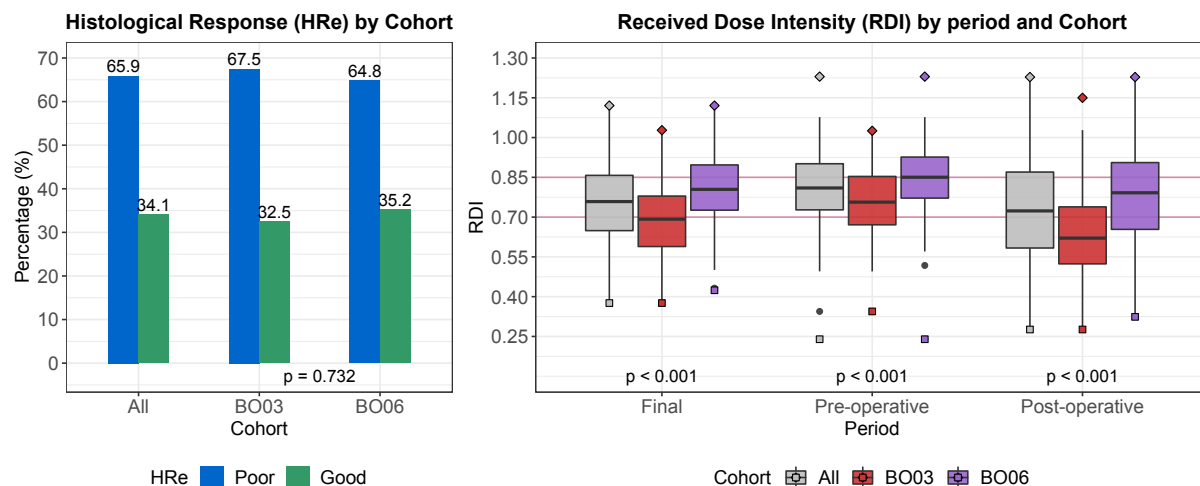
As mentioned in Section 7.1.1, the level of tumour necrosis for each patient was assessed after surgical resection (planned at the end of cycle 3/2 in BO03/BO06 – see Figure 7.1) and used to define HRe, as follows:

**Histological Response.** Histological Response (HRe) to pre-operative chemotherapy is defined as *poor* if tumour necrosis is less than 90% (i.e.,  $\geq 10\%$  of viable tumour) or *good* if tumour necrosis is greater than or equal to 90% (i.e.,  $< 10\%$  of viable tumour).

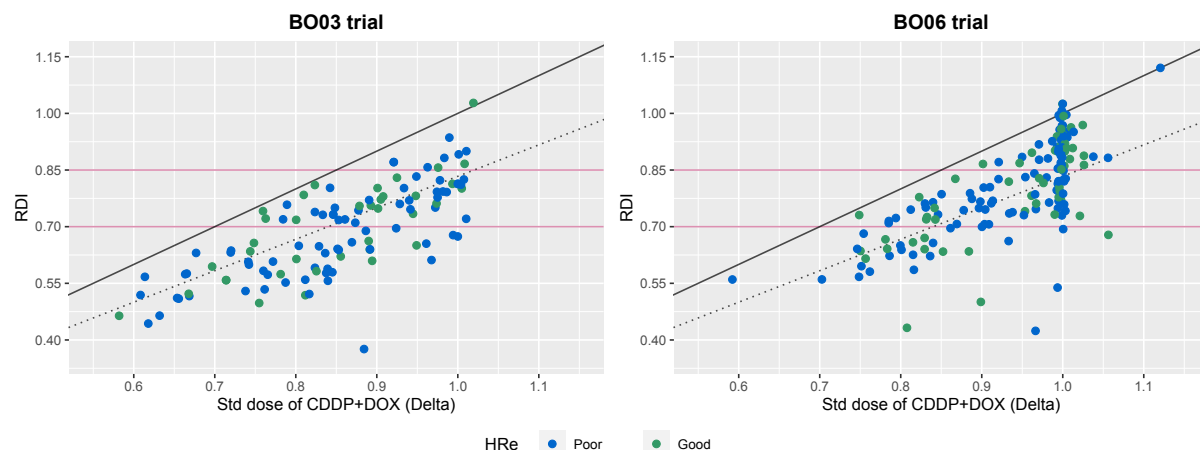
Figure 7.5 reports a summary of treatment exposure characteristics for the whole cohort and conditional on trials. The percentages of patients with a *good* HRe after surgical resection were 34.1% (94 patients) in the whole cohort, 32.5% in BO03 and 35.2% in BO06. Overall, median value of final RDI was 0.759 (IQR=[0.649; 0.857]), with minimum and maximum values of 0.376 and 1.121, respectively. Median percentages of pre-operative and post-operative RDI were 0.810 (IQR=[0.727; 0.901]) and 0.723 (IQR=[0.584;0.870]), respectively, confirming that reductions and delays are usually allocated in the post-operative cycles.

Figure 7.6 shows a scatter plot of RDI at the end of treatment (final  $RDI_i$ ) against the final standardized dose of CDDP+DOX ( $\Delta_i$ ) conditional on trial (left panel: *BO03*; right panel: *BO06*) and HRe (blue: *poor*; green: *good*). The solid horizontal lines in pink vertically divide patients with *normal* RDI levels ( $RDI_i \geq 0.85$ ) from *low* reduction ( $0.70 \leq RDI_i < 0.85$ ) and *high* reduction ( $RDI_i < 0.70$ ) patients. The solid diagonal line in black satisfies equation  $RDI_i = \Delta_i$ , dividing the group of patients with standardized time  $\Gamma_i > 1$  (delayed therapy, below the line) from the group, almost void, of patients with  $\Gamma_i < 1$  (anticipated therapy, above the line). The dotted diagonal line in black satisfies equation  $RDI_i = \Delta_i/1.2$ , dividing the group of patients with therapy delayed by more than 20% of anticipated time (below the dotted line) from the group of patients with therapy delayed by less than 20% of anticipated time (between solid and dotted black lines). This figure shows the lack of a clear association between HRe and RDI. Analogous figures for pre/post-operative RDI against their relative standardized doses can be found in Appendix D.1. Both Figures 7.5 and 7.6 clearly display the difference of treatment delivery in BO03 and BO06 trials.

## 7. Causal effects of joint-exposure on survival in presence of time-varying confounders



**Figure 7.5.** Patients treatment exposure characteristics. Left panel: barplots of Histological Response (HRe) by cohort (All, BO03, BO06) coloured according to HRe level (blue: *poor*; green: *good*). P-value  $p$  refers to the chi-squared test for the association between HRe and BO03/BO06 trial. Right panel: boxplots of final, pre-operative and post-operative Received Dose Intensity (i.e.,  $RDI_i$ ,  $RDI_{i,pre}$ ,  $RDI_{i,post}$ ) grouped by cohort (gray: *All*; red: *BO03*; purple: *BO06*). Squares and diamonds represent minimum and maximum values, respectively. P-values  $p$  refer to Mann-Whitney U tests for the distribution of RDI values in BO03 vs BO06 cohorts.



**Figure 7.6.** Scatter plots of RDI at the end of treatment (i.e.,  $RDI_i$  in Equation (7.4)) against the final standardized dose of CDDP+DOX ( $\Delta_i$ ) conditional on trial (*BO03*: left panel; *BO06*: right panel) and HRe (blue points: *poor*; green points: *good*).

## 7.2. Causal inference structure and methods

Since negative feedback between therapy administration and toxicities acts as a (generally time-dependent) confounder for the effect of chemotherapy exposure on outcome, the idea of this study is to create a pseudo-population in which medical history no longer predicts exposure through IPTW. In that framework, Cox MSMs can be used to estimate the joint causal effect of HRe and dose intensity on Event-Free-Survival (EFS). In order to create such a pseudo-population, outcome, exposure, confounders and their mutual relationships have to be defined. EFS outcome is defined in Section 7.2.1. Causal inference assumptions for MSMs are introduced in Section 7.2.2. A suitable characterisation of the causal structure of the chemotherapy data is given in Section 7.2.3. Two alternative

definitions of joint-exposure with their relative models are finally introduced in Sections 7.2.4 and 7.2.5.

### 7.2.1. Event-Free Survival Outcome

The endpoint of this study is Event-Free Survival (EFS), defined as time from the end of therapy until the first event (local recurrence, evidence of new or progressive metastatic disease, second malignancy, death, or a combination of those events) or censoring at last contact. In particular:

**EFS outcome.** The time-to-event outcome for patient  $i \in \{1, \dots, N\}$  is denoted as  $(T_i, D_i)$ , where  $T_i = \min(T_i^*, C_i)$  is the observed EFS time,  $T_i^*$  is the true event time,  $C_i$  is the censoring time (i.e., the time from the end of the therapy until the last visit) and  $D_i = I(T_i^* \leq C_i)$  is the event indicator, with  $I(\cdot)$  being the indicator function that takes the value 1 when  $T_i^* \leq C_i$ , and 0 otherwise.

### 7.2.2. Causal inference assumptions for marginal structural models

Marginal structural Cox models allow the estimation of the causal associations between treatment exposure  $\mathbf{A}$  and time-to-event response  $T$  in the presence of time-dependent covariates  $\mathbf{L}$  that may be simultaneously confounders and intermediate variables [78, 79, 100]. Cox MSMs target *counterfactual* (or *potential*) time-to-event variables  $T^a$ , i.e., the time at which an event would be observed had the subject, possibly contrary to fact, been administered a treatment exposure  $\mathbf{A} = \mathbf{a}$ . There exist four main assumptions for causal inference with (Cox) MSMs through IPTW [41, 77].

#### 1. *Exchangeability* or *No unmeasured confounding*

Exchangeability (or conditional exchangeability) implies the well-known assumption of *no unmeasured confounding* [41]. It states that exposure allocation is independent of the potential outcomes conditional on pre-treatment covariates (i.e.,  $T^a \perp\!\!\!\perp \mathbf{A} | \mathbf{L}$ ) or, in a longitudinal setting, that treatment is sequentially randomized given the past [41]. This assumption is often referred as “ignorable treatment assignment” or “sequential randomization” in statistics, “selection on observables” in the social sciences or “no omitted variable bias” in econometrics [77].

The main limitation is that, in absence of randomization such as in observational studies, exchangeability is not be testable so there is no guarantee that it holds. Experts knowledge is then necessary for the identification of enough joint predictors of exposure and outcome such that, within the levels of these predictors, associations between exposure and outcome that are due to their common causes will disappear [41].



## 2. Consistency

Consistency means that the outcome observed for each individual is precisely the counterfactual outcome under their observed treatment history, that is  $T^{\mathbf{a}} = T$  for every individual with  $\mathbf{A} = \mathbf{a}$ . This assumption would be violated in the presence of misclassification bias [217] and has two requirements [77]:

- i. since one must be able to explain how a certain level of exposure could hypothetically be assigned to a person exposed to a different level, the exposure must be *defined unambiguously* so that the counterfactual outcomes are well-defined;
- ii. there is a need to *link* the counterfactuals with observed data and thus to reasonably assume that the equality is valid for at least some individuals.

Although consistency can not be empirically verified, it is assumed plausible in observational studies of medical treatments, because one can imagine how to hypothetically manipulate an individual's treatment status [40].

## 3. Positivity

Positivity states that there is a non-zero (i.e., positive) probability of receiving every level of exposure for every combination of values of exposure and covariate histories that occur among individuals in the population [41]. If this assumption is violated, then the weights in IPTW are undefined leading to biased estimates of the causal effect.

If someone cannot be exposed to one or more levels of the confounders (e.g., it cannot be treated in the presence of recommendations from guidelines or established contraindications), then positivity is violated due to a *structural* zero probability of receiving the exposure. A solution is to restrict the inference to the subset with a positive probability of exposure, whenever possible [40]. Even in the absence of structural zeros, *random* zeros may occur by chance due to small sample sizes or highly stratified data by numerous confounders. The inclusion of weak or highly-stratified confounders can provide a better confounding adjustment but may cause severe non-positivity, increasing the bias and variance of the estimated effect. An indication of non-positivity may be the presence of estimated weights with the mean far from one or very extreme values [40].

## 4. No misspecification of both weighting and outcome models

The final assumption of MSMs is that both the weighting model for IPTW and the structural outcome model, which links the outcome to the exposure history, must be correctly specified. This assumption has similar roots in essentially all statistical models [217], as model misspecification leads to instability in the Cox MSM estimator [100, 101].

Since the presence of estimated stabilized weights with the mean far from one or very extreme values are indicative of non-positivity or misspecification of the weight

model [40], correctness of the weighting model specifications can be checked by exploring the distribution of weights [41]. In addition, quantitative (e.g., weighted standardized difference to compare means or prevalences) and qualitative graphical methods can be used to assess whether measured covariates are balanced between treatment groups in the weighted sample [19].

If these assumptions hold, causal inference is possible from MSMs through IPTW. In particular, IPTW creates a pseudo-population by weighting each patient with the inverse probability of observing a certain treatment allocation given the past treatment and confounders history. In the context of chemotherapy treatment, a pseudo-population created in this way has the following two properties:

- i. the past history of pseudo-patients no longer predicts exposure to chemotherapy in the next cycle;
- ii. the association between exposure and outcome is the same in both the original and the pseudo-population, so that causal effect of treatment modifications can be just obtained by a crude analysis on the pseudo-population.

In the following sections, joint-exposure, confounders and Cox MSMs are introduced through a thoughtful process designed to make the four assumptions acceptable. Section 7.2.3 describes a suitable characterisation of the causal structure of the chemotherapy data through the introduction of appropriate Direct Acyclic Graphs (DAGs) that identify all possible (time-dependent) confounders and their relationships with exposure and outcome. In fact, once defined the appropriate DAGs according to clinical and statistical knowledge, it can be reasonably assumed that *exchangeability* is approximately true within confounding strata. Sections 7.2.4 and 7.2.5 introduce two alternative unambiguous definitions of exposure which meet *consistency* according to experts, along with their corresponding counterfactual EFS outcomes and relative proposed Cox MSMs to estimate the association between them. *Positivity* and *no misspecification* will be finally checked for data application results in Section 7.3.

### 7.2.3. Causal structure of chemotherapy data

Relationships between random variables (i.e., exposure, confounders and outcome) is usually represented using DAGs in causal inference [67, 77]. Both clinical/oncological expertise in osteosarcoma treatment and statistical competence in variables definitions and mathematical modelling are required to construct an appropriate DAG for the problem under analysis, where the main interest is to estimate the joint causal effect of HRe and dose intensity reduction on EFS.

In both trials, HRe level was measured after surgery and can be considered as a consequence of patient's pre-operative characteristics. Only the most severe CTCAE grades were recorded in BO03, while data from BO06 are fully longitudinal in both exposure and side-effects (see Section 7.1.2). This fact posed a modelling issue, because the therapy

## 7. Causal effects of joint-exposure on survival in presence of time-varying confounders

adjustment cannot be modelled cycle-by-cycle. Two alternative options are then plausible for dose intensity:

1. *time-fixed final RDI*: the final value of RDI (i.e., the value at the end of treatment) can be seen as the result of the most severe toxicities experienced by the patient throughout the therapy;
2. *time-dependent pre/post-operative RDI*: therapy adjustment can be modelled by pre- and post-operative periods, considering the values of pre/post-operative RDI as results of the most severe overall toxicities experienced by the patient during pre/post-operative cycles.

The first option leads to a *time-fixed joint-exposure* of HRe and final RDI, whereas the second one to a *time-varying joint-exposure* given by HRe and time-varying pre/post-operative RDI.

Confounders were identified according to protocol guidelines and oncological experts knowledge. Conditioning chemotherapy administration over treatment as mentioned in Section 7.1.2, both *rule-specific* and *generic* multiple overall toxicities represent time-dependent confounders. Influencing the drug metabolism, and so being risk factors for increased toxicity, age and gender are baseline confounders because they were also clinically considered independent predictors of mortality. Although the trial does not represent a proper risk factor for failures (p-value of log-rank test for Kaplan-Meier estimators stratified by trial is about 1), it can be considered as a baseline confounder, being both an independent predictor for HRe (through number of preoperative cycles [119] and therapy starting days) and for dose intensity (see Figures 7.4, 7.5, 7.6), and influencing EFS through the way CTCAE grades were assessed and therapy modifications allocated (see Section 7.1.2). Furthermore, since there is usually a tendency not to delay surgery in the case of disease progression, the surgery timing may influence HRe.

According to the literature on MSMs, where the roman capital letter  $L$  is used to indicate a confounder, the following variables denote the characteristics of the  $i$ -th patient that influence both exposure and outcome.

**Time-fixed confounders** for the  $i$ -th patient are represented by vectors of baseline and surgery characteristics, i.e.,  $\mathbf{L}_i^{base}$  and  $\mathbf{L}_i^{surg}$  with elements:

- $L_i^{base,1}$ : trial number (BO03; BO06);
- $L_i^{base,2}$ : gender (*female*; *male*);
- $L_i^{base,3}$ : age group defined according to Collins *et al.* (2013) [43] (*child*: 0–12/0–11 years for males/females; *adolescent*: 13–17/12–16 years for males/females; *adult*: 18/17 or older for males/females);
- $L_i^{surg}$ : surgery time category with respect to schedule (0: *delayed*; 1: *on time* – see Table 7.1).

**Time-varying confounders** for the  $i$ -th patient are represented by the vectors of Multiple Overall Toxicity burden during pre/post-operative periods  $k \in \{pre, post\}$ , i.e.,  $L_{i,pre}^{tox}$  and  $L_{i,post}^{tox}$  with elements

- $L_{i,k}^{tox,1} = MOTox_{i,k}^{(rule)}$ : MOTox score related to period  $k$  based on *rule-specific* conditions  $\mathcal{T}^{(rule)}$  (see Section 7.1.2)
- $L_{i,k}^{tox,2} = MOTox_{i,k}^{(gen)}$ : MOTox score related to period  $k$  based on *generic* conditions  $\mathcal{T}^{(gen)}$  (see Section 7.1.2).

The choice of MOTox scores instead of individual CTCAE grades for the various toxicities is motivated both by the *positivity/confounders trade-off* and by the clinical protocols. By considering the individual grades for each toxicity, the number of possible confounders combinations would be too high leading to non-positivity. This choice also meets the clinical rationale, in the case of multiple toxicities, of adapting treatment according to the overall toxic burden of the patient (see Section 7.1.2).

According to experts knowledge, these characteristics have been believed to form a set of variables that satisfies the hypothesis of *no unmeasured confounding*. In particular, baseline and pre-operative MOTox confounders affect both HRe and RDI. As the delay in the surgery time already included in the calculation of the RDI (it concurs to standardized time), surgery confounder only affects HRe (p-value of chi-squared test for association is 0.023). Being HRe the response to pre-operative treatment, post-operative MOTox confounders only influence RDI.

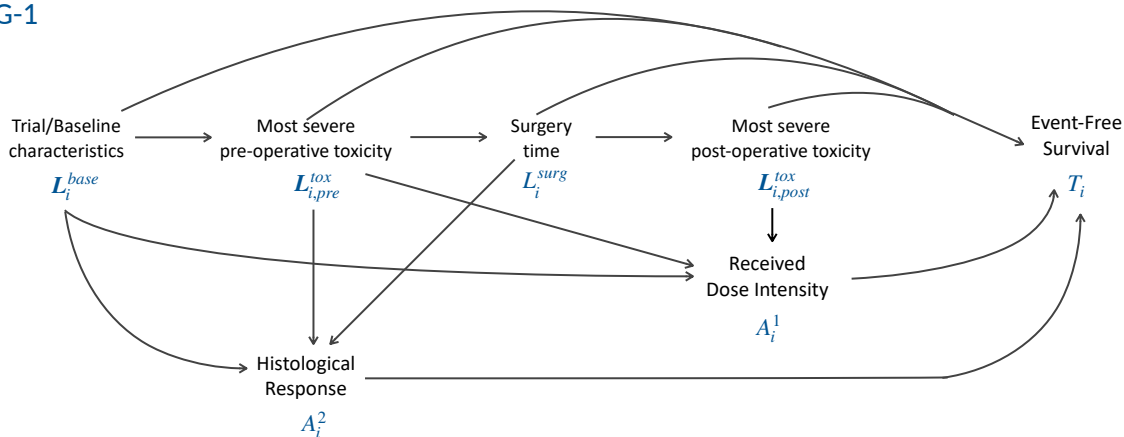
Figure 7.7 shows two alternative DAGs resulting from the causal structure described above. DAG-1 (top panel) is characterized by EFS outcome, aforementioned confounders, and the time-fixed joint-exposure given by both HRe and final RDI. DAG-2 (bottom panel) identifies a relationship among EFS outcome, confounders and a time-varying joint-exposure given by HRe and pre/post-operative RDIs. Both DAGs rely upon the hypothesis that HRe and RDI(s) are conditionally independent on the patient's toxicity-history. In other words, given two patients with the same toxicity history but different values of HRe, the probability of observing a reduction in RDI, say of 15%, is the same in the two patients regardless of one being *poor* responder and the other *good* responder. This assumption can be defended on the following two facts:

- HRe is typically not known until several weeks since chemotherapy is resumed after surgery, i.e., HRe could influence the decision to reduce therapy intensity only in the very last cycles;
- in a randomized trial clinicians can be expected to be rather committed to following the trial protocol.

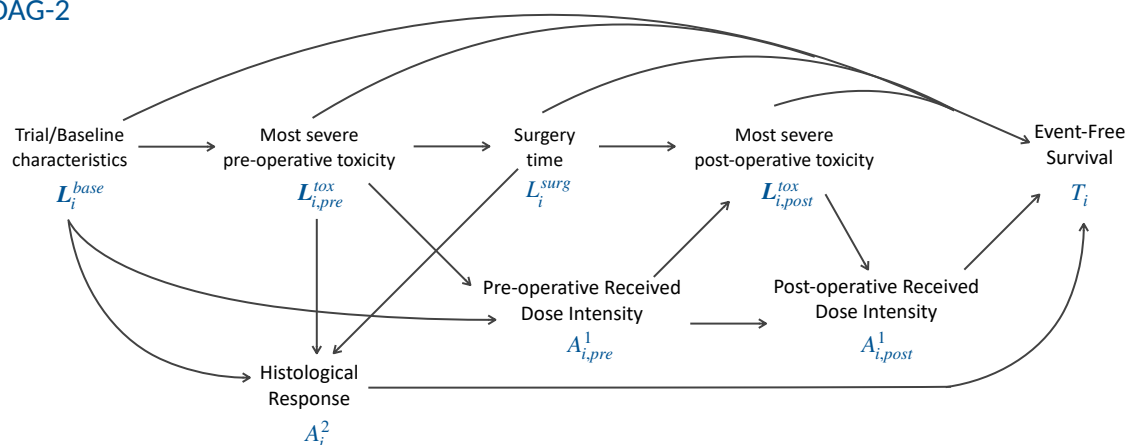
Moreover, both modelling choices do not allow for a fine continuous analysis of RDI, as this would not guarantee the assumptions of *consistency* and *positivity*. Therefore, an unambiguous well-defined categorization according to a clinical rationale of RDI exposure variables must be introduced.

## 7. Causal effects of joint-exposure on survival in presence of time-varying confounders

DAG-1



DAG-2



**Figure 7.7.** Directed Acyclic Graphs (DAGs) used to represent the causal relationships between event free survival outcome  $T_i$ , joint-exposure  $A_i$ , time-fixed confounders  $L_i$  (baseline and surgery) and time-varying confounders  $L_{i,k}$  (pre/post-operative multiple overall toxicities). Top panel (DAG-1): joint exposure is characterized by HRe and time-fixed final RDI. Bottom panel (DAG-2): joint-exposure is characterized by HRe and time-varying pre/post-operative RDI.

### 7.2.4. Joint-exposure and marginal structural Cox model for DAG-1

DAG-1 (top panel in Figure 7.7) is characterized by the EFS outcome  $T_i$ , the time-fixed and time-varying confounders ( $L_i^{base}, L_i^{surg}, L_i^{tox}, L_i^{tox}$ ), and a joint-exposure  $A_i$  given by HRe and final RDI, both time-fixed. According to expert knowledge, a *normal* RDI level (i.e.,  $RDI_i \geq 0.85$ ) can be analysed in contrast to *low-reduction* (from 15% to 30%) and *high-reduction* (more than 30%) categories. Joint-exposure  $A_i$  for DAG-1 can hence be defined as follows.

**Joint-exposure.** The *time-fixed joint-exposure* administered for subject  $i$  is denoted by the vector

$$A_i = (A_i^1, A_i^2) \quad (7.5)$$

where

- $A_i^1$  is the three-level exposure related to final RDI

$$A_i^1 = \begin{cases} 0 & \text{if } RDI_i \geq 0.85 \\ 1 & \text{if } 0.70 \leq RDI_i < 0.85 \\ 2 & \text{if } RDI_i < 0.70 \end{cases}$$

- $A_i^2$  is the binary exposure related to HRe

$$A_i^2 = \begin{cases} 0 & \text{if tumour necrosis}_i < 90\% \\ 1 & \text{if tumour necrosis}_i \geq 90\% \end{cases}$$

that is,  $A_i^2 = 1$  is equivalent to a “good” HRe, while  $A_i^2 = 0$  denotes a “poor” HRe.

Once joint-exposure is defined unambiguously, the counterfactual EFS outcome, i.e., the outcome that would be observed had the subject followed, possibly contrary-to-fact, a given treatment, is also well-defined:

**Counterfactual outcome.** Let  $T_i^{\mathbf{a}} = T_i^{(a_1, a_2)}$  denote the counterfactual EFS time that would be observed in a subject  $i$  with joint-exposure treatment

$$A_i^1 = a_1, \quad a_1 \in \{0, 1, 2\}, \quad \text{and} \quad A_i^2 = a_2, \quad a_2 \in \{0, 1\}.$$

In particular, there are exactly six joint-exposure  $(a_1, a_2)$  that can be realised according definition in Equation (7.5):

- (0,0): *poor* responder without significant reduction (i.e., *normal* RDI level);
- (1,0): *poor* responder with final *low-reduction* of 15-30%;
- (2,0): *poor* responder with final *high-reduction* of more than 30%;
- (0,1): *good* responder without significant reduction;
- (1,1): *good* responder with final *low-reduction* of 15-30%;
- (2,1): *good* responder with final *high-reduction* of 30%.

Within a counterfactual framework, i.e., in the pseudo-population, Cox MSMs enable the conceptual comparison of the hazard functions for different treatment level  $\mathbf{a} = (a_1, a_2)$ . No baseline/experimental covariates are included in the model because there is no clinical interest in assessing the causal effect of changes in chemotherapy exposure within specific population strata. The main interest consists in proposing a Cox MSM that represents the causal RDI analogue of the Intention-To-Treat (ITT) Cox model presented by Lewis *et al.* (2007) [119], which included HRe, intended treatment, and their interaction. A Cox MSM with interactions between  $a_1$  and  $a_2$ , where the treatment binary variable is replaced by the actual final RDI level, is hence proposed as follows:

## 7. Causal effects of joint-exposure on survival in presence of time-varying confounders

**Cox MSM 1.** The Marginal Structural Cox Model for EFS time under treatment level  $\mathbf{a} = (a_1, a_2)$  is

$$h_{T_i^{\mathbf{a}}}(t) = h_0(t) \exp \left\{ \beta_1 \mathbb{1}_{(a_1=1)} + \beta_2 \mathbb{1}_{(a_1=2)} + \beta_3 a_2 + \beta_4 \mathbb{1}_{(a_1=1)} a_2 + \beta_5 \mathbb{1}_{(a_1=2)} a_2 \right\} \quad (7.6)$$

To estimate the causal parameters  $\beta$  of the Cox MSM in Equation (7.6), a weighted Cox model [30, 126] can be fitted to the pseudo-population obtained through IPTW, as follows

$$h_{T_i}^{SW_i}(t|\mathbf{A}_i) = h_0(t) \exp \left\{ \theta_1 \mathbb{1}_{(A_i^1=1)} + \theta_2 \mathbb{1}_{(A_i^1=2)} + \theta_3 A_i^2 + \theta_4 \mathbb{1}_{(A_i^1=1)} A_i^2 + \theta_5 \mathbb{1}_{(A_i^1=2)} A_i^2 \right\} \quad (7.7)$$

with subject-specific stabilized weights

$$SW_i = SW_i^{A^1} \cdot SW_i^{A^2} \quad (7.8)$$

where

$$SW_i^{A^1} = \frac{P(A_i^1)}{P(A_i^1 | \mathbf{L}_i^1)} = \frac{P(A_i^1)}{P(A_i^1 | \mathbf{L}_i^{base}, \mathbf{L}_{i,pre}^{tox}, \mathbf{L}_{i,post}^{tox})};$$

$$SW_i^{A^2} = \frac{P(A_i^2)}{P(A_i^2 | \mathbf{L}_i^2)} = \frac{P(A_i^2)}{P(A_i^2 | \mathbf{L}_i^{base}, \mathbf{L}_{i,pre}^{tox}, L_i^{surg})}.$$

In both  $SW_i^{A^1}$  and  $SW_i^{A^2}$  cases, numerators are the probability that a subject  $i$  received observed exposures  $A_i^1$  and  $A_i^2$  respectively, whereas denominators are the probability that the subject received observed exposures given relative time-fixed and time-dependent confounders. Regression models have to be chosen appropriately, according to the type of exposure. In particular, multinomial logistic regression models are used for both numerator and denominator of  $SW_i^{A^1}$ , whereas binary logistic regression models are adopted for  $SW_i^{A^2}$ .

Under causal inference assumptions (see Section 7.2.2), association is causation in the pseudo-population and the estimates of the associational parameters  $\theta$  are consistent for the causal parameters  $\beta$ . In applying this methodology to the chemotherapy data, different model specifications in terms of confounding covariate features must be compared to satisfy the final assumptions of *positivity* and *no misspecification of the weight-generating models* and guarantee an unbiased estimation of the results.

### 7.2.5. Joint-exposure and marginal structural Cox model for DAG-2

DAG-2 (bottom panel in Figure 7.7) is characterized by the EFS outcome  $T_i$ , the time-fixed and time-varying confounders  $(\mathbf{L}_i^{base}, L_i^{surg}, \mathbf{L}_{i,pre}^{tox}, \mathbf{L}_{i,post}^{tox})$ , and a joint-exposure  $\bar{\mathbf{A}}_i$  given by HRe and time-varying pre/post-operative RDI. As in the previous section, a *normal* RDI level can be analysed in contrast to *low* and *high* reductions. Time-varying joint-exposure  $\bar{\mathbf{A}}_i$  for DAG-2 is hence defined as follows.

**Joint-exposure.** The *time-varying joint-exposure* administered for subject  $i$  is denoted by the vector

$$\bar{\mathbf{A}}_i = (\bar{\mathbf{A}}_i^1, A_i^2) = ((A_{i,pre}^1, A_{i,post}^1), A_i^2) \quad (7.9)$$

where

- $\bar{\mathbf{A}}_i^1$  is the time-varying three-level exposure vector related to pre/post-operative RDI with elements

$$A_{i,k}^1 = \begin{cases} 0 & \text{if } RDI_{i,k} \geq 0.85 \\ 1 & \text{if } 0.70 \leq RDI_{i,k} < 0.85 \\ 2 & \text{if } RDI_{i,k} < 0.70 \end{cases}$$

where  $k \in \{pre, post\}$  indicating the *pre-operative* and *post-operative* periods, respectively;

- $A_i^2$  is the binary exposure related to HRe

$$A_i^2 = \begin{cases} 0 & \text{if tumour necrosis}_i < 90\% \\ 1 & \text{if tumour necrosis}_i \geq 90\% \end{cases}$$

that is,  $A_i^2 = 1$  is equivalent to a “good” HRe, while  $A_i^2 = 0$  denotes a “poor” HRe.

Once joint-exposure is defined unambiguously, the counterfactual EFS outcome, i.e., the outcome that would be observed had the subject followed – possibly contrary-to-fact – a given treatment, is also well-defined:

**Counterfactual outcome.** Let  $T_i^{\bar{\mathbf{a}}} = T_i^{((a_{11}, a_{12}), a_2)}$  denote the counterfactual EFS time that would be observed in a subject  $i$  with time-varying joint-exposure

$$A_{i,pre}^1 = a_{11}, \quad A_{i,post}^1 = a_{12}, \quad a_{11}, a_{12} \in \{0, 1, 2\}, \quad \text{and} \quad A_i^2 = a_2, \quad a_2 \in \{0, 1\}.$$

In particular, there are exactly 18 time-varying joint-exposure combinations  $\bar{\mathbf{a}} = (\bar{\mathbf{a}}_1, a_2) = ((a_{11}, a_{12}), a_2)$  that can be realised according definition in Equation (7.9). To avoid too many combinations, we specify a model that combines information from many strategies to help estimate the causal effects. For example, we can hypothesize a cumulative treatment effects under sub-strategy  $\bar{\mathbf{a}}_1$ , named *cumulative RDI-exposure*

$$\text{cum}(\bar{\mathbf{a}}_1) = \sum_{k=1}^2 a_{1k}$$

which could takes value

- 0: if no reduction, neither pre nor post surgery;
- 1: if only one *low* reduction pre or post surgery;
- 2: if *low* reductions both pre and post surgery or *high* reduction pre or post surgery;



## 7. Causal effects of joint-exposure on survival in presence of time-varying confounders

- 3: if both pre-operative (or post-operative) *low* reduction and post-operative (or pre-operative) *high* reduction;
- 4: if *high* reductions both pre and post surgery.

Therefore  $\text{cum}(\bar{\mathbf{a}}_1)$  represents the number of reductions by a value of 15-30%, where a single *high* reduction of at least 30% can be seen as twice a *low* reduction of 15-30%. In the following, the *time-varying joint-exposure* levels and values for patient  $i$  (with 18 different possible combinations) are indicated by  $\bar{\mathbf{a}} = (\bar{\mathbf{a}}_1, a_2)$  and  $\bar{\mathbf{A}}_i = (\bar{\mathbf{A}}_i^1, A_i^2)$  respectively, whereas the *cumulative joint-exposure* levels and values for patient  $i$  (with 10 different possible combinations) are indicated by  $\tilde{\mathbf{a}} = (\text{cum}(\bar{\mathbf{a}}_1), a_2)$  and  $\tilde{\mathbf{A}}_i = (\text{cum}(\bar{\mathbf{A}}_i^1), A_i^2)$ , respectively.

Within a counterfactual framework, Cox MSMs enable the conceptual comparison of the hazard functions for different treatment exposure  $\bar{\mathbf{a}} = (\bar{\mathbf{a}}_1, a_2)$ . As in Section 7.2.4, no baseline/trial covariates are included in the proposed structural model. Since the interests is in analysing the causal RDI analogue of the ITT Cox model presented by Lewis *et al.* (2007) [119] according to pre/post-operative RDI definitions, a Cox MSM with interactions between  $\text{cum}(\bar{\mathbf{a}}_1)$  and  $a_2$  is hence proposed.

**Cox MSM 2.** The Marginal Structural Cox Model for EFS time under cumulative treatment level  $\tilde{\mathbf{a}} = (\text{cum}(\bar{\mathbf{a}}_1), a_2)$  is

$$h_{T_i^{\tilde{\mathbf{a}}}}(t) = h_0(t) \exp \{ \beta_1 \text{cum}(\bar{\mathbf{a}}_1) + \beta_2 a_2 + \beta_3 \text{cum}(\bar{\mathbf{a}}_1) a_2 \} \quad (7.10)$$

To estimate the causal parameters  $\beta$  of the Cox MSM in Equation (7.10), a weighted Cox model [30, 126] can be fitted to the pseudo-population obtained through IPTW, as follows

$$h_{T_i}^{SW_i}(t|\bar{\mathbf{A}}_i) = h_0(t) \exp \{ \theta_1 \text{cum}(\bar{\mathbf{A}}_i^1) + \theta_2 A_i^2 + \theta_3 \text{cum}(\bar{\mathbf{A}}_i^1) A_i^2 \} \quad (7.11)$$

where  $\text{cum}(\bar{\mathbf{A}}_i^1)$  is the cumulative RDI-exposure vector

$$\text{cum}(\bar{\mathbf{A}}_i^1) = \sum_{k \in \{pre, post\}} A_{i,k}^1$$

and  $SW_i$  are the subject-specific stabilized weights given by

$$SW_i = SW_i^{\bar{\mathbf{A}}^1} \cdot SW_i^{A^2} \quad (7.12)$$

with

$$SW_i^{\bar{\mathbf{A}}^1} = SW_i^{A_{i,pre}^1} \cdot SW_i^{A_{i,post}^1} = \frac{P(A_{i,pre}^1)}{P(A_{i,pre}^1 | \mathbf{L}_i^{base}, \mathbf{L}_{i,pre}^{tox})} \cdot \frac{P(A_{i,post}^1 | A_{i,pre}^1)}{P(A_{i,post}^1 | A_{i,pre}^1, \mathbf{L}_i^{base}, \mathbf{L}_{i,pre}^{tox}, \mathbf{L}_{i,post}^{tox})};$$

$$SW_i^{A^2} = \frac{P(A_i^2)}{P(A_i^2 | \mathbf{L}_i^{base}, \mathbf{L}_{i,pre}^{tox}, L_i^{surg})}.$$

As in the previous section, multinomial logistic regression models can be used for both numerators and denominators of  $SW_i^{\bar{\mathbf{A}}^1}$ , whereas binary logistic regression models can be adopted for  $SW_i^{A^2}$ .

Under causal inference assumptions, association is causation in the pseudo-population and the estimates of the associational parameters  $\theta$  are consistent for the causal parameters  $\beta$ . In applying this methodology to the chemotherapy data, different model specifications must be compared to satisfy *positivity* and *no misspecification of the weight-generating models* and guarantee an unbiased estimation of the results.

## 7.3. Results

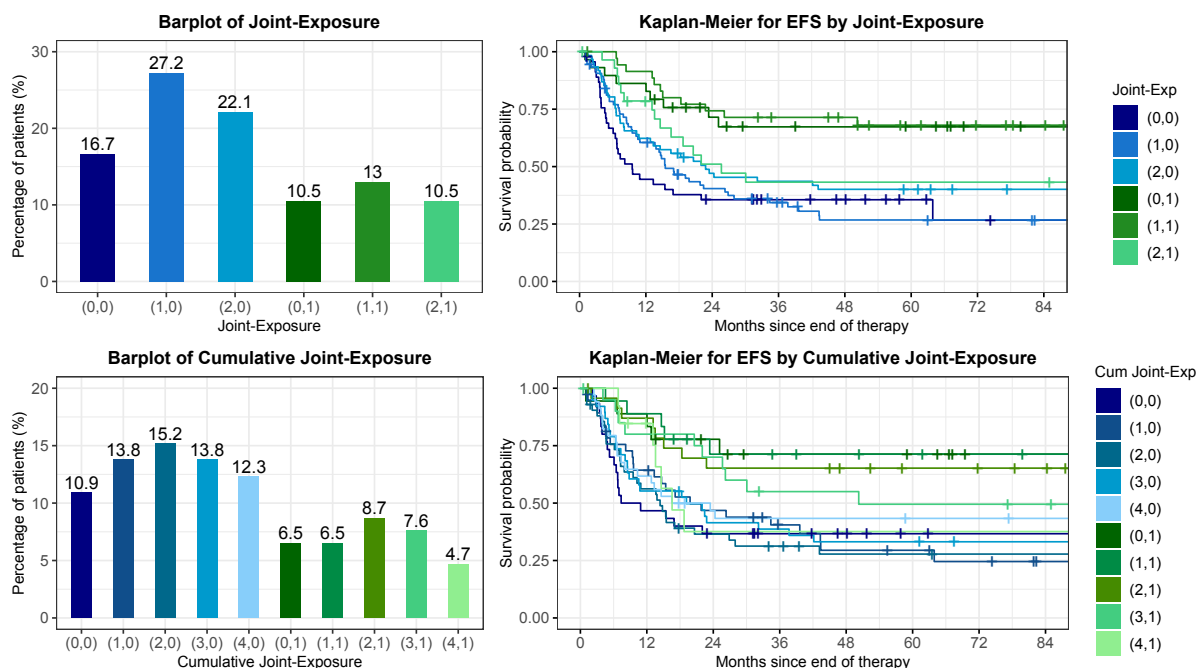
IPTW-based causal methodologies introduced in Section 7.2 are now applied to BO03-BO06 chemotherapy data presented in Section 7.1. In Section 7.3.1 joint-exposures for DAG-1 and DAG-2 are explored in terms of percentages of patients in each exposure-level and association with EFS in the original population. Different IPTW model specifications to determine the subject-specific standardized weights for the pseudo-population are presented in Section 7.3.2. Results of causal Cox MSMs fitted on the pseudo-population are presented in Section 7.3.3, along with their relative unweighted Cox results to show the *toxicity-treatment-adjustment* bias present in the original data. Statistical analyses were performed in the R-software environment [161], in particular using `ipw` [210] and `survival` [201] packages. R code for the current study is provided here: <https://github.com/mspreafico/B00x-CoxMSM>.

### 7.3.1. Joint-exposure descriptive and association with EFS

Once computed the *time-fixed* and *time-varying* joint-exposures for each subject, the percentage of patients in each level and the naive association with survival were observed. Overall, median EFS time computed using the reverse Kaplan-Meier method by Schemper and Smith (1996) [182] was 89.59 months (IQR = [50.33; 146.30]) and 155 patients (55.1%) experienced an event after the end of the therapy. Figure 7.8 shows both *time-fixed*  $\mathbf{A}_i$  (top panels) and *cumulative*  $\tilde{\mathbf{A}}_i$  (bottom panels) joint-exposure characteristics. In both cases, left panels report percentage of patients according to the various joint-exposure levels and right panels display Kaplan-Meier estimators for EFS curves stratified by joint-exposure levels with time expressed in months since end of therapy. As expected, *Good* Responders (GRs) (green curves) presented a better survival with respect to *Poor* ones (PRs – blue curves). In particular, in GRs an increased final/cumulative RDI level seemed associated with better survival, whereas a reversed association was observed in the group of PRs. However, in both cases the curve of GRs with the highest reduction overlapped PRs curves, suggesting the possibility of a non-negligible interaction between the joint-exposure components and validating the Cox MSMs proposed in Equations (7.6) and (7.10).

To further investigate these findings and analyse the causal effect of time-fixed/time-varying joint-exposure on EFS through Cox MSMs, subject-specific standardized weights must be computed from correctly specified IPTW models which take into account all the confounding factors identified in Section 7.2.3.

## 7. Causal effects of joint-exposure on survival in presence of time-varying confounders



**Figure 7.8.** Joint-exposure characteristics. Top panels refer to *time-fixed joint-exposure*  $\mathbf{A}_i = (A_i^1, A_i^2)$  introduced in Section 7.2.4, where  $A_i^1$  is the final RDI level (0: *normal*  $RDI_i \geq 0.85$ ; 1: *low-reduction*  $0.70 \leq RDI_i < 0.85$ ; 2: *high-reduction*  $RDI_i < 0.70$ ) and  $A_i^2$  is the HRe (0: *poor*: 1: *good*). Bottom panels refer to *cumulative joint-exposure*  $\tilde{\mathbf{A}}_i = (\text{cum}(\tilde{\mathbf{A}}_i^1), A_i^2)$ , where  $\text{cum}(\tilde{\mathbf{A}}_i^1)$  is the cumulative pre/post-operative RDI level described in Section 7.2.5 and  $A_i^2$  is the HRe (0: *poor*: 1: *good*). In both cases, left panels report percentage of patients by joint-exposure levels and right panels display Kaplan-Meier estimators for EFS curves stratified by joint-exposure levels.

### 7.3.2. IPTW diagnostics

Different specifications of the subject-specific standardized weights for final RDI level  $SW_i^{A^1}$ , HRe category  $SW_i^{A^2}$  and pre/post-operative RDI levels  $SW_i^{\tilde{\mathbf{A}}^1} = SW_i^{A_{pre}^1} \cdot SW_i^{A_{post}^1}$  were investigated in order to check whether and which models best satisfied *positivity* and *no misspecification*. As mentioned in Sections 7.2.4 and 7.2.5, multinomial logistic regression models were used for both numerators and denominators of  $SW_i^{A^1}$  and  $SW_i^{\tilde{\mathbf{A}}^1}$ , whereas binary logistic regression models were adopted for  $SW_i^{A^2}$ . In all cases, the following four different model specifications in terms of confounding features for the denominators were compared:

1. each confounding covariate entered the IPTW model as a main effect only and the MOTox scores were linearly related to the log-odds;
2. specification 1 + two interaction terms linearly related to the log-odds, that is (i) interaction between pre-operative rule-specific and generic MOTox scores and (ii) interaction between post-operative rule-specific and generic MOTox scores
3. specification 1 + four interaction terms between the four MOTox scores and the trial assumed linearly related to the log-odds;
4. each categorical/binary confounding covariates entered the IPTW model as a main effect only and cubic smoothing B-splines with 3 internal knots were used to model

the relationship between each of the continuous MOTox scores and the log-odds of treatment.

Table 7.3 reports the summaries of the stabilized weights obtained with the different specifications for final RDI level  $SW_i^{A^1}$ , HRe category  $SW_i^{A^2}$  and pre/post-operative RDI levels  $SW_i^{\bar{A}^1}$ .

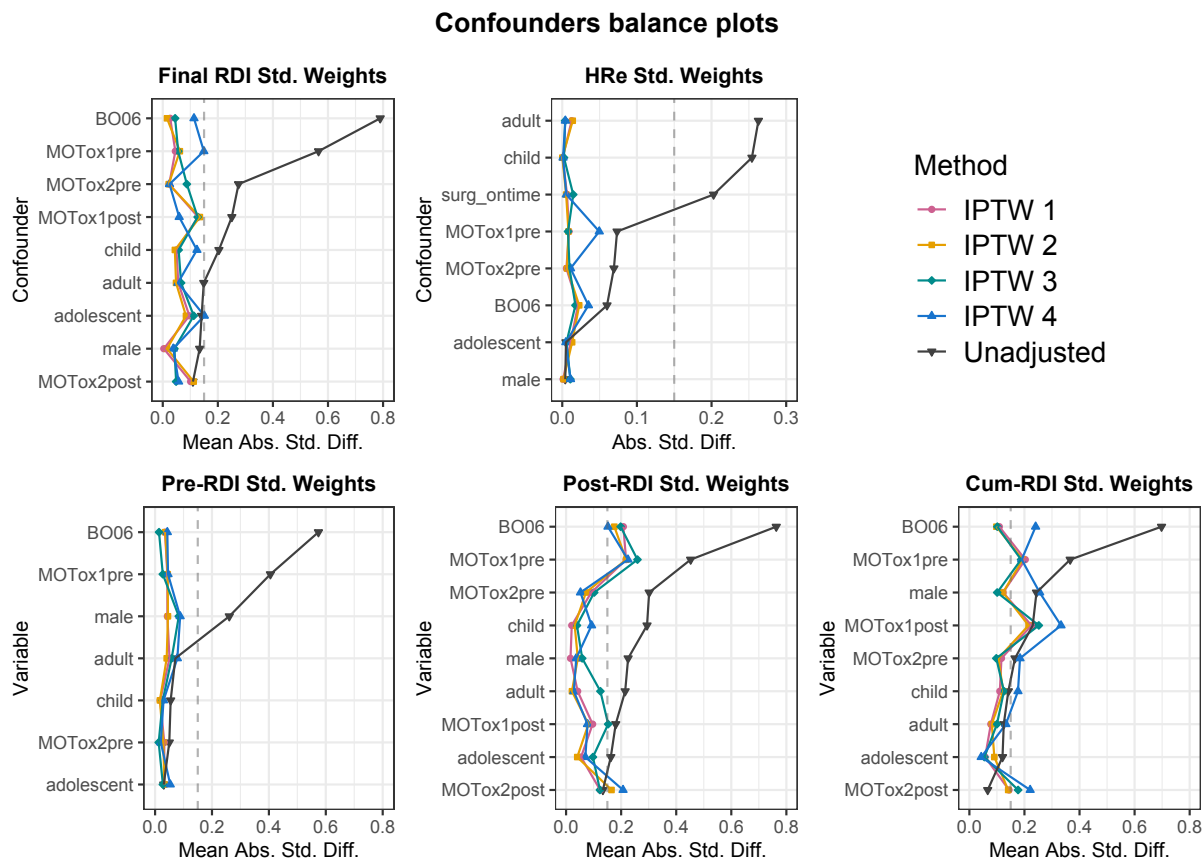
By examining the distributions of the standardized weights for final RDI, there was no evidence of non-positivity or of misspecification for IPTW methods 1 and 2 (mean values of about 0.99 without extreme values), whereas methods 3 and 4 presented lower mean values and higher standard deviations. The same was confirmed by the diagnostics balance plot in top-left panel of Figure 7.9, where the mean absolute standardized differences for final RDI confounders in the unweighted sample (black points) always exceeded those in the weighted samples, and the lowest values were observed for IPTW 1 and 2. IPTW model 1 was finally selected among the two as it had a mean value closer to 1 and lower standard deviation.

Similarly, according to the distributions of the standardized weights for HRe models, there was no evidence of non-positivity or misspecification in the four IPTW methods: they all presented a mean value of 1 with standard deviation from 0.22 to 0.25. In terms of covariates balance (top-right panel in Figure 7.9, all IPTW methods performed better than the unweighted sample (black) but IPTW 4 (blue) was worse than the others. In the absence of any particular contraindications, IPTW model 1 was finally selected as it was simpler (in terms of features) and had lower standard deviation weights.

**Table 7.3.** Inverse Probability of Treatment Weighting (IPTW) diagnostics based on summaries of stabilized weights related to final RDI level  $SW_i^{A^1}$ , HRe category  $SW_i^{A^2}$  and pre/post-operative RDI levels  $SW_i^{\bar{A}^1}$  computed using the four different specifications listed in Section 7.3.2.

Final RDI level: $SW_i^{A^1}$		
Specification	Mean (s.d.)	Min/Max
IPTW 1	0.988 (0.668)	0.330/5.252
IPTW 2	0.987 (0.682)	0.354/5.189
IPTW 3	0.979 (0.700)	0.324/5.469
IPTW 4	0.968 (0.797)	0.326/6.946
HRe: $SW_i^{A^2}$		
Specification	Mean (s.d.)	Min/Max
IPTW 1	1.001 (0.200)	0.598/1.746
IPTW 2	1.001 (0.201)	0.603/1.780
IPTW 3	1.001 (0.242)	0.578/2.116
IPTW 4	0.999 (0.250)	0.531/2.373
Pre/Post RDI levels: $SW_i^{\bar{A}^1}$		
Specification	Mean (s.d.)	Min/Max
IPTW 1	0.988 (0.839)	0.285/7.109
IPTW 2	0.994 (0.910)	0.267/8.555
IPTW 3	0.998 (1.101)	0.266/11.438
IPTW 4	0.998 (1.245)	0.161/12.959

7. Causal effects of joint-exposure on survival in presence of time-varying confounders

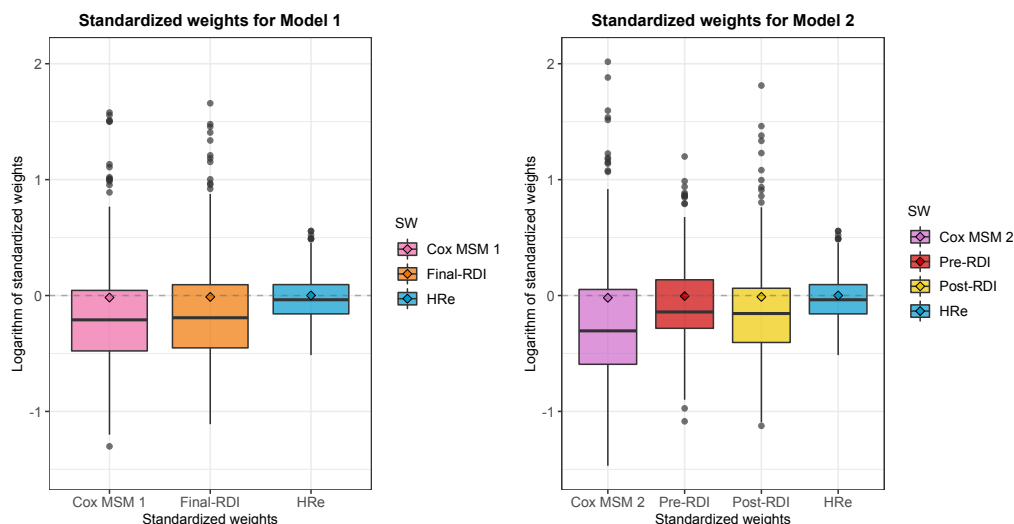


**Figure 7.9.** Diagnostic balance plot for Inverse Probability of Treatment Weighting (IPTW). Lines represent the (mean) absolute standardized differences for each exposure-related confounder according to the four different specification methods introduced in Section 7.3.2 and their unadjusted versions (pink: *IPTW 1*; orange: *IPTW 2*; green: *IPTW 3*; blue: *IPTW 4*; black: *Unadjusted*).

IPTW methods for pre/post-operative RDI levels was selected as trade-off between the two product components. No evidence of assumptions violation was observed according to the distributions of the standardized weights  $SW_i^{\bar{A}^1}$ . IPTW 4 method resulted in a worse balance of confounders in terms of mean absolute standardised differences for cumulative-RDI levels based on  $SW_i^{\bar{A}^1}$  (see bottom-right panel in Figure 7.9). The same was valid for post-RDI levels using  $SW_i^{A^1_{post}}$  obtained through IPTW 3 (bottom-centre panel). Between IPTW methods 1 and 2, both with a mean value of about 0.99, IPTW 1 was selected as it was simpler (in terms of features) and had lower standard deviation weights.

The formulas of the denominators of  $SW_i^{A^1}$ ,  $SW_i^{A^1_{pre}}$ ,  $SW_i^{A^1_{post}}$  and  $SW_i^{A^2}$  related to the selected IPTW specifications are reported in Appendix D.2.

Left panel of Figure 7.10 shows the standardized weights  $SW_i$  in Equation (7.8) obtained as product of  $SW_i^{A^1}$  and  $SW_i^{A^2}$  to be used for create the pseudo-population in case of *time-fixed joint exposure*. The y-axis is in logarithmic scale. Mean value of  $SW_i$  was 0.983 (*s.d.* = 0.694) with minimum and maximum values of 0.272 and 4.849. Analogously, right panel of Figure 7.10 shows the standardized weights  $SW_i$  in Equation (7.12) obtained as product of  $SW_i^{\bar{A}^1}$  and  $SW_i^{A^2}$  to be used for create the pseudo-population in case of *time-*



**Figure 7.10.** Diagnostic boxplots of subject-specific standardized weights computed via Equations (7.8) (left panel) and (7.12) (right panel). The scale on the y-axis is logarithmic. Diamonds represent the mean values (in logarithmic scale).

*varying joint exposure.* Mean value of  $SW_i$  was 0.981 ( $s.d. = 0.865$ ) with minimum and maximum values of 0.230 and 7.518. These weights satisfied all the required assumptions and were finally used to fit on the relative pseudo-populations the IPT weighted Cox models in Equations (7.7) and (7.11).

### 7.3.3. Causal inference through marginal structural Cox models

Once met causal inference assumptions, association was causation in both pseudo populations. The causal parameters  $\beta$  in Cox MSMs (7.6) and (7.10) were hence estimated through their consistent associational parameters  $\theta$  in IPT weighted Cox models (7.7) and (7.11) fitted on the relative pseudo-populations. Obtained estimates were finally compared to the results obtained by fitting the corresponding standard (i.e., unweighted) Cox models on the original population.

Estimated parameters for both Cox MSMs and their unweighted versions are reported in Table 7.4. In Cox MSM 1 and 2 robust standard errors for computing the confidence interval of each coefficient were obtained via the option `robust=TRUE` in R function `coxph` [201]. Figure 7.11 graphically displays the Hazard Ratios related to the different joint-exposure levels for Cox MSMs in 7.6 and 7.10 fitted on the pseudo population (left panels) and the results for corresponding unweighted models (right panels).

Top panels refers to the causal structure of DAG-1 presented in Section 7.2.4. Reference level was PRs with *normal* RDI level at the end of treatment, i.e.,  $(a_1, a_2) = (0, 0)$ . Considering the unweighted Cox model 1 (top-right), which represents the RDI-analogue of the ITT Cox model presented by Lewis *et al.* (2007) [119] without considering IPT weights, in PRs the RDI reductions appeared associated with an improvement in EFS, even if not statistically significant. With respect to GRs receiving a *normal* RDI, GRs receiving a *low-reduction* experienced an event 12% slower ( $HR = 0.273/0.309 = 0.88$ ) whereas those

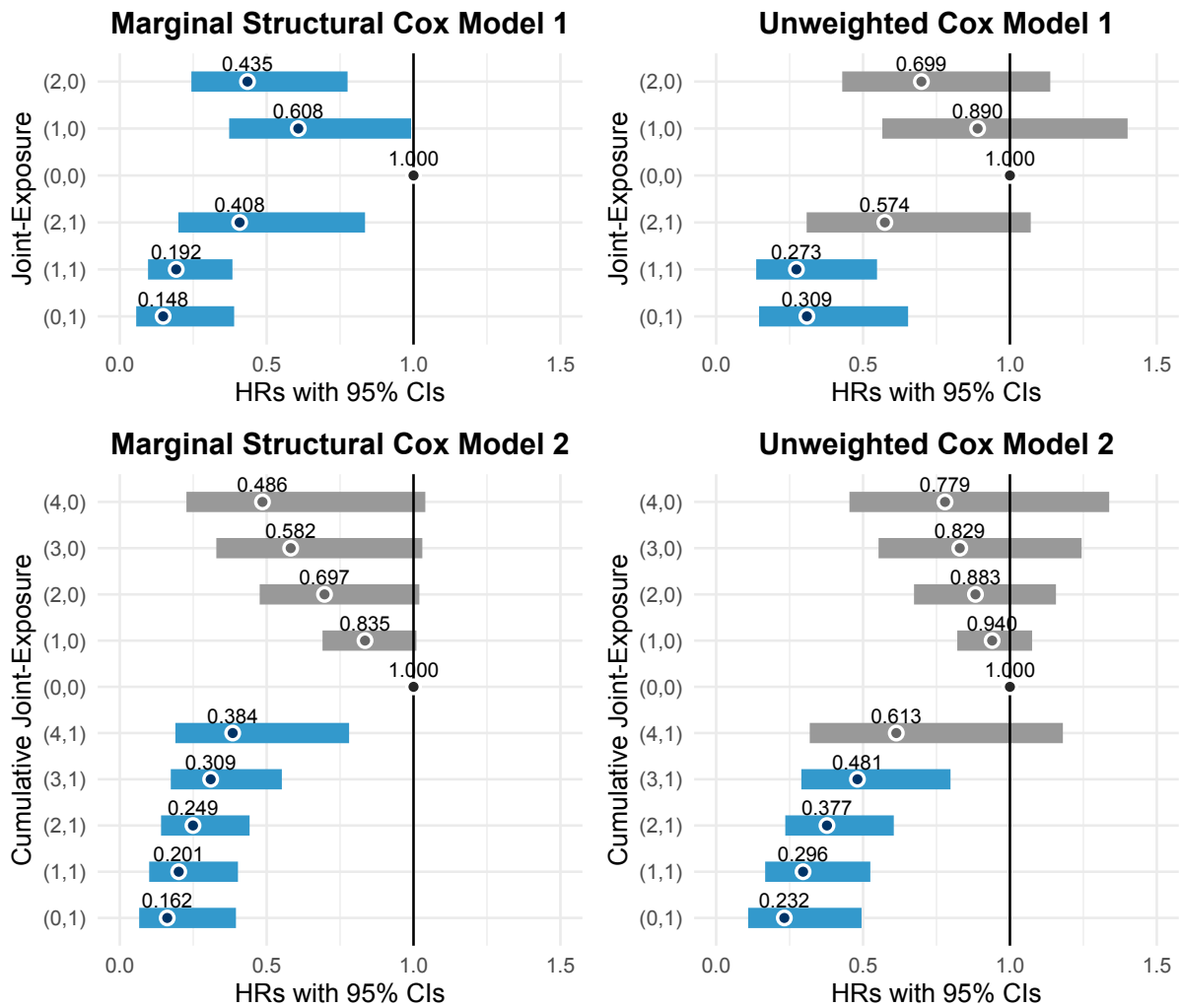
## 7. Causal effects of joint-exposure on survival in presence of time-varying confounders

**Table 7.4.** Estimated parameters  $\hat{\beta}$  along with their 95% Confidence Intervals (CIs) for Cox MSMs 1 and 2 in Equation 7.6 and 7.10, respectively, and for their corresponding unweighted versions.

Treatment	Cox MSM 1		Unweighted Cox model 1	
	$\hat{\beta}$	95% CIs	$\hat{\beta}$	95% CIs
$a_1 = 1$	-0.498	[-0.986; -0.010]	-0.116	[-0.568; 0.335]
$a_1 = 2$	-0.833	[-1.409; -0.257]	-0.359	[-0.844; 0.127]
$a_2 = 1$	-1.914	[-2.880; -0.948]	-1.175	[-1.921; -0.429]
$a_1 = 1 \times a_2 = 1$	0.762	[-0.399; 1.923]	-0.006	[-0.997; 0.984]
$a_1 = 2 \times a_2 = 1$	1.850	[0.643; 3.057]	0.979	[0.020; 1.938]
Treatment	Cox MSM 2		Unweighted Cox model 2	
	$\hat{\beta}$	95% CIs	$\hat{\beta}$	95% CIs
cum ( $\bar{a}_1$ )	-0.181	[-0.370; 0.009]	-0.062	[-0.197; 0.072]
$a_2 = 1$	-1.823	[-2.714; -0.932]	-1.461	[-2.215; -0.707]
cum ( $\bar{a}_1$ ) $\times$ $a_2 = 1$	0.397	[0.052; 0.743]	0.305	[0.010; 0.601]

receiving a *high-reduction* experienced an event 86% faster ( $HR = 0.574/0.309 = 1.86$ ). However, these results were affected by the toxicity-treatment-adjustment bias and could not be interpreted in a causal way. In fact, the final value of RDI was the realisation of the treatment trajectory as result of both the severity of the overall toxicity experienced by each patient and the side-effects handling operated by physicians. To overcome these issues, Cox MSM 1 (top-left) represented a clear improvement with respect to its unweighted version. At the same final RDI level, a *good* response caused an 85.2% decrease in the risk of an event ( $\exp(\hat{\beta}_3) = 0.148$ ) with respect to a *poor* one. Reductions in the final RDI caused better EFS for PRs, whereas a reverse causal association was founded in GRs. In particular, the higher the final reduction, the better the survival for PRs (estimated HRs were 0.608 and 0.435 for *low* and *high* reduction PRs, respectively). On the contrary, the higher the final reduction, the worsen the survival for GRs: GRs with *low* or *high* reduction experienced an event 1.30 ( $HR = 0.192/0.148 = 1.30$ ) or 2.76 ( $HR = 0.408/0.148 = 2.76$ ) times faster than GRs with *normal*-RDI.

Bottom panels refers to the causal structure of DAG-2 presented in Section 7.2.5, where reference level was PRs without reduction, neither pre nor post surgery, i.e., (cum ( $\bar{a}_1$ ),  $a_2$ ) is (0,0). Results were in line with previous results: (i) GRs presented better survival with respect to PRs; (ii) an increasing number of pre/post-operative reductions in RDI showed opposite trends for PRs and GRs, improving and worsening EFS, respectively. This was even more evident in the Cox MSM 2 (bottom-left) than in its unweighted version (bottom-right) affected by the *toxicity-treatment-adjustment* bias: point estimates with respect to reference level dramatically improved even if statistical significance did not change, again showing the bias due to *toxicity-treatment-adjustment*. Considering parameter estimates for Cox MSM 2 (see Table 7.4), at the same RDI level, a *good* response caused an 83.8% decrease in the risk of an event ( $\exp(\hat{\beta}_3) = 0.162$ ) with respect to a *poor* one. Moreover, 1-unit increase in the number of reductions of 15-30% (i.e., 1-unit increase in cum ( $\bar{a}_1$ )) caused a decrease of 16.5% in the risk of an event for PRs ( $\exp(\hat{\beta}_1) = 0.835$ ) and an increase of 24.1% for GRs ( $\exp(\hat{\beta}_1 + \hat{\beta}_3) = 1.241$ ).



**Figure 7.11.** Graphical displays of Hazard Ratios (HRs) along with their 95% Confidence Intervals (CIs) for Marginal Structural Cox Models (left panels) and corresponding unweighted Cox models (right panels). Top panels refer to Cox MSM 1 in Equation (7.6), where reference level is *poor* responder with *normal* RDI level at the end of treatment (i.e.,  $(a_1, a_2) = (0, 0)$ ). Bottom panels refer to Cox MSM 2 in Equation (7.10), where reference level is *poor* responder without reduction, neither pre nor post surgery, i.e.,  $(\text{cum}(\bar{a}_1), a_2) = (0, 0)$ .

One possible clinical explanation for these reverse behaviours could lie in the fact that chemotherapy also damages non-cancerous cells and processes of the immune system that can detect and kill cancer cells. In PRs, for whom chemotherapy is less effective, this negative effect is not offset by treatment efficacy, and an increase in RDI may be detrimental to survival due to the impact on the immune system.

## 7.4. Final remarks

In cancer trials, longitudinal chemotherapy data are problematic to analyse due to the presence of negative feedback between exposure to cytotoxic drugs and consequent toxic side effects. Therapy administration is usually complicated by the dynamical adjustment of the treatment based on patients' clinical picture, especially on chemotherapy-induced



## 7. Causal effects of joint-exposure on survival in presence of time-varying confounders

multi-systemic toxicities. For this reason, chemotherapy is usually modelled by Intention-To-Treat (ITT) analysis [70], although the introduction of the Received Dose Intensity (RDI) [86] marked a significant departure from ITT in the direction of a closer description of the actual clinical practice. The main issue in analysing actual treatment lies in the fact that toxicities act as time-dependent confounders for the effect of chemotherapy intensity exposure on survival, determining the *toxicity-treatment-adjustment* bias if not properly considered. Suitable methodologies are hence needed to control for exposure-affected (time-varying) toxicity confounding in longitudinal chemotherapy data. In addition, since the assignment of dose reductions/interruptions or delays in administration during treatment is determined not by individual toxicities but by the overall toxic burden of each patient, the different types and number of side effects must be adequately summarized to be included in the analysis.

Motivated by a sharp yet delicate clinical question on the effect of treatment modifications on Event-Free Survival (EFS) in osteosarcoma patients, this chapter proposed Marginal Structural Models (MSMs) in combination with Inverse-Probability-of-Treatment Weighted (IPTW) estimators to assess the causal effects of chemotherapy intensity exposure seen in terms of both Histological Response (HRe) and RDI reductions compared to protocol. Control arms data from BO03 and BO06 trials for osteosarcoma were analyzed. Since only the most severe side-effects were recorded in BO03, the analysis of such mixed longitudinal/non-longitudinal data required both an original analytical strategy and an unconventional model formulation. First, pre and post-operative toxicity data were summarized using a Multiple Overall Toxicity (MOTox) approach [190] based on most severe CTACE grades of both *rule-specific* and *generic* side effects. This allowed (i) to reduce the number of possible confounders combinations, avoiding non-positivity and highly-correlated data, and (ii) to meet the clinical rationale of tailoring treatment according to the patient's overall toxic burden in the case of multiple toxic side effects. Then, two different joint-exposure characterizations – which met *consistency* according to experts – were defined unambiguously based on *time-fixed final RDI* or *time-dependent pre/post-operative RDI* combined with HRe. This led to the introduction of two alternative Direct Acyclic Graphs (DAGs) to identify all possible (time-dependent) confounders and their relationships with both joint-exposure and EFS outcome, validating the assumption of *no unmeasured confounding*. Suitable IPTW-based techniques and Cox MSMs, representing the causal RDI analogues of the ITT Cox model presented by Lewis *et al.* (2007) [119], were finally designed to mimic a randomized trial where the joint-exposure intensity was no longer confounded by toxicities. Once *positivity* and *no misspecification* were satisfied, in the pseudo-population thus created, a crude analysis sufficed to estimate the causal effect of joint-exposure modifications on EFS.

Regardless of RDI-level, in both Cox MSMs all estimated HRs were lower for *Good* Responders (GRs) than for *Poor* ones (PRs), showing that GRs presented a better EFS than PRs in all cases. This was not surprising because HRe is the strongest prognostic survival factor known to date in osteosarcoma [31]. Increasing RDI-reductions created two opposite trends for PRs and GRs: the higher the reduction in final or pre/post-operative RDI, the better (worsen) was the EFS in PRs (GRs). One possible clinical explanation for

these inverse behaviours could lie in the effect of chemotherapy on non-cancerous cells. By targeting a broad spectrum of cells, chemotherapy also damages the processes and mechanisms of the immune system that can detect and kill cancer cells. While in GRs this negative effect may be largely offset by the efficacy of the tumour therapy, in PRs – for whom chemotherapy is less effective – an increased RDI may be harmful to survival due to the impact on the immune system.

This study highlighted both the confounding nature of toxicity data and the detrimental effect of not considering them in the analysis, showing the potential pitfalls of a naive RDI-based analysis of chemotherapy data. When ITT models were translated into RDI-based ones by simply neglecting the role of toxicities as in the unweighted Cox models, results were clearly affected by the *toxicity-treatment-adjustment* bias. The use of Cox MSMs allowed to model the contribution of patient's toxicity history to EFS through the realisation of the (cumulative) joint-exposures. In other words, the use of IPTW-based Cox MSMs broke the feedback between side effects and therapy adjustments, resulting in unbiased estimates of the effect of treatment modifications on EFS and describing better the effect of low-intensity regimens.

The presented IPTW-based MSMs have clear limitations. The property of MSMs to give unbiased estimates relies on the four main assumptions presented in Section 7.2.2, which are often unverifiable and mostly based on experts knowledge. This is really the potential weakness of both the analysis presented above and the methodology based on IPTW and MSMs in general. In addition, the lack of longitudinal confounders in BO03 forced the causal structures represented by the DAGs in Figure 7.7. These DAGs relied on the assumption that the most severe CTCAE grades of each toxicity in pre- and post-operative treatment predicted well the final and pre/post-operative RDI values, thus flattening the toxicity history. This assumption might still be challenged, since severe toxicities might look simultaneous producing some significant interactions. However, there is no guarantee that severe CTCAE grades occurred simultaneously, so these interactions were not considered. The development of appropriate causal structures and methodologies for studying chemotherapy data using a cycle-by-cycle longitudinal perspective would be of great interest for future analyses, as it would overcome this issue, but the need for adequate toxicity data collection still remains.

In summary, this chapter showed the difficulty of analysing chemotherapy data on a RDI-based approach, mostly originated from data quality. The main contribution of this work is the presentation of an all-round analysis of complex chemotherapy data, with tutorial-like explanations of the difficulties encountered and the problem-solving strategies deployed. Focusing on a way of analysing chemotherapy data that is RDI-based rather than ITT-based, it illustrated the key role played by toxicities in this transition and showed the detrimental effect of neglecting them. To the best of our knowledge, this is the first application of IPTW-based techniques to survival data from a randomized trial of chemotherapy in order to eliminate the *toxicity-treatment-adjustment* bias.

## D. Appendix to Chapter 7

### D.1. Pre/Post-operative Received Dose Intensity definitions

For each patient  $i$ , let  $np_i$  denote the number of cycles performed before the surgery. *Pre-operative* period is made up of cycles performed before surgery, i.e.,  $j \in \{1, \dots, np_i\}$ . *Post-operative* period is made up of cycles performed after surgery, i.e.,  $j \in \{np_i+1, \dots, 6\}$ . To consider the two periods separately, definitions in Equations (7.2), (7.3) and (7.4) in Section 7.1.3 can be adapted as follows.

**Pre/Post-operative standardized dose.** The pre-operative and post-operative standardized doses  $\Delta_{i,pre}$  and  $\Delta_{i,post}$  for patient  $i$  are defined as

$$\Delta_{i,pre} = \frac{1}{2} (\Delta_{i,pre}^{CDDP} + \Delta_{i,pre}^{DOX}) = \frac{1}{2 \cdot np_i} \left( \sum_{j=1}^{np_i} \delta_{ij}^{CDDP} + \sum_{j=1}^{np_i} \delta_{ij}^{DOX} \right),$$

$$\Delta_{i,post} = \frac{1}{2} (\Delta_{i,post}^{CDDP} + \Delta_{i,post}^{DOX}) = \frac{1}{2 \cdot (5 - np_i)} \left( \sum_{j=np_i+1}^6 \delta_{ij}^{CDDP} + \sum_{j=np_i+1}^6 \delta_{ij}^{DOX} \right).$$

**Pre/Post-operative standardized time.** The pre-operative standardized time for the  $i$ -th patient is

$$\Gamma_{i,pre} = \frac{\text{actual pre-operative time}}{\text{anticipated pre-operative time}}$$

where

- *actual pre-operative time* is the difference in days between the starting date of cycle 1 and the date of the surgery,
- *anticipated pre-operative time* is  $21 \times np_i$  days, i.e.,  $np_i$  cycles lasting 21 days each.

Similarly, the post-operative standardized time for patient  $i$  is

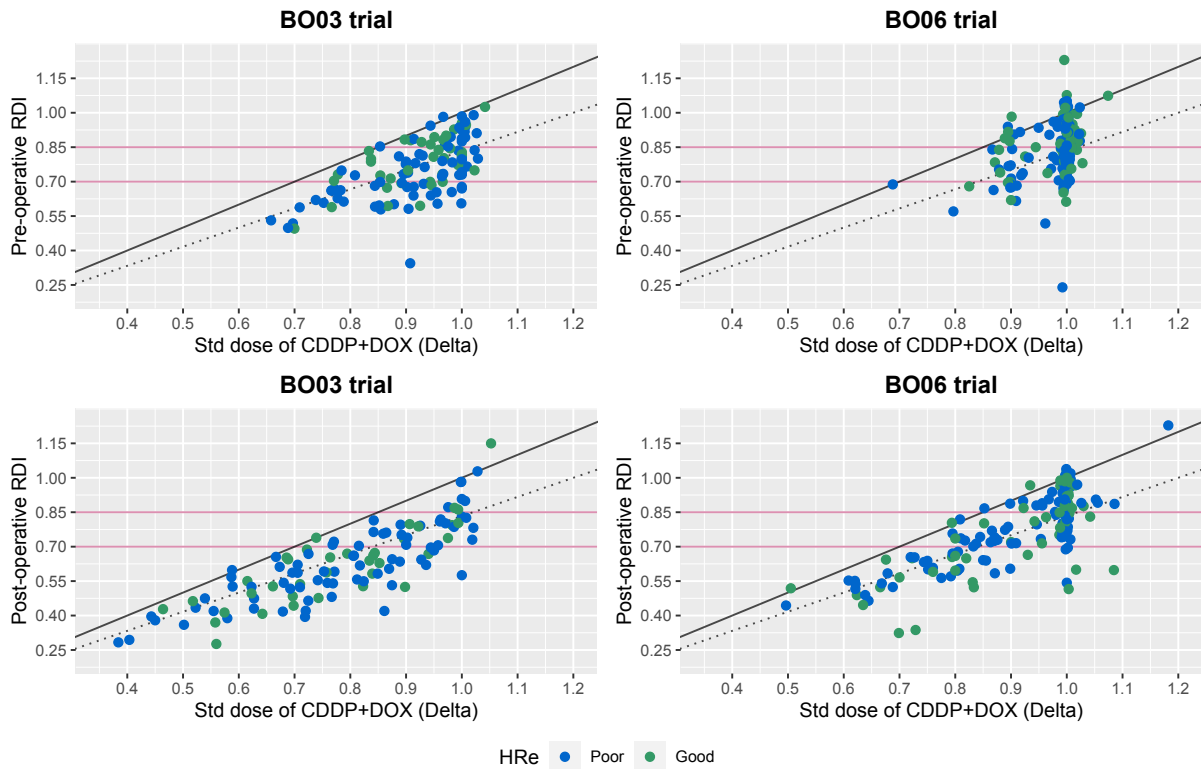
$$\Gamma_{i,post} = \frac{\text{actual post-operative time}}{\text{anticipated post-operative time}}$$

where

- *actual post-operative time* is the difference in days between the surgery date and the 3rd day after the start of cycle 6,
- *anticipated post-operative time* is  $14 + (5 - np_i) \times 21 + 3$  days, i.e., 14 days of surgery,  $5 - np_i$  cycles lasting 21 days each and 3 days after the start of cycle 6.

**Pre/Post-operative Received Dose Intensity.** The pre-operative and post-operative Received Dose Intensities (RDIs) for patient  $i$  are defined as

$$RDI_{i,pre} = \frac{\Delta_{i,pre}}{\Gamma_{i,pre}}, \quad (7.13)$$



**Figure 7.12.** Top panels: Scatter plots of pre-operative RDI (i.e.,  $RDI_{i,pre}$  in Equation (7.13)) against pre-operative standardized dose of CDDP+DOX ( $\Delta_{i,pre}$ ) conditional on trial (*BO03*: left panel; *BO06*: right panel) and HRe (blue: *poor*; green: *good*).

Bottom panels: Scatter plots of post-operative RDI (i.e.,  $RDI_{i,post}$  in Equation (7.14)) against post-operative standardized dose of CDDP+DOX ( $\Delta_{i,post}$ ) conditional on trial (*BO03*: left panel; *BO06*: right panel) and HRe (blue: *poor*; green: *good*).

$$RDI_{i,post} = \frac{\Delta_{i,post}}{\Gamma_{i,post}}. \quad (7.14)$$

The RDI computed on the whole treatment as in Equation (7.4) is not the sum of pre-operative and post-operative RDIs, i.e.,  $RDI_i \neq RDI_{i,pre} + RDI_{i,post}$ .

A summary of  $RDI_{i,pre}$  and  $RDI_{i,post}$  exposure characteristics for the whole cohort and conditional on trials is reported in Figure 7.6. Figure 7.12 shows the scatter plots of pre- (top panels) and post- (bottom panels) operative RDI ( $RDI_{i,k}$ ) against their relative standardized doses of CDDP+DOX ( $\Delta_{i,k}$ ) conditional on trial (left panel: *BO03*; right panel: *BO06*) and HRe (blue: *poor*; green: *good*). The solid horizontal lines in pink vertically divide patients with *normal* RDI levels ( $RDI_{i,k} \geq 0.85$ ) from *low* reduction ( $0.70 \leq RDI_{i,k} < 0.85$ ) and *high* reduction ( $RDI_{i,k} < 0.70$ ) patients. The solid diagonal line in black satisfies equation  $RDI_{i,k} = \Delta_{i,k}$ , dividing the group of patients with standardized time  $\Gamma_{i,k} > 1$  (delayed therapy, below the line) from the group, almost void, of patients with  $\Gamma_{i,k} < 1$  (anticipated therapy, above the line). The dotted diagonal line in black satisfies equation  $RDI_{i,k} = \Delta_{i,k}/1.2$ , dividing the group of patients with therapy delayed by more than 20% of anticipated time (below the dotted line) from the group of patients with therapy delayed by less than 20% of anticipated time (between solid and dotted black lines). This figure clearly shows the difference of treatment delivery in *BO03*

vs. BO06, also considering pre/post-operative periods separately. It also shows the lack of a clear association between HRe and pre/post-operative RDI.

## D.2 Denominator specifications for selected IPTW models

In Section 7.3.2, different model specifications (in terms of confounding covariates  $\mathbf{L}$  for the denominators) to determine the subject-specific standardized weights for final RDI level  $SW_i^{A^1}$ , HRe category  $SW_i^{A^2}$  and pre/post-operative RDI levels  $SW_i^{A^1} = SW_i^{A^1_{pre}}$ .  $SW_i^{A^1_{post}}$  were investigated. The following denominator formulas were selected:

- multinomial logistic regression model for denominator of final RDI level  $SW_i^{A^1}$ :

$$\begin{aligned} \log \frac{\Pr(A_i^1 = a | \mathbf{L}_i^1)}{\Pr(A_i^1 = 0 | \mathbf{L}_i^1)} &= \gamma_{a0} + \gamma_{a1} \cdot \mathbb{1}_{(\text{trial}_i = \text{BO06})} + \gamma_{a2} \cdot \mathbb{1}_{(\text{age}_i = \text{adolescent})} + \\ &\quad \gamma_{a3} \cdot \mathbb{1}_{(\text{age}_i = \text{adult})} + \gamma_{a4} \cdot \mathbb{1}_{(\text{gender}_i = \text{male})} + \\ &\quad \gamma_{a5} \cdot \text{MOTOx}_{i,pre}^{(gen)} + \gamma_{a6} \cdot \text{MOTOx}_{i,pre}^{(rule)} + \\ &\quad \gamma_{a7} \cdot \text{MOTOx}_{i,post}^{(gen)} + \gamma_{a8} \cdot \text{MOTOx}_{i,post}^{(rule)} \end{aligned}$$

where confounding covariates are

$$\begin{aligned} \mathbf{L}_i^1 &= (\mathbf{L}_i^{base}, \mathbf{L}_{i,pre}^{tox}, \mathbf{L}_{i,post}^{tox}) = \\ &= (\text{trial}_i, \text{age}_i, \text{gender}_i, \text{MOTOx}_{i,pre}^{(gen)}, \text{MOTOx}_{i,pre}^{(rule)}, \text{MOTOx}_{i,post}^{(gen)}, \text{MOTOx}_{i,post}^{(rule)}); \end{aligned}$$

- binary logistic regression model for denominator of HRe category  $SW_i^{A^2}$

$$\begin{aligned} \log \frac{\Pr(A_i^2 = 1 | \mathbf{L}_i^2)}{1 - \Pr(A_i^2 = 1 | \mathbf{L}_i^2)} &= \alpha_0 + \alpha_1 \cdot \mathbb{1}_{(\text{trial}_i = \text{BO06})} + \alpha_2 \cdot \mathbb{1}_{(\text{age}_i = \text{adolescent})} + \\ &\quad \alpha_3 \cdot \mathbb{1}_{(\text{age}_i = \text{adult})} + \alpha_4 \cdot \mathbb{1}_{(\text{gender}_i = \text{male})} + \\ &\quad \alpha_5 \cdot \text{MOTOx}_{i,pre}^{(gen)} + \alpha_6 \cdot \text{MOTOx}_{i,pre}^{(rule)} + \\ &\quad \alpha_7 \cdot \mathbb{1}_{(\text{surgery}_i = \text{on time})} \end{aligned}$$

where confounding covariates are

$$\begin{aligned} \mathbf{L}_i^2 &= (\mathbf{L}_i^{base}, \mathbf{L}_{i,pre}^{tox}, L_i^{surg}) = \\ &= (\text{trial}_i, \text{age}_i, \text{gender}_i, \text{MOTOx}_{i,pre}^{(gen)}, \text{MOTOx}_{i,pre}^{(rule)}, \text{surgery}_i); \end{aligned}$$

- multinomial logistic regression model for denominator of pre-operative RDI level  $SW_i^{A^1_{pre}}$ :

$$\begin{aligned} \log \frac{\Pr(A_{i,pre}^1 = a | \mathbf{L}_{i,pre}^1)}{\Pr(A_{i,pre}^1 = 0 | \mathbf{L}_{i,pre}^1)} &= \gamma_{a0} + \gamma_{a1} \cdot \mathbb{1}_{(\text{trial}_i = \text{BO06})} + \gamma_{a2} \cdot \mathbb{1}_{(\text{age}_i = \text{adolescent})} + \\ &\quad \gamma_{a3} \cdot \mathbb{1}_{(\text{age}_i = \text{adult})} + \gamma_{a4} \cdot \mathbb{1}_{(\text{gender}_i = \text{male})} + \\ &\quad \gamma_{a5} \cdot \text{MOTOx}_{i,pre}^{(gen)} + \gamma_{a6} \cdot \text{MOTOx}_{i,pre}^{(rule)} \end{aligned}$$

where confounding covariates are

$$\mathbf{L}_{i,pre}^1 = (\mathbf{L}_i^{base}, \mathbf{L}_{i,pre}^{tox}) = \left( \text{trial}_i, \text{age}_i, \text{gender}_i, \text{MOTox}_{i,pre}^{(gen)}, \text{MOTox}_{i,pre}^{(rule)} \right);$$

- multinomial logistic regression model for denominator of post-operative RDI level  $SW_i^{A_{i,post}^1}$ :

$$\begin{aligned} \log \frac{\Pr(A_{i,post}^1 = a | \bar{\mathbf{L}}_{i,post}^1, A_{i,pre}^1)}{\Pr(A_{i,post}^1 = 0 | \bar{\mathbf{L}}_{i,post}^1, A_{i,pre}^1)} &= \gamma_{a0} + \gamma_{a1} \cdot \mathbb{1}_{(\text{trial}_i = BO06)} + \gamma_{a2} \cdot \mathbb{1}_{(\text{age}_i = \text{adolescent})} + \\ &\quad \gamma_{a3} \cdot \mathbb{1}_{(\text{age}_i = \text{adult})} + \gamma_{a4} \cdot \mathbb{1}_{(\text{gender}_i = \text{male})} + \\ &\quad \gamma_{a5} \cdot \text{MOTox}_{i,pre}^{(gen)} + \gamma_{a6} \cdot \text{MOTox}_{i,pre}^{(rule)} + \\ &\quad \gamma_{a7} \cdot \mathbb{1}_{(A_{i,pre}^1 = 1)} + \gamma_{a8} \cdot \mathbb{1}_{(A_{i,pre}^1 = 2)} + \\ &\quad \gamma_{a9} \cdot \text{MOTox}_{i,post}^{(gen)} + \gamma_{a10} \cdot \text{MOTox}_{i,post}^{(rule)}. \end{aligned}$$

where  $A_{i,pre}^1$  is the pre-operative RDI level and confounding covariates are

$$\begin{aligned} \bar{\mathbf{L}}_{i,post}^1 &= (\mathbf{L}_i^{base}, \mathbf{L}_{i,pre}^{tox}, \mathbf{L}_{i,post}^{tox}) \\ &= \left( \text{trial}_i, \text{age}_i, \text{gender}_i, \text{MOTox}_{i,pre}^{(gen)}, \text{MOTox}_{i,pre}^{(rule)}, \text{MOTox}_{i,post}^{(gen)}, \text{MOTox}_{i,post}^{(rule)} \right). \end{aligned}$$



# 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.

# Bibliography

- [1] O. Aalen. Nonparametric inference for a family of counting processes. *Annals of Statistics*, 6(4):701–726, 1978.
- [2] O. O. Aalen, R. J. Cook, and K. Røysland. Does Cox analysis of a randomized survival study yield a causal treatment effect? *Lifetime data analysis*, 21(4):579–593, 2015.
- [3] AIFA – Agenzia Italiana del Farmaco. *Accesso al farmaco, Autorizzazione dei farmaci*. <https://www.aifa.gov.it/autorizzazione-dei-farmaci>
- [4] AIFA – Agenzia Italiana del Farmaco. *Banca Dati Farmaci, Ricerca per Principio Attivo*. <https://farmaci.agenziafarmaco.gov.it/bancadatifarmaci/cerca-per-principio-attivo> (2018-09-08).
- [5] AIFA – Agenzia Italiana del Farmaco. *Linee guida per la classificazione e conduzione degli studi osservazionali sui farmaci*. [http://www.agenziafarmaco.gov.it/sites/default/files/det\\_20marzo2008.pdf](http://www.agenziafarmaco.gov.it/sites/default/files/det_20marzo2008.pdf)
- [6] J. Aitchison. *The Statistical Analysis of Compositional Data*. Chapman & Hall Ltd, 1986.
- [7] H. Akaike. Information Theory and an Extension of the Maximum Likelihood Principle. In B. N. Petrov and F. Csaki, editors, *Proceedings of the 2nd International Symposium on Information Theory*, pages 267–281. Budapest: Akademiai Kiado, 1973.
- [8] H. S. Al-malky, S. E. A. Harthi, and A.-M. M. Osman. Major obstacles to doxorubicin therapy: Cardiotoxicity and drug resistance. *Journal of Oncology Pharmacy Practice*, 26(2):434–444, 2020.
- [9] S. A. Aldossary. Review on pharmacology of cisplatin: Clinical use, toxicity and mechanism of resistance of cisplatin. *Biomedical and Pharmacology Journal*, 12(1):07–15, 2019.
- [10] V. Allan, S. V. Ramagopalan, J. Mardekian, *et al.* Propensity score matching and inverse probability of treatment weighting to address confounding by indication in comparative effectiveness research of oral anticoagulants. *Journal of comparative effectiveness research*, 9(9):603–614, 2020.
- [11] L. D. A. F. Amorim and J. Cai. Modelling recurrent events: a tutorial for analysis in epidemiology. *International Journal of Epidemiology*, 44(1):324–333, 2014.
- [12] P. K. Andersen and R. D. Gill. Cox’s regression model for counting processes: A large sample study. *The Annals of Statistics*, 10(4):1100–1120, 1982.
- [13] P. K. Andersen and N. Keiding. Multi-state models for event history analysis. *Statistical Methods in Medical Research*, 11(2):91–115, 2002.
- [14] S. E. Andrade, K. H. Kahler, F. Frech, and K. A. Chan. Methods for evaluation of medication adherence and persistence using automated databases. *Pharmacoepidemiology & Drug Safety*, 15(8):565–574, 2006.
- [15] J. K. Anninga, H. Gelderblom, M. Fiocco, *et al.* Chemotherapeutic adjuvant treatment for osteosarcoma: Where do we stand? *European Journal of Cancer*, 47(16):2431–2445, 2011.

## Bibliography

- [16] M. W. Arisido, L. Antolini, D. Bernasconi, *et al.* Joint model robustness compared with the time-varying covariate Cox model to evaluate the association between a longitudinal marker and a time-to-event endpoint. *BMC Med Res Methodol*, 19:222, 2019.
- [17] I. Arnet, M. Greenland, M. W. Knuiman, *et al.* Operationalization and validation of a novel method to calculate adherence to polypharmacy with refill data from the australian pharmaceutical benefits scheme (PBS) database. *Clinical Epidemiology*, 10:1181–1194, 2018.
- [18] P. C. Austin, A. Latouche, and J. P. Fine. A review of the use of time-varying covariates in the Fine-Gray subdistribution hazard competing risk regression model. *Statistics in Medicine*, 39(2):103–113, 2020.
- [19] P. C. Austin and E. A. Stuart. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Statistics in Medicine*, 34(28):3661–3679, 2015.
- [20] S. Bacci, S. Pandolfi, and F. Pennoni. A Comparison of Some Criteria for States Selection in the Latent Markov Model for Longitudinal Data. *Advances in Data Analysis and Classification*, 8(2):125–145, 2014.
- [21] S. Baraldo, F. Ieva, A. M. Paganoni, and V. Vitelli. Outcome prediction for heart failure telemonitoring via generalized linear models with functional covariates. *Scandinavian Journal of Statistics*, 40(3):403–416, 2013.
- [22] F. Bartolucci, A. Farcomeni, and F. Pennoni. *Latent Markov Models for Longitudinal Data*. Chapman & Hall/CRC, Boca Raton, 2013.
- [23] F. Bartolucci, A. Farcomeni, and F. Pennoni. Latent Markov models: a review of a general framework for the analysis of longitudinal data with covariates. *TEST*, 23:433–465, 2014.
- [24] F. Bartolucci, M. Lupporelli, and G. E. Montanari. Latent Markov model for longitudinal binary data: An application to the performance evaluation of nursing homes. *The Annals of Applied Statistics*, 3(2):611 – 636, 2009.
- [25] F. Bartolucci, G. E. Montanari, and S. Pandolfi. Three-step estimation of latent Markov models with covariates. *Computational Statistics & Data Analysis*, 83:287–301, 2015.
- [26] F. Bartolucci, S. Pandolfi, and F. Pennoni. LMest: An R Package for Latent Markov Models for Longitudinal Categorical Data. *Journal of Statistical Software*, 81(4):1–38, 2017.
- [27] P. C. Baumgartner, R. B. Haynes, K. E. Hersberger, and I. Arnet. A systematic review of medication adherence thresholds dependent of clinical outcomes. *Frontiers in pharmacology*, 9:1290, 2018.
- [28] B. N. Bekele and P. F. Thall. Dose-Finding Based on Multiple Toxicities in a Soft Tissue Sarcoma Trial. *Journal of the American Statistical Association*, 99(465):26–35, 2004.
- [29] M. J. Bijlsma, F. Janssen, and E. Hak. Estimating time-varying drug adherence using electronic records: extending the proportion of days covered (PDC) method. *Pharmacoepidemiology & Drug Safety*, 25(3):325–332, 2016.
- [30] D. A. Binder. Fitting cox’s proportional hazards models to survey data. *Biometrika*, 79:139–147, 1992.
- [31] M. W. Bishop, Y. C. Chang, M. D. Krailo, *et al.* Assessing the Prognostic Significance of Histologic Response in Osteosarcoma: A Comparison of Outcomes on CCG-782 and INT0133-A Report From the Children’s Oncology Group Bone Tumor Committee. *Pediatric blood & cancer*, 63(10):1737–1743, 2016.
- [32] N. E. Breslow. Analysis of survival data under the proportional hazards model. *International Statistical Review/Revue Internationale de Statistique*, pages 45–57, 1975.

- [33] E. R. Brown, J. G. Ibrahim, and V. DeGruttola. A flexible b-spline model for multiple longitudinal biomarkers and survival. *Biometrics*, 61(1):64–73, Mar. 2005.
- [34] J. Cai and D. E. Schaube. Marginal Means/Rates Models for Multiple Type Recurrent Event Data. *Lifetime Data Analysis*, 10:121–138, 2004.
- [35] M. Carhini, M. Suárez-Fariñas, and R. G. Maki. A Method to Summarize Toxicity in Cancer Randomized Clinical Trials. *Clinical Cancer Research*, 24(20):4968–4975, 2018.
- [36] A. Charles-Nelson, S. Katsahian, and C. Schramm. How to analyze and interpret recurrent events data in the presence of a terminal event: An application on readmission after colorectal cancer surgery. *Statistics in Medicine*, 38(18):3476–3502, 2019.
- [37] X. Chen, Q. Wang, J. Cai, and V. Shankar. Semiparametric additive marginal regression models for multiple type recurrent events. *Lifetime data analysis*, 18(4):504–527, 2012.
- [38] Y. Chi and J. G. Ibrahim. Joint Models for Multivariate Longitudinal and Multivariate Survival Data. *Biometrics*, 62(2):432–445, 2006.
- [39] P. J. Clare, T. A. Dobbins, and R. P. Mattick. Causal models adjusting for time-varying confounding—a systematic review of the literature. *International journal of epidemiology*, 48(1):254–265, 2019.
- [40] S. R. Cole and C. E. Frangakis. The Consistency Statement in Causal Inference. *Epidemiology*, 20(3-5), 2009.
- [41] S. R. Cole and M. A. Hernán. Constructing Inverse Probability Weights for Marginal Structural Models. *American Journal of Epidemiology*, 168(6):656–664, 2008.
- [42] L. M. Collins and S. Lanza. *Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences*. John Wiley and Sons Inc, 2010.
- [43] M. Collins, M. Wilhelm, R. Conyers, *et al.* Benefits and Adverse Events in Younger Versus Older Patients Receiving Neoadjuvant Chemotherapy for Osteosarcoma: Findings From a Meta-Analysis. *Journal of Clinical Oncology*, 31(18):2303–2312, 2013.
- [44] R. J. Cook and J. F. Lawless. *The Statistical Analysis of Recurrent Events*. New York: Springer, 2007.
- [45] G. Corrao and G. Mancia. Generating evidence from computerized healthcare utilization databases. *Hypertension*, 65:490–498, 2015.
- [46] D. Cox. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B*, 34(2):187–220, 1972.
- [47] D. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes*. Springer-Verlag New York, 2nd edition, 2003.
- [48] R. M. Daniel, S. N. Cousens, B. L. De Stavola, *et al.* Methods for dealing with time-dependent confounding. *Statistics in medicine*, 32(9):1584–1618, 2013.
- [49] E. Dantan, P. Joly, J.-F. Dartigues, and H. Jacqmin-Gadda. Joint model with latent state for longitudinal and multistate data. *Biostatistics*, 12(4):723–736, 2011.
- [50] K. Dickstein, A. Cohen-Solal, G. Filippatos, *et al.* ESC guidelines for the diagnosis and treatment of acute and chronic heart failure 2008: the task force for the diagnosis and treatment of acute and chronic heart failure 2008 of the european society of cardiology. developed in collaboration with the heart failure association of the ESC (HFA) and endorsed by the european society of intensive care medicine (ESICM). *European Journal of Heart Failure*, 10(10):933–989, 2008.

## Bibliography

- [51] N. A. G. dos Santos, R. S. Ferreira, and A. C. dos Santos. Overview of cisplatin-induced neurotoxicity and ototoxicity, and the protective agents. *Food and Chemical Toxicology*, 136:111079, 2020.
- [52] L. Duchateau, P. Janssen, I. Kezic, and C. Fortpiet. Evolution of recurrent asthma event rate over time in frailty models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 52(3):355–363, 2003.
- [53] M. Elseviers, M. Andersen, A. B. Almarsdottir, *et al.* *Drug utilization research*. John Wiley & Sons, Nashville, TN, 2016.
- [54] T. Emura, M. Nakatochi, K. Murotani, and V. Rondeau. A joint frailty-copula model between tumour progression and death for meta-analysis. *Statistical methods in medical research*, 26(6):2649–2666, 2017.
- [55] T. Fawcett. An introduction to ROC analysis. *Pattern Recognit. Lett.*, 27(8):861–874, June 2006.
- [56] F. Ferraty and P. Vieu. *Nonparametric functional data analysis: theory and practice*. Springer Series in Statistics. Springer New York, 2006.
- [57] G. Fitzmaurice, M. Davidian, G. Verbeke, and G. Molenberghs. *Longitudinal data analysis*. Chapman & Hall, CRC, London, 2009.
- [58] C. A. Forbes, S. Deshpande, F. Sorio-Vilela, *et al.* A systematic literature review comparing methods for the measurement of patient persistence and adherence. *Current Medical Research and Opinion*, 34(9):1613–1625, 2018.
- [59] F. Gasperoni, F. Ieva, G. Barbati, *et al.* Multi-state modelling of heart failure care path: A population-based investigation from italy. *PLoS One*, 12(6):e0179176, 2017.
- [60] F. Gasperoni, F. Ieva, A. Paganoni, *et al.* Non-parametric frailty cox models for hierarchical time-to-event data. *Biostatistics*, 12 2018.
- [61] E. Gayat, M. Arrigo, S. Littnerova, *et al.* Heart failure oral therapies at discharge are associated with better outcome in acute heart failure: a propensity-score matched study. *European Journal of Heart Failure*, 20(2):345–354, 2018.
- [62] J. E. Gellar, E. Colantuoni, D. M. Needham, and C. M. Crainiceanu. Cox regression models with functional covariates for survival data. *Statistical Modelling*, 15(3):256–278, 2015.
- [63] A. Giardini, M. T. Martin, C. Cahir, *et al.* Toward appropriate criteria in medication adherence assessment in older persons: Position paper. *Aging Clinical and Experimental Research*, 28(3):371–381, 2016.
- [64] J. R. González, E. Fernandez, V. Moreno, *et al.* Sex differences in hospital readmission among colorectal cancer patients. *Journal of epidemiology and community health*, 59(6):506–511, 2005.
- [65] L. A. Gould, M. E. Boye, M. J. Crowther, *et al.* Joint modeling of survival and longitudinal non-survival data: current methods and issues. Report of the DIA Bayesian joint modeling working group. *Statistics in Medicine*, 34(14):2181–2195, 2015.
- [66] E. Graf, C. Schmoor, W. Sauerbrei, and M. Schumacher. Assessment and comparison of prognostic classification schemes for survival data. *Statistics in Medicine*, 18(17-18):2529–2545, 1999.
- [67] S. Greenland, J. Pearl, and J. M. Robins. Causal diagrams for epidemiologic research. *Epidemiology*, 10(1):37–48, 1999.
- [68] R. W. Gregg, J. M. Molepo, V. J. Monpetit, *et al.* Cisplatin neurotoxicity: the relationship between dosage, time, and platinum concentration in neurologic tissues, and morphologic evidence of toxicity. *Journal of Clinical Oncology*, 10(5):795–803, 1992.

- [69] T. Grimmsmann and W. Himmel. Discrepancies between prescribed and defined daily doses: a matter of patients or drug classes? *European Journal of Clinical Pharmacology*, 67(8):847–854, 2011.
- [70] S. K. Gupta. Intention-to-treat concept: A review. *Perspectives in clinical research*, 2(3):109–112, 2011.
- [71] H. Hao, L. Chen, D. Huang, *et al.* Meta-analysis of alkaline phosphatase and prognosis for osteosarcoma. *European Journal of Cancer Care*, 26(5):e12536, 2017.
- [72] C. Happ and S. Greven. Multivariate Functional Principal Component Analysis for Data Observed on Different (Dimensional) Domains. *Journal of the American Statistical Association*, 113(522):649–659, 2018.
- [73] X. He and P. Ng. Cobs: Qualitatively constrained smoothing via linear programming. *Computational Statistics*, 14(3):315–338, 1999.
- [74] P. J. Heagerty and Y. Zheng. Survival model predictive accuracy and ROC curves. *Biometrics*, 61(1):92–105, 2005.
- [75] P. A. Heidenreich, B. Bozkurt, D. Aguilar, *et al.* 2022 AHA/ACC/HFSA guideline for the management of heart failure: A report of the american college of Cardiology/American heart association joint committee on clinical practice guidelines. *Circulation*, 145(18):e895–e1032, 2022.
- [76] R. Henderson, P. Diggle, and A. Dobson. Joint modelling of longitudinal measurements and event time data. *Biostatistics*, 1(4):465–480, 2000.
- [77] M. Hernán and J. Robins. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC, 2020.
- [78] M. A. Hernán, B. Brumback, and J. M. Robins. Marginal Structural Models to Estimate the Causal Effect of Zidovudine on the Survival of HIV-Positive Men. *Epidemiology*, 11(5):561–570, 2000.
- [79] M. A. Hernán, B. Brumback, and J. M. Robins. Marginal Structural Models to Estimate the Joint Causal Effect of Nonrandomized Treatments. *Journal of the American Statistical Association*, 96(454):440–448, 2001.
- [80] L. M. Hess, M. A. Raebel, D. A. Conner, and D. C. Malone. Measurement of adherence in pharmacy administrative databases: a proposal for standard definitions and preferred measures. *Annals of Pharmacotherapy*, 40(7-8):1280–1288, 2006.
- [81] G. L. Hickey, P. Philipson, A. Jorgensen, and R. Kolamunnage-Dona. Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues. *BMC Medical Research Methodology*, 16(117), 2016.
- [82] G. L. Hickey, P. Philipson, A. Jorgensen, and R. Kolamunnage-Dona. A comparison of joint models for longitudinal and competing risks data, with application to an epilepsy drug randomized controlled trial. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 181(4):1105–1123, 2018.
- [83] G. L. Hickey, P. Philipson, A. Jorgensen, and R. Kolamunnage-Dona. Joint models of longitudinal and time-to-event data with more than one event time outcome: A review. *The International Journal of Biostatistics*, 14(1), 2018.
- [84] A. Hirakawa, K. Sudo, K. Yonemori, *et al.* A Comparative Study of Longitudinal Toxicities of Cytotoxic Drugs, Molecularly Targeted Agents, Immunomodulatory Drugs, and Cancer Vaccines. *Clinical Pharmacology & Therapeutics*, 106(4):803–809, 2019.
- [85] P. Hougaard. Frailty models for survival data. *Lifetime Data Analysis*, 1:255–273, 1995.



## Bibliography

- [86] W. M. Hryniuk and M. Goodyear. The calculation of received dose intensity. *Journal of Clinical Oncology*, 8(12):1935–1937, 1990.
- [87] S. A. Hunt, W. T. Abraham, M. H. Chin, *et al.* 2009 focused update incorporated into the ACC/AHA 2005 guidelines for the diagnosis and management of heart failure in adults: a report of the american college of cardiology Foundation/American heart association task force on practice guidelines: developed in collaboration with the international society for heart and lung transplantation. *Circulation*, 119(14):e391–479, 2009.
- [88] J. Ibrahim, M. Chen, and D. Sinha. *Bayesian Survival Analysis*. Springer Series in Statistics. Springer, New York, NY, 2001.
- [89] F. Ieva and F. Gasperoni. Discussion of the paper "statistical challenges of administrative and transaction data" by david j. han. *Journal of the Royal Statistical Society - Series A*, 181(3):591–592, 2018.
- [90] F. Ieva, C. Jackson, and L. Sharples. Multi-state modelling of repeated hospitalisation and death in patients with heart failure: the use of large administrative databases in clinical epidemiology. *Statistical Methods in Medical Research*, 26(3):1350–1372, 2017.
- [91] F. Ieva and A. M. Paganoni. Risk prediction for myocardial infarction via generalized functional regression models. *Statistical Methods in Medical Research*, 25(4):1648–1660, 2016.
- [92] F. Ieva, A. M. Paganoni, D. Pigoli, and V. Vitelli. Multivariate functional clustering for the morphological analysis of electrocardiograph curves. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 62(3):401–418, 2013.
- [93] A. Iorio, F. Rea, G. Barbati, *et al.* HF progression among outpatients with HF in a community setting. *International Journal of Cardiology*, 277:140–146, 2019.
- [94] A. Iorio, M. Senni, G. Barbati, *et al.* Prevalence and prognostic impact of non-cardiac co-morbidities in heart failure outpatients with preserved and reduced ejection fraction: a community-based study. *European Journal of Heart Failure*, 20(9):1257–1266, 2018.
- [95] A. Jahn-Eimermacher. Comparison of the Andersen–Gill model with poisson and negative binomial regression on recurrent event data. *Computational Statistics & Data Analysis*, 52(11):4989–4997, 2008.
- [96] B. Juang and L. Rabiner. Hidden Markov models for speech recognition. *Technometrics*, 33:251–272, 1991.
- [97] J. Kalbfleisch and R. Prentice. *The Statistical Analysis of Failure Time Data, 2nd Edition*. John Wiley & Sons, 360, 2011.
- [98] A. Kalogeropoulos, A. Samman-Tahhan, J. Hedley, *et al.* Progression to Stage D Heart Failure Among Outpatients With Stage C Heart Failure and Reduced Ejection Fraction. *JACC Heart Failure*, 5(7):528–537, 2017.
- [99] E. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *Journal of American Statistical Association*, 53:457–481, 1958.
- [100] M. E. Karim, P. Gustafson, J. Petkau, *et al.* Marginal Structural Cox Models for Estimating the Association Between  $\beta$ -Interferon Exposure and Disease Progression in a Multiple Sclerosis Cohort. *American Journal of Epidemiology*, 180(2):160–171, 2014.
- [101] M. E. Karim, J. Petkau, P. Gustafson, *et al.* On the application of statistical learning approaches to construct inverse probability weights in marginal structural cox models: Hedging against weight-model misspecification. *Communications in Statistics - Simulation and Computation*, 46(10):7668–7697, 2017.

- [102] S. Karve, M. A. Cleves, M. Helm, *et al.* Prospective validation of eight different adherence measures for use with administrative claims data among patients with schizophrenia. *Value in Health*, 12(6):989–995, 2009.
- [103] P. J. Kelly and L. L.-Y. Lim. Survival analysis for recurrent event data: an application to childhood infectious diseases. *Statistics in Medicine*, 19(1):13–33, 2000.
- [104] B. Kennedy. Repeated hospitalizations and self-rated health among the elderly: A multivariate failure time analysis. *American Journal of Epidemiology*, 153:232–241, 02 2001.
- [105] R. H. Keogh, S. R. Seaman, J. M. Gran, and S. Vansteelandt. Simulating longitudinal data from marginal structural models using the additive hazard model. *Biometrical Journal*, 63(7):1526–1541, 2021.
- [106] D. G. Kleinbaum and M. Klein. *Survival Analysis: A Self-Learning Text*. Statistics for Biology and Health. Springer New York, 2013.
- [107] M. Komajda, S. D. Anker, M. R. Cowie, *et al.* Physicians’ adherence to guideline-recommended medications in heart failure with reduced ejection fraction: data from the QUALIFY global survey. *European Journal of Heart Failure*, 18(5):514–522, May 2016.
- [108] M. Komajda, M. R. Cowie, L. Tavazzi, Filippatos, and QUALIFY Investigators. *et al.* Physicians’ guideline adherence is associated with better prognosis in outpatients with heart failure with reduced ejection fraction: the QUALIFY international registry. *European Journal of Heart Failure*, 19(11):1414–1423, 2017.
- [109] D. Kong, J. G. Ibrahim, E. Lee, and H. Zhu. Flcrm: Functional linear cox regression model. *Biometrics*, 74(1):109–117, 2018.
- [110] C. Lancia, J. Anninga, M. R. Sydes, *et al.* A novel method to address the association between received dose intensity and survival outcome: benefits of approaching treatment intensification at a more individualised level in a trial of the European Osteosarcoma Intergroup. *Cancer chemotherapy and pharmacology*, 83(5):951–962, 2019.
- [111] C. Lancia, J. Anninga, M. R. Sydes, *et al.* Method to measure the mismatch between target and achieved received dose intensity of chemotherapy in cancer trials: a retrospective analysis of the MRC BO06 trial in osteosarcoma. *BMJ open*, 9(5), 2019.
- [112] C. Lancia, C. Spitoni, J. Anninga, *et al.* Marginal structural models with dose-delay joint-exposure for assessing variations to chemotherapy intensity. *Statistical Methods in Medical Research*, 28(9):2787–2801, 2019.
- [113] G. Last and A. Brandt. *Marked Point Processes on the Real Line: The Dynamical Approach*. Springer-Verlag New York, 1995.
- [114] J. G. Le-Rademacher, S. Hillman, E. Storricks, *et al.* Adverse event burden score—a versatile summary measure for cancer clinical trials. *Cancers (Basel)*, 12(11):3251, 2020.
- [115] D. Lee, L. Donovan, P. Austin, *et al.* Comparison of coding of heart failure and comorbidities in administrative and clinical data for use in outcomes research. *Medical Care*, 43:182–188, 2005.
- [116] E. Lee, H. Zhu, D. Kong, *et al.* BFLCRM: A Bayesian functional linear Cox regression model for predicting time to conversion to Alzheimer’s disease. *The Annals of Applied Statistics*, 9(4):2153–2178, 2015.
- [117] S. Lee, D. Hershman, P. Martin, *et al.* Toxicity burden score: a novel approach to summarize multiple toxic effects. *Annals of Oncology*, 23(2):537–541, 2012.
- [118] A. Lehmann, P. Aslani, R. Ahmed, *et al.* Assessing medication adherence: options to consider. *International Journal of Clinical Pharmacy*, 36(1):55–69, 2014.

## Bibliography

- [119] I. Lewis, M. Nooij, J. Whelan, *et al.* Improvement in Histologic Response But Not Survival in Osteosarcoma Patients Treated With Intensified Chemotherapy: A Randomized Phase III Trial of the European Osteosarcoma Intergroup. *JNCI: Journal of the National Cancer Institute*, 99(2):112–128, 2007.
- [120] I. J. Lewis, S. Weeden, D. Machin, *et al.* Received dose and dose-intensity of chemotherapy and outcome in nonmetastatic extremity osteosarcoma. *Journal of Clinical Oncology*, 18(24):4028–4037, 2000.
- [121] K. Li and S. Luo. Dynamic prediction of Alzheimer’s disease progression using features of multiple longitudinal outcomes and time-to-event data. *Statistics in Medicine*, 38(24):4804–4818, 2019.
- [122] L. Li, T. Greene, and B. Hu. A simple method to estimate the time-dependent receiver operating characteristic curve and the area under the curve with right censored data. *Statistical Methods in Medical Research*, 27(8):2264–2278, 2018.
- [123] L. Li and C. Wu. *tdROC: Nonparametric Estimation of Time-Dependent ROC Curve from Right Censored Survival Data*, 2016. R package version 1.0, <https://CRAN.R-project.org/package=tdROC>
- [124] Z. Li, V. M. Chinchilli, and M. Wang. A Bayesian joint model of recurrent events and a terminal event. *Biometrical Journal*, 61(1):187–202, 2019.
- [125] Z. Li, V. M. Chinchilli, and M. Wang. A time-varying Bayesian joint hierarchical copula model for analysing recurrent events and a terminal event: an application to the Cardiovascular Health Study. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 69(1):151–166, 2020.
- [126] D. Y. Lin. On fitting Cox’s proportional hazards models to survey data. *Biometrika*, 87:37–47, 2000.
- [127] B. Liu, Y. Huang, Y. Sun, *et al.* Prognostic value of inflammation-based scores in patients with osteosarcoma. *Scientific Reports*, 6(1), 2016.
- [128] X. Liu and M. C. K. Yang. Identifying temporally differentially expressed genes through functional principal components analysis. *Biostatistics*, 10(4):667–679, 2009.
- [129] D. Lloyd-Jones, R. Adams, T. Brown, *et al.* Executive summary: heart disease and stroke statistics–2010 update: a report from the American Heart Association. *Circulation*, 121:948–954, 2010.
- [130] A. P. Maggioni, S. D. Anker, U. Dahlström, *et al.* Are hospitalized or ambulatory patients with heart failure treated in accordance with european society of cardiology guidelines? evidence from 12,440 patients of the ESC heart failure Long-Term registry. *European Journal of Heart Failure*, 15(10):1173–1184, 2013.
- [131] A. P. Maggioni and F. Spandonaro. Lo scompenso cardiaco acuto in italia. *Giornale italiano di cardiologia (Rome)*, 15(2 Suppl 2):3S–4S, 2014.
- [132] L. Mancini and A. M. Paganoni. Marked point process models for the admissions of heart failure patients. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 12(2):125–135, 2019.
- [133] B. C. Martin, E. K. Wiley-Exley, S. Richards, *et al.* Contrasting measures of adherence with simple drug use, medication switching, and therapeutic duplication. *Annals of Pharmacotherapy*, 43(1):36–44, 2009.
- [134] A. Martino, A. Ghiglietti, F. Ieva, and A. M. Paganoni. A k-means procedure based on a Mahalanobis type distance for clustering multivariate functional data. *Statistical Methods & Applications*, 28(2):301–322, 2019.

- [135] J. A. Martín-Fernández, J. Palarea-Albaladejo, and R. A. Olea. Dealing with Zeros. In V. Pawlowsky-Glahn and A. Buccianti, editors, *Compositional Data Analysis*, chapter 4, pages 43–58. John Wiley & Sons, Ltd, 2011.
- [136] C. Mazzali, A. M. Paganoni, F. Ieva, *et al.* Methodological issues on the use of administrative data in healthcare research: The case of heart failure hospitalizations in Lombardy region, 2000 to 2012. *BMC Health Services Research*, 16(234), 2016.
- [137] P. McCullagh and J. Nelder. *Generalized Linear Models. Second edn.* Monographs on Statistics and Applied Probability. Chapman and Hall, 1989.
- [138] J. McMurray, A. Cohen-Solal, R. Dietz, *et al.* Practical recommendations for the use of ace inhibitors, beta-blockers, aldosterone antagonists and angiotensin receptor blockers in heart failure: Putting guidelines into practice. *European Journal of Heart Failure*, 7(5):710–721, 2005.
- [139] J. J. McMurray, S. Adamopoulos, S. D. Anker, *et al.* ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure 2012: The Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure 2012 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association (HFA) of the ESC. *European Heart Journal*, 14:803–869, 2012.
- [140] A. McTiernan, R. C. Jinks, M. R. Sydes, *et al.* Presence of chemotherapy-induced toxicity predicts improved survival in patients with localised extremity osteosarcoma treated with doxorubicin and cisplatin: A report from the European Osteosarcoma Intergroup. *European Journal of Cancer*, 48(5):703–712, 2012.
- [141] MERIT-HF Study Group. Effect of metoprolol CR/XL in chronic heart failure: Metoprolol CR/XL Randomised Intervention Trial in Congestive Heart Failure (MERIT-HF). *The Lancet*, 353:2001–2007, 1999.
- [142] P. A. Meyer. A decomposition theorem for supermartingales. *Illinois J. Math.*, 6:193–205, 1962.
- [143] Ministero della Salute della Repubblica Italiana. *Scompenso Cardiaco*, 2022. [https://www.salute.gov.it/portale/salute/p1\\_5.jsp?area=Malattie\\_cardiovascolari&id=43&lingua=italiano](https://www.salute.gov.it/portale/salute/p1_5.jsp?area=Malattie_cardiovascolari&id=43&lingua=italiano)
- [144] H. Müller. Functional modelling and classification of longitudinal data. *Scandinavian Journal of Statistics*, 32:223–240, 2005.
- [145] W. Nelson. Hazard plotting for incomplete failure data. *Journal of Quality Technology.*, 1(1):27–52, 1969.
- [146] W. Nelson. Theory and applications of hazard plotting for censored failure data. *Technometrics*, 14(4):945–966, 1972.
- [147] L. Osterberg and T. Blaschke. Adherence to medication. *N. Engl. J. Med.*, 353(5):487–497, Aug. 2005.
- [148] W. Ouwerkerk, A. A. Voors, S. D. Anker, *et al.* Determinants and clinical outcome of uptitration of ACE-inhibitors and beta-blockers in patients with heart failure: a prospective european study. *European Heart Journal*, 38(24):1883–1890, 2017.
- [149] A. Ozga, M. Kieser, and G. Rauch. A systematic comparison of recurrent event models for application to composite endpoints. *BMC Medical Research Methodology*, 18(2), 2018.
- [150] P. Pazos-López, J. Peteiro-Vázquez, A. Carcía-Campos, *et al.* The causes, consequences, and treatment of left or right heart failure. *Vascular Health Risk Management*, 7:237–254, 2011.
- [151] M. J. Pencina and R. B. D’Agostino. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Statistics in Medicine*, 23(13):2109–2123, 2004.

## Bibliography

- [152] M. J. Pencina, R. B. D'Agostino, Sr, R. B. D'Agostino, Jr, and R. S. Vasan. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Statistics in Medicine*, 27(2):157–72; discussion 207–12, Jan. 2008.
- [153] A. Peters and T. Hothorn. *ipred: Improved Predictors*, 2019. R package version 0.9-9, <https://CRAN.R-project.org/package=ipred>
- [154] P. Ponikowski, A. A. Voors, S. D. Anker, *et al.* 2016 esc guidelines for the diagnosis and treatment of acute and chronic heart failure: The task force for the diagnosis and treatment of acute and chronic heart failure of the european society of cardiology (esc). developed with the special contribution of the heart failure association (hfa) of the esc. *European Heart Journal*, 37(27):2129–2200, 2016.
- [155] G. Prada-Ramallal, B. Takkouche, and A. Figueiras. Bias in pharmacoepidemiologic studies using secondary health care databases: a scoping review. *BMC Medical Research Methodology*, 19(1):53, 2019.
- [156] R. L. Prentice, B. J. Williams, and A. V. Peterson. On the regression analysis of multivariate failure time data. *Biometrika*, 68(2):373–379, 1981.
- [157] C. Proust-Lima, J. Dartigues, and H. Jacqmin-Gadda. Joint modeling of repeated multivariate cognitive measures and competing risks of dementia and death: a latent process and latent class approach. *Statistics in Medicine*, 35(3):382–398, 2016.
- [158] X. Qin, J. Hung, M. Knuiman, *et al.* Evidence-based pharmacotherapies used in the postdischarge phase are associated with improved one-year survival in senior patients hospitalized with heart failure. *Cardiovascular Therapeutics*, 36(6):e12464, Dec. 2018.
- [159] S. Qu, J. L. Wang, and X. Wang. Optimal estimation for the functional Cox model. *Annals of Statistics*, 44(4):1708–1738, 2016.
- [160] H. Quan, V. Sundararajan, P. Halfon, *et al.* Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Medical Care*, 43(11):1130–1139, 2005.
- [161] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018. <https://www.R-project.org/>
- [162] J. Ramsay and B. W. Silverman. *Functional Data Analysis*. Springer Series in Statistics. Springer, New York, NY, 2005.
- [163] J. O. Ramsay and B. W. Silverman. *Applied Functional Data Analysis: Methods and Case Studies*. Springer Series in Statistics. Springer New York, 2002.
- [164] Regione Lombardia. *HFDData project: Utilization of Regional Health Source Databases for Evaluating Epidemiology, short- and medium-term outcome and process indicators in patients hospitalized for heart failure*. Progetto di Ricerca Finalizzata di Regione Lombardia - HFDData-RF-2009-1483329, 2012.
- [165] H. Ren, L. Sun, H. Li, and Z. Ye. Prognostic Significance of Serum Alkaline Phosphatase Level in Osteosarcoma: A Meta-Analysis of Published Data. *BioMed Research International*, 2015(Article ID 160835), 2015.
- [166] J. Ritter and S. S. Bielack. Osteosarcoma. *Annals of Oncology*, 21(suppl 7):vii320–vii325, 2010.
- [167] D. Rizopoulos. *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R*. Chapman and Hall/CRC, 2012.
- [168] D. Rizopoulos. The R Package JMbayes for fitting joint models for longitudinal and time-to-event data using MCMC. *Journal of Statistical Software*, 72(7), 2016.

- [169] J. Robins. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical Modelling*, 7(9):1393–1512, 1986.
- [170] J. M. Robins, D. Blevins, G. Ritter, and M. Wulfsohn. G-estimation of the effect of prophylaxis therapy for *Pneumocystis carinii* pneumonia on the survival of AIDS patients. *Epidemiology*, 3(4):319–336, 1992.
- [171] J. M. Robins, M. A. Hernán, and B. Brumback. Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11(5):550–560, 2000.
- [172] A. Rogatko, J. S. Babb, H. Wang, *et al.* Patient Characteristics Compete with Dose as Predictors of Acute Treatment Toxicity in Early Phase Clinical Trials. *Clinical Cancer Research*, 10(14):4645–4651, 2004.
- [173] J. K. Rogers, A. Yaroshinsky, S. J. Pocock, *et al.* Analysis of recurrent events with an associated informative dropout time: Application of the joint frailty model. *Statistics in Medicine*, 35(13):2195–2205, 2016.
- [174] V. Rondeau. Statistical models for recurrent events and death: Application to cancer events. *Mathematical and Computer Modelling*, 52(7):949–955, 2010.
- [175] V. Rondeau, D. Commenges, and P. Joly. Maximum Penalized Likelihood Estimation in a Gamma-Frailty Model. *Lifetime Data Analysis*, 9:139–153, 2003.
- [176] V. Rondeau, L. Filleul, and P. Joly. Nested frailty models using maximum penalized likelihood estimation. *Statistics in Medicine*, 25(23):4036–4052, 2006.
- [177] V. Rondeau, S. Mathoulin-Pelissier, H. Jacqmin-Gadda, *et al.* Joint frailty models for recurring events and death using maximum penalized likelihood estimation: application on cancer events. *Biostatistics*, 8(4):708–721, 2007.
- [178] G. Rosen and A. Nirenberg. Neoadjuvant chemotherapy for osteogenic sarcoma: a five year follow-up (t-10) and preliminary report of new studies (t-12). *Progress in clinical and biological research*, 201:39–51, 1985.
- [179] T. M. Ruppap, P. S. Cooper, D. R. Mehr, *et al.* Medication adherence interventions improve heart failure mortality and readmission rates: Systematic review and meta-analysis of controlled trials. *Journal of the American Heart Association*, 5(6), 2016.
- [180] J. Saczynski, S. Andrade, L. Harrold, *et al.* A systematic review of validated methods for identifying heart failure using administrative data. *Pharmacoepidemiology & Drug Safety*, 21:129–140, 2012.
- [181] M. Schemper and R. Henderson. Predictive accuracy and explained variation in cox regression. *Biometrics*, 56(1):249–255, 2000.
- [182] M. Schemper and T. L. Smith. A note on quantifying follow-up in studies of failure time. *Controlled Clinical Trials*, 17(4):343–346, 1996.
- [183] G. Schwarz. Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2):461–464, 1978.
- [184] S. Sivendran, A. Latif, R. B. McBride, *et al.* Adverse Event Reporting in Cancer Clinical Trial Publications. *Journal of Clinical Oncology*, 32(2):83–89, 2014.
- [185] S. Smeland, S. S. Bielack, J. Whelan, *et al.* Survival and prognosis with osteosarcoma: outcomes in more than 2000 patients in the EURAMOS-1 (European and American Osteosarcoma Study) cohort. *European Journal of Cancer*, 109:36–50, 2019.

## Bibliography

- [186] R. L. Souhami, A. W. Craft, J. W. Van der Eijken, *et al.* Randomised trial of two regimens of chemotherapy in operable osteosarcoma: a study of the European Osteosarcoma Intergroup. *The Lancet*, 350(9082):911–917, 1997.
- [187] M. Spreafico, F. Gasperoni, G. Barbati, *et al.* Adherence to Disease-Modifying Therapy in Patients Hospitalized for HF: Findings from a Community-Based Study. *American Journal of Cardiovascular Drugs*, 20:179–190, 2020.
- [188] M. Spreafico and F. Ieva. Dynamic monitoring of the effects of adherence to medication on survival in heart failure patients: A joint modeling approach exploiting time-varying covariates. *Biometrical Journal*, 63(2):305–322, 2021.
- [189] M. Spreafico and F. Ieva. Functional modeling of recurrent events on time-to-event processes. *Biometrical Journal*, 63(5):948–967, 2021.
- [190] M. Spreafico, F. Ieva, F. Arlati, *et al.* Novel longitudinal Multiple Overall Toxicity (MOTox) score to quantify adverse events experienced by patients during chemotherapy treatment: a retrospective analysis of the MRC BO06 trial in osteosarcoma. *BMJ Open*, 11(12):e053456, 2021.
- [191] M. Spreafico, F. Ieva, and M. Fiocco. Longitudinal Latent Overall Toxicity (LOTox) profiles in osteosarcoma: a new taxonomy based on latent Markov models. *arXiv*, 2107.12863, 2021. <https://arxiv.org/abs/2107.12863>
- [192] M. Spreafico, F. Ieva, and M. Fiocco. Modelling time-varying covariates effect on survival via Functional Data Analysis: application to the MRC BO06 trial in osteosarcoma. *Statistical Methods & Applications*, 2022. <https://doi.org/10.1007/s10260-022-00647-0>
- [193] J. F. Steiner. Measuring adherence with medications: time is of the essence. *Pharmacoepidemiology & Drug Safety*, 25(3):333–335, 2016.
- [194] B. L. Strom, editor. *Pharmacoepidemiology*. John Wiley & Sons, Ltd, Chichester, UK, 2003.
- [195] B. L. Strom and S. E. Kimmel, editors. *Textbook of Pharmacoepidemiology*. John Wiley & Sons, Chichester, England, 2006.
- [196] C. L. Su and F. C. Lin. Analysis of cyclic recurrent event data with multiple event types. *Japanese Journal of Statistics and Data Science*, 4(2):895–915, 2021.
- [197] L. Sun, L. Zhu, and J. Sun. Regression analysis of multivariate recurrent event data with time-varying covariate effects. *Journal of Multivariate Analysis*, 100(10):2214 – 2223, 2009.
- [198] G. Thanarajasingam, P. J. Atherton, P. J. Novotny, *et al.* Longitudinal adverse event assessment in oncology clinical trials: the Toxicity over Time (ToxT) analysis of Alliance trials NCCTG N9741 and 979254. *The Lancet Oncology*, 17(5):663–670, 2016.
- [199] G. Thanarajasingam, J. M. Hubbard, J. A. Sloan, and A. Grothey. The Imperative for a New Approach to Toxicity Analysis in Oncology Clinical Trials. *JNCI: Journal of the National Cancer Institute*, 107(10), 08 2015. djv216.
- [200] G. Thanarajasingam, J. P. Leonard, T. E. Witzig, *et al.* Longitudinal Toxicity over Time (ToxT) analysis to evaluate tolerability: a case study of lenalidomide in the CALGB 50401 (Alliance) trial. *The Lancet Haematology*, 7(6):e490–e497, 2020.
- [201] T. M. Therneau. *A Package for Survival Analysis in R*, 2020. R package version 3.1-11, <https://CRAN.R-project.org/package=survival>
- [202] T. M. Therneau and P. Grambsch. *Modeling Survival Data: Extending the Cox Model*. Springer, New York, 2000.

- [203] T. M. Therneau, P. M. Grambsch, and T. R. Fleming. Martingale-based residuals for survival models. *Biometrika*, 77(1):147–160, 03 1990.
- [204] A. Trotti, A. Colevas, A. Setser, *et al.* CTCAE v3.0: development of a comprehensive grading system for the adverse effects of cancer treatment. *Seminars in Radiation Oncology*, 13(3):176–181, 2003.
- [205] A. Trotti, T. F. Pajak, C. K. Gwede, *et al.* TAME: development of a new method for summarising adverse events of cancer treatment by the Radiation Therapy Oncology Group. *The Lancet Oncology*, 8(7):613–624, 2007.
- [206] A. A. Tsiatis and M. Davidian. Joint modelling of longitudinal and time-to-event data: an overview. *Statistica Sinica*, 14(3):809–834, 2004.
- [207] S. Ullah and C. F. Finch. Applications of functional data analysis: A systematic review. *BMC Medical Research Methodology*, 13(43), 2013.
- [208] U.S. Department of Health and Human Services. *Common Terminology Criteria for Adverse Events v3.0 (CTCAE)*, 2006. [https://www.eortc.be/services/doc/ctc/ctcae\\_v3.pdf](https://www.eortc.be/services/doc/ctc/ctcae_v3.pdf).
- [209] S. van Buuren and K. Groothuis-Oudshoorn. mice: Multivariate imputation by chained equations inr. *J. Stat. Softw.*, 45(3), 2011.
- [210] W. M. van der Wal and R. B. Geskus. ipw: An R package for inverse probability weighting. *Journal of Statistical Software*, 43(13):1–23, 2011.
- [211] A. Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, 13(2):260–269, 1967.
- [212] B. Vrijens, S. De Geest, D. A. Hughes, *et al.* A new taxonomy for describing and defining adherence to medications. *British Journal of Clinical Pharmacology*, 73(5):691–705, 2012.
- [213] L. J. Wei, D. Y. Lin, and L. Weissfeld. Regression Analysis of Multivariate Incomplete Failure Time Data by Modeling Marginal Distributions. *Journal of the American Statistical Association*, 84(408):1065–1073, 1989.
- [214] WHO Collaborating Centre for Drug Statistics Methodology. <https://www.whocc.no>
- [215] WHO, International League Against Epilepsy and International Bureau for Epilepsy. *Atlas: Epilepsy Care in the World 2005*. Geneva: World Health Organization, 2005.
- [216] L. Wiggins. *Panel analysis: latent probability models for attitude and behaviour processes*. Elsevier, Amsterdam, 1973.
- [217] T. Williamson and P. Ravani. Marginal structural models in clinical research: when and how to use them? *Nephrology Dialysis Transplantation*, 32(suppl 2):ii84–ii90, 2017.
- [218] M. A. Williamson, L. M. Snyder, and J. B. Wallach. *Wallach’s Interpretation of Diagnostic Tests: Pathways to Arriving at a Clinical Diagnosis. Tenth Edition*. Philadelphia: Wolters Kluwer Health, 2015.
- [219] World Health Organization. *Adherence to Long-Term Therapies: Evidence for Action*. WHO Library Cataloguing-in-Publication Data, 2003.
- [220] World Health Organization, WHO International Working Group for Drug Statistics Methodology, WHO Collaborating Centre for Drug Statistics Methodology, and WHO Collaborating Centre for Drug Utilization Research and Clinical Pharmacological Service. *Introduction to Drug Utilization Research*. WHO Library Cataloguing-in-Publication Data, 2003.



## Bibliography

- [221] WRITING COMMITTEE MEMBERS, C. W. Yancy, M. Jessup, *et al.* 2016 ACC/AHA/HFSA focused update on new pharmacological therapy for heart failure: An update of the 2013 ACCF/AHA guideline for the management of heart failure: A report of the american college of cardiology/american heart association task force on clinical practice guidelines and the heart failure society of america. *Circulation*, 134(13):e282–93, 2016.
- [222] C. W. Yancy, M. Jessup, B. Bozkurt, *et al.* 2013 ACCF/AHA guideline for the management of heart failure: executive summary: a report of the american college of cardiology Foundation/American heart association task force on practice guidelines. *Circulation*, 128(16):1810–1852, Oct. 2013.
- [223] F. Yao, H. Müller, and J. Wang. Functional Data Analysis for Sparse Longitudinal Data. *Journal of the American Statistical Association*, 100(470):577–590, 2005.
- [224] P. Ye, X. Zhao, L. Sun, and W. Xu. A semiparametric additive rates model for multivariate recurrent events with missing event categories. *Computational Statistics & Data Analysis*, 89:39 – 50, 2015.
- [225] B. S. Yoo, J. Oh, B. K. Hong, *et al.* SURvey of Guideline Adherence for Treatment of Systolic Heart Failure in Real World (SUGAR): a multi-center, retrospective, observational study. *PLoS One*, 9(1):e86596, Jan. 2014.
- [226] S. Zhang, Q. Chen, and Q. Wang. The use of and adherence to CTCAE v3.0 in cancer clinical trial publications. *Oncotarget*, 7(40):65577–65588, 2016.
- [227] L. Zhao and B. Zhang. Doxorubicin induces cardiotoxicity through upregulation of death receptors mediated apoptosis in cardiomyocytes. *Scientific Reports*, 7:44735, 2017.
- [228] S. Zhou, A. Starkov, M. K. Froberg, *et al.* Cumulative and irreversible cardiac mitochondrial dysfunction induced by doxorubicin. *Cancer research*, 61(2):771–777, 2001.

# 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

## *Summary*

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.

# Samenvatting

*Tijdsafhankelijke covariaten* zijn van groot belang in klinisch onderzoek. Ze geven namelijk zowel inzicht in de evolutie van dynamische patronen, als in de invloed daarvan op de gezondheidsstatus en ziekteprogressie van een patiënt. De literatuur beschouwt zulke gegevens meestal als stuksgewijs-constant of met vaste baseline. Daarbij gaat een potentieel aan informatie verloren, wat wél benut kan worden wanneer goed wordt gekeken naar de associatie tussen de tijdsafhankelijke covariaat enerzijds, en de *tijd tot een bepaalde* gebeurtenis anderzijds. De genoemde fenomenen zijn complex, wat vraagt om een geavanceerd analytisch kader waarbinnen men zowel de evolutie van de ziekte als de relatie met de dynamische aard van tijdsafhankelijke kenmerken adequaat kan modelleren. Een dergelijk analytisch kader wordt nog zelden overwogen, terwijl het nieuwe inzichten kan verschaffen voor medisch onderzoek. Dit vormt een uitdaging van zowel klinische als statistische relevantie.

Dit proefschrift richt zich op het ontwikkelen van wiskundige en statistische methoden om tijdsafhankelijke processen adequaat weer te geven en te modelleren binnen de context van de analyse van de tijd tot een gebeurtenis. Hiervoor wordt gebruik gemaakt van toegespitste “Cox”modellen. Het voornaamste doel is het verrijken van de huidige kennis op dit gebied, door kenmerken toe te voegen die relevant zijn voor de betreffende tijdsafhankelijke covariaten. De inspanningen van dit werk zijn tweeledig: ze richtten zich op de complexiteit van (i) de ontwikkeling van adequate dynamische karakterisering van de bestudeerde processen (d.w.z. het *representatieprobleem*) en (ii) de identificatie en kwantificering van het verband tussen tijdsafhankelijke processen en de overleving van patiënten (d.w.z. het *tijd-tot-gebeurtenis modelleringsprobleem*). In beide gevallen is het voornaamste probleem het hanteren van complexe gegevensbronnen, waarbij men twee aspecten in ogenschouw moet houden: ten eerste de aard van de processen en hun bijzondere kenmerken – denk aan temporele dynamiek, categorische niveaus of terugkerende gebeurtenissen – en ten tweede de complexe balans tussen klinische interpreteerbaarheid en wiskundige formulering.

Afhankelijk van de context van het onderzoek worden verschillende benaderingen voorgesteld. Wat betreft *representaties* wordt ten eerste gebruikgemaakt van complexe data-integratie en “functionele” data-analyse. Nieuwe longitudinale of functionele kenmerken maken het mogelijk om trends en variaties in de evolutie van processen in de tijd op te nemen. Ten tweede worden nieuwe statistische methoden geïntroduceerd waarmee men bijkomende informatie, om precies te zijn longitudinale dan wel functionele gegevens, kan ontleen aan verschillende bronnen. Om antwoord te geven op onderzoeksvragen

## *Samenvatting*

worden onder andere stochastische procestheorie, latente Markovmodellen en compositionele data-analyse gebruikt. Wat betreft het *modellieren* van de tijd tot de gebeurtenis worden geavanceerde versies van regressiemodellen van het Cox-type voorgesteld. Op die manier is het mogelijk om dynamische covariaten in de overlevingsanalyse op te nemen. Afhankelijk van het type studie en gegevens wordt gekozen voor gecombineerde of functionele overlevingsmodellen, die technieken gebruiken om de dimensie te verminderen, of marginale structurele modellen.

Met het oplossen van bovengenoemde statistische uitdagingen heeft dit werk niet alleen impact voor de wiskundige en statistische onderzoeksgemeenschap. Het tracht daarnaast artsen en andere klinici te ondersteunen door instrumenten aan te bieden die nuttig zijn voor hun beroepspraktijk. Alle onderzoeksonderwerpen zijn ingegeven door specifieke klinische vragen, met als doel om meer inzicht te krijgen in gepersonaliseerde behandelingen voor cardiologische en oncologische patiënten. Deel I van dit proefschrift is gericht op de studie van de processen van medicijngebruik, therapietrouw aan medicatie en heropnames bij patiënten met hartfalen. Deel II betreft het onderzoek van tijdsafhankelijke aspecten van chemotherapiebehandeling bij patiënten met een osteosarcoom. Denk hierbij aan dosiswijzigingen, biomarkers en accumulatie van toxiciteiten.

De nieuw ontwikkelde, op maat gemaakte methoden verbeteren dus de modellering van de tijd tot een gebeurtenis waarbij tijdsafhankelijke kenmerken worden meegenomen. Dit markeert een belangrijke stap voorwaarts in de definitie van nieuwe, aangepaste én flexibele monitoringinstrumenten. De in dit werk ontwikkelde instrumenten vinden hun toepassing in onderzoek naar diverse ziektebeelden, waarbij men beschikt over complexe gegevensbronnen.

# Sommario

Nella ricerca clinica, le *covariate tempo-dipendenti* sono di grande interesse in quanto racchiudono informazioni sull'evoluzione di elementi dinamici che possono avere un impatto sullo stato di salute del paziente e sulla progressione della malattia. In letteratura, queste caratteristiche vengono solitamente analizzate utilizzando un approccio di “baseline” o “costante a tratti”, perdendo così il potenziale contenuto informativo che potrebbero fornire se l'associazione tra i dati tempo-dipendenti e quelli di *tempo all'evento* venisse catturata in maniera appropriata. Data la complessità dei fenomeni, è necessario formulare adeguatamente i modelli per l'evoluzione della malattia e caratterizzare in un contesto analitico avanzato la relazione tra di essa e la natura dinamica delle caratteristiche tempo-varianti. Questi aspetti sono raramente affrontati e possono fornire nuovi spunti per la ricerca medica, rappresentando una sfida sia di rilevanza clinica che di interesse statistico.

Questa tesi si concentra sullo sviluppo di metodi matematici e statistici per rappresentare appropriatamente i processi tempo-dipendenti ed includerli in un contesto di analisi di tempo all'evento, mediante modelli di sopravvivenza di tipo “Cox”. Lo scopo principale è quello di arricchire le informazioni disponibili per studiare la sopravvivenza con caratteristiche rilevanti legate alle covariate tempo-varianti di interesse. Questo lavoro affronta quindi le difficoltà sia di (i) sviluppare adeguate caratterizzazioni dinamiche dei processi oggetto di studio (ovvero il problema di *rappresentazione*), sia di (ii) identificare e quantificare l'associazione tra i processi tempo-dipendenti e la sopravvivenza dei pazienti (ovvero il problema della *modellazione* del tempo all'evento). In entrambi i casi, la sfida principale consiste nel trattare fonti di dati complesse, sia tenendo conto della natura dei processi e dei loro aspetti peculiari (come le dinamiche temporali, i livelli categorici o gli eventi ricorrenti), sia bilanciando in modo adeguato l'interpretabilità clinica e la formulazione matematica.

Lungo il corso della tesi vengono proposti approcci diversificati a seconda del contesto dello studio. In termini di *rappresentazione*, grazie all'integrazione di fonti di dati complesse e all'utilizzo di tecniche di analisi dei dati funzionali, vengono elaborate nuove covariate tempo-dipendenti (longitudinali o funzionali) in grado di riflettere l'evoluzione dei processi nel tempo. Inoltre, vengono introdotte nuove metodologie statistiche per estrarre dai dati informazioni aggiuntive, impiegando, tra gli altri, la teoria dei processi stocastici, i modelli di Markov latenti e l'analisi dei dati composizionali. In termini di *modellazione* del tempo all'evento, i modelli di regressione di tipo “Cox” vengono estesi al fine di includere le covariate dinamiche nell'analisi di sopravvivenza. Le tecniche di riduzione della

dimensionalità ed i modelli proposti (“joint” o “marginal structural”) sono implementati a secondo del tipo di studio e di dati a disposizione.

Oltre ad avere un impatto sulla comunità dei ricercatori in matematica e statistica, questo lavoro mira a fornire strumenti utili per supportare i medici nella pratica clinica. Tutti i temi di ricerca che vengono affrontati sono motivati da domande cliniche specifiche riguardanti i trattamenti personalizzati in pazienti cardiologici ed oncologici. La parte I di questa tesi si concentra sullo studio dei processi di re-ospedalizzazione, consumo ed aderenza ai farmaci in pazienti con scompenso cardiaco. La parte II tratta lo studio degli aspetti tempo-varianti della chemioterapia, come le modifiche del dosaggio, i biomarcatori o l’accumulo di tossicità, in pazienti affetti da osteosarcoma.

In conclusione, lo sviluppo di nuove metodologie per dati di tempo all’evento che includano informazioni dinamiche può rappresentare un significativo passo avanti nella definizione di nuovi strumenti di monitoraggio personalizzati, che potrebbero poi essere applicati allo studio di diverse patologie caratterizzate da fonti di dati complesse.

# List of Publications

## Peer reviewed journal articles

- **M. Spreafico**, F. Gasperoni, G. Barbati *et al.* Adherence to Disease-Modifying Therapy in Patients Hospitalized for HF: Findings from a Community-Based Study. *American Journal of Cardiovascular Drugs*, 20:179–190, 2020.
- **M. Spreafico** and F. Ieva. Dynamic monitoring of the effects of adherence to medication on survival in heart failure patients: A joint modeling approach exploiting time-varying covariates. *Biometrical Journal*, 63(2):305–322, 2021.
- **M. Spreafico** and F. Ieva. Functional modeling of recurrent events on time-to-event processes. *Biometrical Journal*, 63(5):948–967, 2021.
- **M. Spreafico**, F. Ieva, F. Arlati *et al.* Novel longitudinal Multiple Overall Toxicity (MOTox) score to quantify adverse events experienced by patients during chemotherapy treatment: a retrospective analysis of the MRC BO06 trial in osteosarcoma. *BMJ Open*, 11(12):e053456, 2021.
- **M. Spreafico**, F. Ieva and M. Fiocco. Modelling time-varying covariates effect on survival via Functional Data Analysis: application to the MRC BO06 trial in osteosarcoma. *Statistical Methods & Applications*, 2022. <https://doi.org/10.1007/s10260-022-00647-0>
- **M. Spreafico**, F. Ieva and M. Fiocco. Longitudinal Latent Overall Toxicity (LO-Tox) profiles in osteosarcoma: a new taxonomy based on latent Markov models. arXiv:2107.12863, 2021. <https://arxiv.org/abs/2107.12863> [Submitted]
- **M. Spreafico**, C. Spitoni, C. Lancia *et al.* Causal effects of chemotherapy regimen intensity on survival outcome in osteosarcoma patients through Marginal Structural Cox Models, 2022. [Submitted]

## Book chapters

- F. Ieva, **M. Spreafico** and D. Burba. *Modeling the Effect of Recurrent Events on Time-to-event Processes by Means of Functional Data*. In: G. Aneiros, I. Horová, M. Hušková and P. Vieu (eds) *Functional and High-Dimensional Statistics and Related Fields*. IWFOS 2020. Contributions to Statistics. Springer, Cham, 2020.



## Conference proceedings

- **M. Spreafico**, F. Gasperoni, G. Barbati *et al.* *Target dosages and adherence to PolyPharmacy therapy: a case study based on a regional real-world HF community*. Atti del Congresso congiunto SISMEC & Associazione Alessandro Liberati – Network italiano Cochrane 2018 “Linee guida e percorsi diagnostico-terapeutici assistenziali (PDTA): metodi, aderenza e responsabilità”, May 2019, pp 29–34. ISBN: 9788894345612.
- **M. Spreafico** and F. Ieva. *Joint Models: a smart way to include functional data in healthcare analytics*. In: G. Arbia, S. Peluso, A. Pini and G. Rivellini (eds) *Smart Statistics for Smart Applications – Book of Short Paper SIS2019*, Pearson Italia, pp 1089–1094, June 2019. ISBN: 9788891915108.
- **M. Spreafico** and F. Ieva. *Investigating the role of Proteinuria in Renal Disease: a real-world clinical case study*. In: 40<sup>th</sup> Annual Conference of the International Society for Clinical Biostatistics – Book of Abstracts, pp 553–554, July 2019. ISBN: 9789461652874.
- **M. Spreafico** and F. Ieva. *Impact of time-dependent medication adherence on Heart Failure patients using a Joint Modelling framework*. In: Atti del X Congresso Nazionale SISMEC 2019 “Nuovi disegni nella ricerca clinica: la sfida della complessità tra etica e salute”, pp 193–197, March 2020. ISBN: 9788894345629.
- **M. Spreafico**, F. Ieva and M. Fiocco. *A functional approach to study the relationship between dynamic covariates and survival outcomes: an application to a randomized clinical trial on osteosarcoma*. In: A. Pollice, N. Salvati and F. Schirripa Spagnolo (eds) *Book of Short Paper – SIS 2020*, Pearson Italia, pp 727–732, 2020. ISBN: 9788891910776.
- **M. Spreafico**, F. Ieva and M. Fiocco. *Modelling longitudinal latent toxicity profiles evolution in osteosarcoma patients*. In: C. Perna, N. Salvati and F. Schirripa Spagnolo (eds) *Book of Short Papers – SIS 2021*, Pearson Italia, pp 566–571, 2021. ISBN: 9788891927361.
- **M. Spreafico** and F. Ieva. *Functional modelling of recurrent drug purchases and re-hospitalizations on survival in Heart Failure patients: a real-world case study*. In: Atti del XI Congresso Nazionale SISMEC 2021 “Dati, modelli, decisioni: metodi a servizio dell’organizzazione sanitaria nell’emergenza epidemiologica”, pp 112–114, 2021. ISBN: 9791280503244.
- M. A. Cassia, F. Ieva, **M. Spreafico** *et al.* *Predire l’outcome nella nefrite lupica: il ruolo dell’andamento della proteinuria o della singola misurazione attraverso gli alberi di regressione*. *Giornale Italiano di Nefrologia*, 38(S78):198–199, 2021.
- C. Gregorio, **M. Spreafico** and F. Ieva. *Optimal timing of bone-marrow transplant in myelodysplastic syndromes through multi-state modeling and microsimulation*. In: A. Balzanella, M. Bini, C. Cavicchia and R. Verde (eds) *Book of Short Papers – SIS 2022*, Pearson Italia, pp 1436–1441, 2022.

# Acknowledgments

The development of this thesis has been a long, stimulating and growing process and several people should be acknowledged.

I would like to thank my supervisors, Marta Fiocco and Francesca Ieva, for their support and scientific guidance during these years. Both always made sure that my work was the best it could be. They also helped me to deal with my shyness in public speaking by organising many seminars and presentations. Thanks to Francesca for introducing me to the beautiful world of biostatistics and for all the projects she has involved me in. Thanks to Marta for making my writing “less Italian”, for giving me new perspectives and for her thoughtfulness in supporting me during my settlement in Leiden. This experience as cotutelle PhD gave me the opportunity to learn and grow a lot, both scientifically and personally. For this reason, I would also like to acknowledge Peter Steenhagen for the idea of this double program and for his help in all procedures.

I would like to thank the MOXStat group in Milan, the statistics group in Leiden, and all the people who shared this journey with me. Thanks to Anna Paganoni for her enthusiasm into my work and for always being supportive. Thanks to my colleagues at Politecnico, Mario, Agostino, Edoardo, Nicola, Matteo, Riccardo, Antonino, Chiara, Eleonora, Andrea, Ludovica, Caterina and Laura, for making the research work more enjoyable. I am grateful for all the fun breaks, our online meetings during lockdowns, our dinners and pizzas. Many thanks to my colleagues and friends Lara and Michela, for all these years of mutual support and daily sharing, in good and bad times, and for being my gossip girls. Thanks to my office mates at the Mathematical Institute, Daniel, Tyron, Vera and Totò, for being kind and friendly to me from the start, creating a lovely working and living atmosphere. Many thanks to Vera for translating the Samenvatting of this thesis. Thanks to all PhD and Algant colleagues who made my time in Leiden carefree and fun.

I would like to thank my family for always believing in me, with love. They encouraged me to find my own way, even when it took me abroad. Many thanks to my mum who raised me to be independent, giving me the freedom to be myself and follow my dreams. Thanks to my brothers for enlightening my life. Thanks also to my second family in Ome for welcoming me wholeheartedly and making me feel at home. Thanks to my friends, Rachele and Anna, for always being by my side even though we are far apart.

I would like to thank my paranymphs, Carlotta and Riccardo, for their constant presence and never-ending support. Thanks to my aunt, Carlotta, for her enthusiasm in listening to my projects and for her great emotion at each of my successes. Making her proud fills

## *Acknowledgments*

my heart with joy. Thanks to my partner, Riccardo, for this further journey together and for all our scientific discussions that have always inspired me. But most of all I thank him for making my life happy and extraordinary. Having them both by my side is a privilege.

One last special thanks goes to my grandfather, Ermanno<sup>†</sup>. He has always been my number one fan and would have been proud to leaf through this thesis. I dedicate this work to him.

# Curriculum Vitae

Marta Spreafico was born on January 22<sup>nd</sup> 1993 in Bellano, Italy. She grew up in Forlì, where she got her diploma at the Liceo Scientifico Fulcieri Paulucci de Calboli in 2012.

She then started her bachelor in Mathematical Engineering at Politecnico di Milano (Italy), where she got her BSc degree in 2015 and continued her master studies with a major track in applied statistics. In 2016 she was Board member of the *AIM - Associazione Ingegneri Matematici* (Association of Mathematical Engineers). During her Master thesis period she also did an internship at *Indigo AI Srl*, where she worked on machine learning methods applied to natural language processing for conversational agents and chatbots. She completed both the internship and the MSc degree in Mathematical Engineering in April 2018.

Her master thesis project, followed by few months as graduate research fellow, marked the beginning of her research on biostatistics and statistical modelling for medical applications. She started her PhD at Politecnico di Milano in November 2018 and joined a co-tutelle program with Leiden University (The Netherlands) in July 2019, under the supervision of Dr. Francesca Ieva and Prof.dr. Marta Fiocco.

At Politecnico di Milano she has been teaching assistant in various courses of *Statistics* and *Biostatistics* and tutor for several projects in the course of *Applied Statistics*. She also co-supervised a MSc Thesis in Mathematical Engineering and was PhD Representative at the Department of Mathematics. In the last year of her PhD research she was lecturer for the *Analytics for Pharmacoepidemiology* course (Master in Business Analytics and Big Data) at POLIMI Graduate School of Management (Italy).

During the PhD research time, she presented her research at conferences and seminars in Italy, the Netherlands, Belgium, Poland (online), France (online), Latvia and the United Kingdom. She has also been involved in several collaborations with clinicians from the Cardiovascular Centre in Trieste, the Department of Health Sciences at the University of Milan, the Leiden University Medical Center, the Princess Máxima Center for Pediatric Oncology, the Humanitas Research hospital and Humanitas University.

At the time of writing she is lecturer for the *Mathematics for Statisticians* course (MSc in Statistics and Data Science) at Leiden University and she plans to continue her academic career.

