Statistical modelling of time-varying covariates for survival data

Spreafico, M.

Chapter 2

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 overcame 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 irregularity 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].

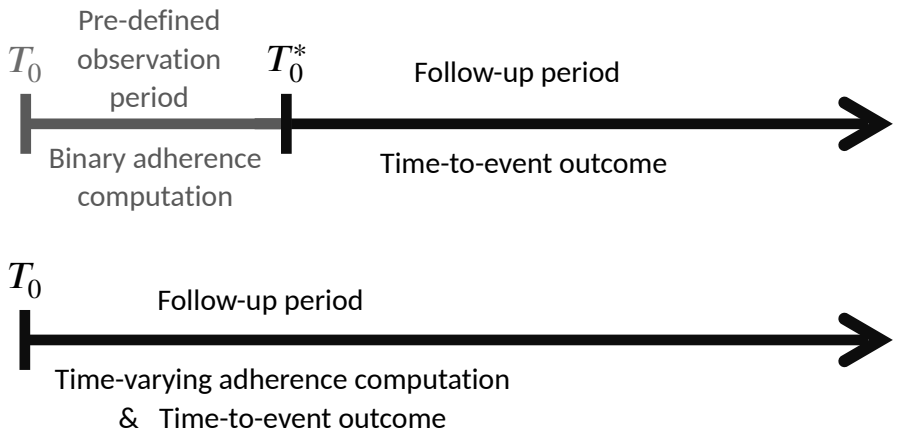

Figure 2.1. *Classical framework*: binary adherence is computed on a pre-defined observation period and time-to-event outcome refers to the follow-up period. *Time-varying framework*: time-varying adherence is computed jointly with the time-to-event outcome. T_0 is the origin event, T_0^* is the end 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 nor 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, $\forall i = 1, ..., n$,

(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, $\forall i = 1, ..., n$

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

where τ 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 $\boldsymbol{y}^{(C)}(t) = \left\{ y_i^{(C)}(t), i = 1, ..., n \right\}$ 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 $\boldsymbol{y}^{(D)}(t) = \left\{ y_i^{(D)}(t), i = 1, ..., n \right\}$ was a dichotomization of $\boldsymbol{y}^{(C)}(t)$ to

identify as adherent those patients with a proportion of therapy consumption up to time t greater or equal to a pre-defined threshold τ, 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 $\boldsymbol{y}^{(C)}(t)$ and $\boldsymbol{y}^{(D)}(t)$ 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 $\boldsymbol{y}^{(D)}(t)$ 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 τ, as usually done in the literature. This allowed us to distinguish patients with *good* $\left(\boldsymbol{y}^{(D)}(t) = 1\right)$ and *poor* $\left(\boldsymbol{y}^{(D)}(t) = 0\right)$ adherence continuity rates during time. On the contrary, the continuous covariates $\boldsymbol{y}^{(C)}(t)$ 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 $\boldsymbol{y}^{(C)}(t)$ 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 $\boldsymbol{y}^{(C)}(t)$ was more descriptive and informative than the dichotomous $\boldsymbol{y}^{(D)}(t)$, we expected that the approach with $\boldsymbol{y}^{(C)}(t)$ 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 $\mathcal{D}_n = \{T_i, D_i, \boldsymbol{y}_i; i = 1, ..., n\}$ denote a sample from the target population, where $\boldsymbol{y}_i$ denote the $n_i \times 1$ longitudinal response vector for the i-th subject $\left(\boldsymbol{y}_i^{(C)} \text{ or } \boldsymbol{y}_i^{(D)} \text{ in our analyses}\right)$, with element y_{il} denoting the value of the longitudinal process taken at time point t_{il}, $l = 1, ..., n_i$. The general form of joint models we used for our analysis is the following:

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

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

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, $\boldsymbol{b}_i$ is the vector of subject-specific random effects, $\boldsymbol{x}_i(t)$ and $\boldsymbol{z}_i(t)$ 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 $\boldsymbol{y}^{(C)}(t) = \left\{ y_i^{(C)}(t), i = 1, ..., n \right\}$, 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:

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

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, $\tilde{\boldsymbol{x}}_i$ is the vector of baseline covariates with fixed effects with regression parameters $\tilde{\boldsymbol{\beta}}$, $\varepsilon_i(t) \sim \mathcal{N}(0, \sigma_\varepsilon^2 \boldsymbol{I}_{n_i})$ is the unknown vector of random errors and $\boldsymbol{b}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{D})$ is the vector of the patient-specific random effects, with $\boldsymbol{D}$ unstructured variance-covariance matrix. Therefore the time-dependent vectors for the fixed and random effects for the i-th patient in (2.1) were $\boldsymbol{x}_i(t) = \left[1, B_n(t, \lambda_1), \ldots, B_n(t, \lambda_4), \tilde{\boldsymbol{x}}_i^T\right]^T$ and $\boldsymbol{z}_i(t) = [1, B_n(t, \lambda_1), \ldots, B_n(t, \lambda_4)]^T$, with relative vectors of regression coefficients $\boldsymbol{\beta} = \left[\beta_0, \ldots, \beta_4, \tilde{\boldsymbol{\beta}}^T\right]^T$ and $\boldsymbol{b}_i = [b_{i0}, \ldots, b_{i4}]^T$.

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

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

where i is the patient's index, $\tilde{\boldsymbol{x}}_i$ is the vector of baseline covariates with fixed effects with regression parameters $\tilde{\boldsymbol{\beta}}$ and $\boldsymbol{b}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{D})$ is the vector of the patient-specific random effects, with $\boldsymbol{D}$ 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 $\boldsymbol{x}_i(t) = \left[1, t, \tilde{\boldsymbol{x}}_i^T\right]^T$ and $\boldsymbol{z}_i(t) = [1, t]^T$, with relative vectors of regression coefficients $\boldsymbol{\beta} = \left[\beta_0, \beta_1, \tilde{\boldsymbol{\beta}}^T\right]^T$ and $\boldsymbol{b}_i = [b_{i0}, b_{i1}]^T$.

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 $\boldsymbol{y}^{(C)}(t)$, giving us information about the patient's current path. Therefore, we focused on three different forms of $f(\mathcal{H}_i(t), \boldsymbol{b}_i, \boldsymbol{\alpha})$ 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:

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

(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:

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

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

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

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.

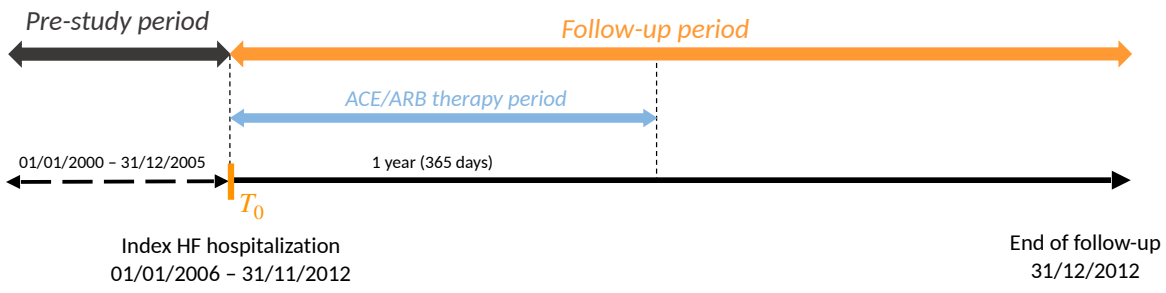

Figure 2.2. Study design. HF = Heart Failure, ACE = Angiotensin-Converting Enzyme inhibitors, ARB = Angiotensin II Receptor Blockers.

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

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

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

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

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 ($t = 1, ..., 12$ 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 $\texttt{cum\_months}_i(t)$ (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 $\texttt{cum\_months}_i(t)$, which indicates the cumulative months covered by ACE/ARB assumption up to time t.

Finally, variable $\texttt{adherence}_i(t)$ (2.9) was a dichotomization of variable $\texttt{cum\_months}_i(t)$ 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, $\texttt{adherence}_i(t)$ was equal to 1 if $\texttt{cum\_months}_i(t)/t \geq 0.8$ 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 $\texttt{cum\_months}_i(t)$ (top panel) and $\texttt{adherence}_i(t)$ (bottom panel) for a random patient.

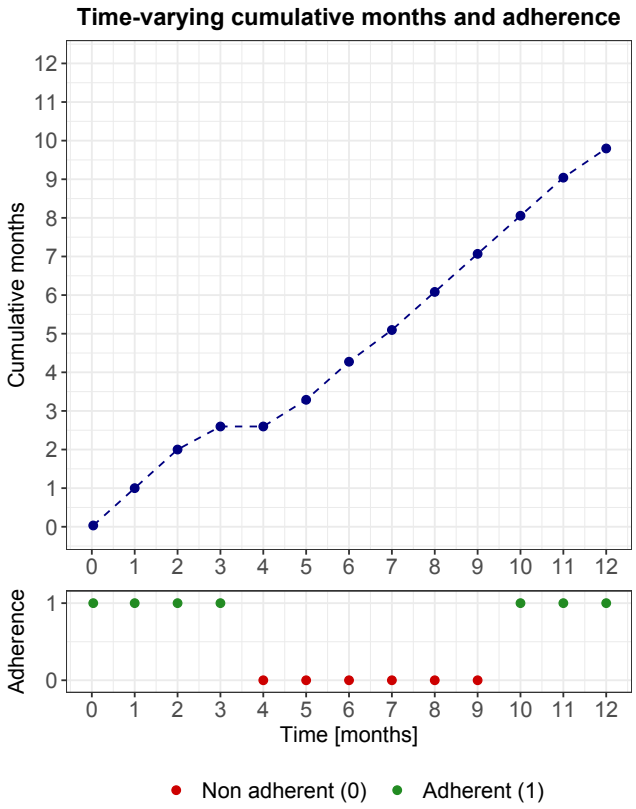

Figure 2.3. Example of time-varying consumption and adherence to ACE/ARB therapy, $\texttt{cum\_months}_i(t)$ and $\texttt{adherence}_i(t)$ respectively.

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$^{\text{st}}$, 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 ≥ 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 $\texttt{cum\_months}_i(12)$, equal to 8.77 months ($s.d.$ = 3.04) and only 2,039 patients (54.2%) with $\texttt{adherence}_i(12) = 1$.

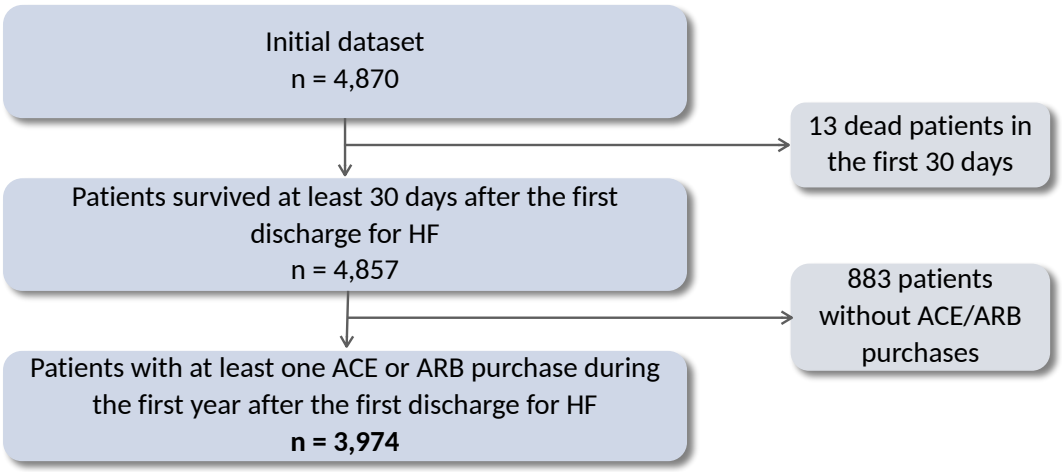

Figure 2.4. Flowchart of cohort selection.

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 $\texttt{cum\_months}_i(t)$, we considered three joint models, namely M_1, M_2 and M_3, with the same longitudinal subprocess (2.3) given by

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

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

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

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

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

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

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

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 $\tilde{\boldsymbol{x}}_i = \texttt{n\_com}_i$.

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 $\boldsymbol{\omega}_i = (\texttt{age}_i,\ \texttt{gender}_i,\ \texttt{n\_com}_i)$.

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 $\boldsymbol{y}^{(C)}(t) = \texttt{cum\_months}(t)$ 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 $\texttt{cum\_months}(t)$ was the following:

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

From parameter estimation of model M_2 we observed that all the 95% credibility intervals did not contain 0, except for the parameters β_0 and α_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.

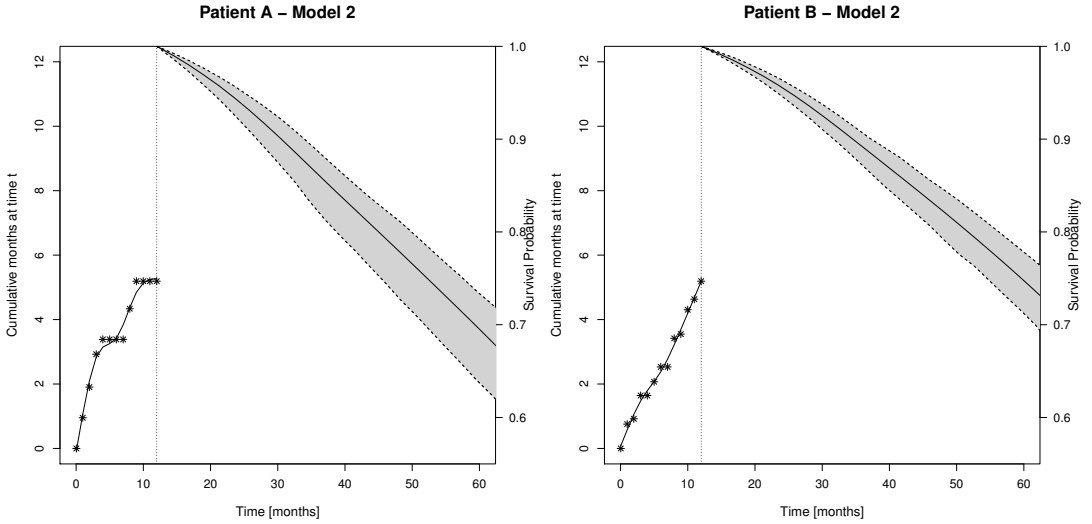

Figure 2.5. Survival probability plots for two male patients aged 72 years and with two comorbidities at the index date using joint model M_2. The values of continuous time-varying covariate $\texttt{cum\_months}(t)$ are reported in the left part of both panels (stars) with their linear predictors $\eta_i(t)$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

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 $\boldsymbol{y}^{(C)}(t) =$`cum_months`$(t)$.

		M_1		M_2		M_3	
Longitudinal Process		**Mean**	**[2.5; 97.5]% CI**	**Mean**	**[2.5; 97.5]% CI**	**Mean**	**[2.5; 97.5]% CI**
Intercept	β_0	-0.0254	[-0.4406; 0.3992]	-0.0213	[-0.4445; 0.3699]	-0.0240	[-0.4253; 0.3885]
$B_n(t, \lambda_1)$	β_1	4.4124	[4.3574; 4.4692]	4.4120	[4.3574; 4.4673]	4.4125	[4.3566; 4.4647]
$B_n(t, \lambda_2)$	β_2	6.0006	[5.9299; 6.0692]	5.9997	[5.9302; 6.0678]	6.0009	[5.9325; 6.0668]
$B_n(t, \lambda_3)$	β_3	9.8316	[9.7227; 9.9390]	9.8296	[9.7246; 9.9359]	9.8324	[9.7302; 9.9328]
$B_n(t, \lambda_4)$	β_4	7.6003	[7.5103; 7.6901]	7.5998	[7.5110; 7.6865]	7.6003	[7.5092; 7.6892]
Comorbidity	β_5	-0.3408	[-0.6370; -0.0448]	-0.3414	[-0.6453; -0.0495]	-0.3417	[-0.6446; -0.0523]
	σ_ε	0.1529	[0.1517; 0.1541]	0.1529	[0.1517; 0.1541]	0.1529	[0.1517; 0.1542]
Event Process		**Mean**	**[2.5; 97.5]% CI**	**Mean**	**[2.5; 97.5]% CI**	**Mean**	**[2.5; 97.5]% CI**
Age	θ_1	0.7391	[0.6494; 0.8280]	0.7332	[0.6483; 0.8129]	0.7365	[0.6525; 0.8207]
Gender (Male)	θ_2	0.1824	[0.0548; 0.3107]	0.1922	[0.0687; 0.3193]	0.1918	[0.0567; 0.3218]
Comorbidity	θ_3	0.1968	[0.1380; 0.2558]	0.1978	[0.1377; 0.2533]	0.1970	[0.1397; 0.2549]
Current Value	α_1	-0.0090	[-0.0133; -0.0046]	0.0056	[-0.0032; 0.0146]		
Slope	α_2			-0.6301	[-0.9670; -0.2964]	-0.4537	[-0.6131; -0.2785]
DIC		61439.06		61372.08		61431.62	

On the other hand, considering the dichotomous time-varying variable as longitudinal process $\boldsymbol{y}^{(D)}(t) = \texttt{adherence}(t)$ 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 $\texttt{adherence}(t)$ was the following:

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

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 $\texttt{cum\_months}(t)$ in M_2 but on the current value of $\texttt{adherence}(t)$ 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, $\texttt{cum\_months}(t)$ is the cumulative months covered by ACE/ARB consumption and its slope reflects current adherence. On the other hand, the current value of $\texttt{adherence}(t)$ 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

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 $\boldsymbol{y}^{(D)}(t) = \texttt{adherence}(t)$.

		M_4		M_5		M_6	
Longitudinal Process		**Mean**	**[2.5; 97.5]% CI**	**Mean**	**[2.5; 97.5]% CI**	**Mean**	**[2.5; 97.5]% CI**
Intercept	β_0	-2.0829	[-2.2211; -1.9405]	-2.0905	[-2.2243; -1.9604]	-2.0826	[-2.2120; -1.9528]
Time t	β_1	0.8368	[0.7563; 0.9232]	0.8386	[0.7519; 0.9220]	0.8372	[0.7572; 0.9181]
Comorbidity	β_2	-0.2955	[-0.4166; -0.1753]	-0.3079	[-0.4365; -0.1835]	-0.3072	[-0.4403; -0.1741]
Event Process		**Mean**	**[2.5; 97.5]% CI**	**Mean**	**[2.5; 97.5]% CI**	**Mean**	**[2.5; 97.5]% CI**
Age	θ_1	0.7394	[0.6570; 0.8210]	0.7342	[0.6443; 0.8176]	0.7354	[0.6517; 0.8224]
Gender (Male)	θ_2	0.1813	[0.0569; 0.3175]	0.1787	[0.0545; 0.3058]	0.1836	[0.0501; 0.3107]
Comorbidity	θ_3	0.1958	[0.1371; 0.2512]	0.1961	[0.1365; 0.2539]	0.1960	[0.1363; 0.2547]
Current Value	α_1	-0.0011	[-0.0019; -0.0003]	0.0000	[-0.0018; 0.0019]		
Slope	α_2			-0.0530	[-0.1328; 0.0228]	-0.0511	[-0.0842; -0.0175]
DIC		74869.36		74879.86		74879.11	

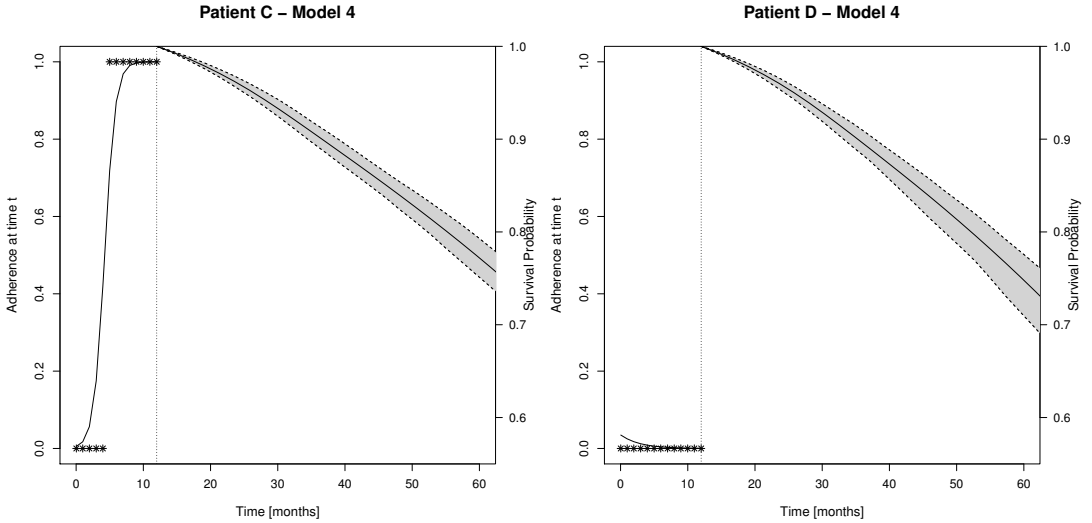

Figure 2.6. Survival probability plots for two male patients aged 72 years and with two comorbidities at the index date using joint model M_4. The values of dichotomous time-varying covariate $\texttt{adherence}(t)$ are in the left part of both panels (stars) with sigmoid transformation of their linear predictors, $\exp(\eta_i(t))/(1+\exp(\eta_i(t))$ (line). In particular, at $t = 12$ months $\texttt{adherence}_C(12) = 1$ and $\texttt{adherence}_D(12) = 0$, 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 $C^{\Delta t}_{dyn}(u)$ introduced by Rizopoulos

Table 2.3. Estimated means along with standard deviations (s.d.) of the integrated prediction errors $IPE(u|t)$ and the dynamic discrimination indexes $C^{\Delta t}_{dyn}(u)$ 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
Calibration	*Mean (s.d.)*	*Mean (s.d.)*
$\widehat{IPE}(u = 12\|t = 3)$	0.0095 (0.0052)	0.0095 (0.0052)
$\widehat{IPE}(u = 60\|t = 12)$	0.1012 (0.0107)	0.1019 (0.0107)
Discrimination	*Mean (s.d.)*	*Mean (s.d.)*
$\widehat{C}^{\Delta t=1}_{dyn}(u = 12)$	0.7051 (0.0794)	0.6994 (0.0342)
$\widehat{C}^{\Delta t=6}_{dyn}(u = 60)$	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 Δt, respectively. For further details on integrated prediction error, dynamic concordance index and their estimates see Rizopoulos (2016) [168]. For each fold k with $k \in \{1, \ldots, 10\}$, we computed the integrated prediction error $IPE_k(u|t)$ and the dynamic concordance index $C^{\Delta t}_{dyn,k}(u)$, 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

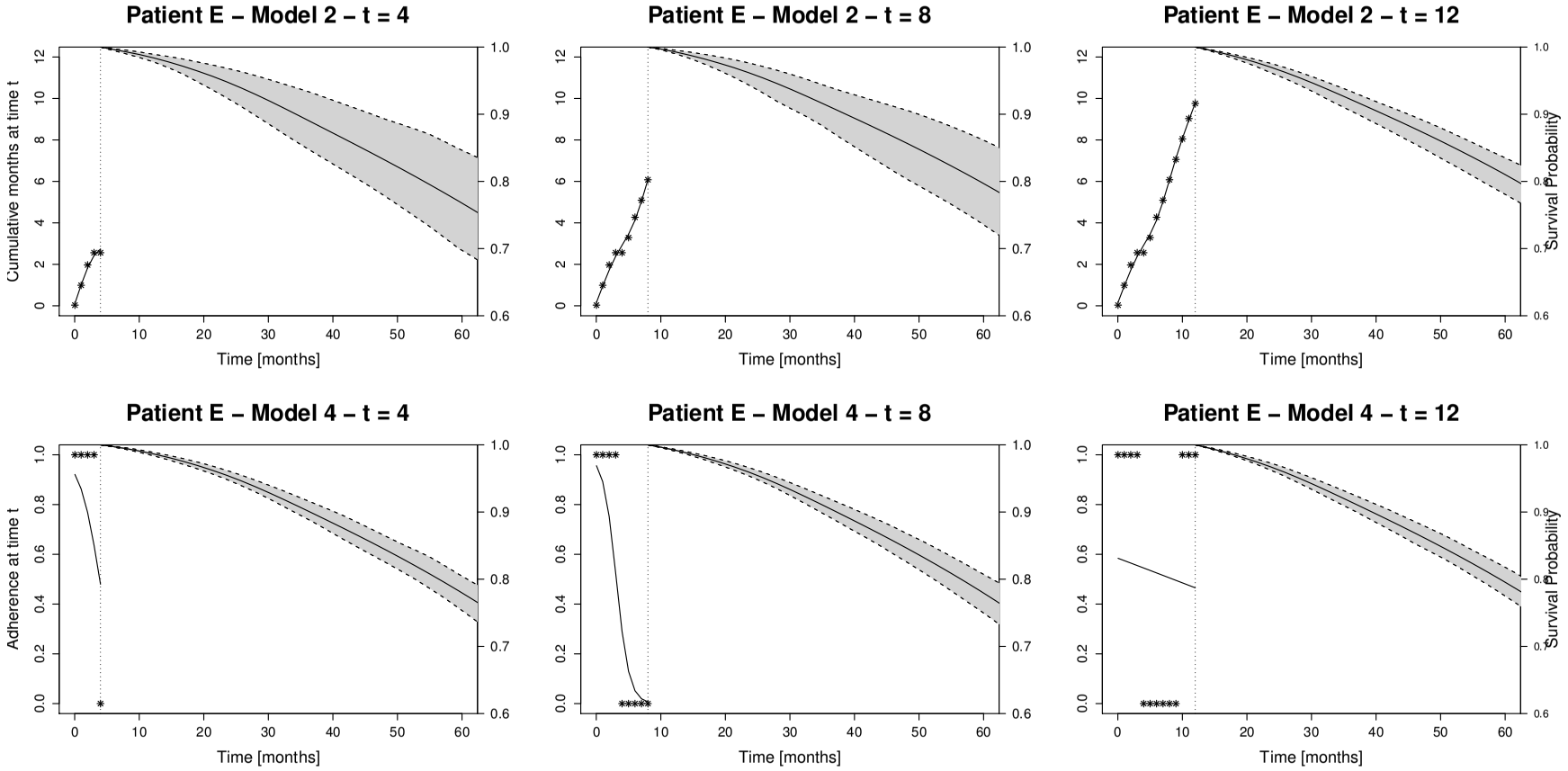

Figure 2.7. Survival probability plots at different time points for the 72 year-old male patient E with two comorbidities at the index date and the pharmacological history shown in Figure 2.3 using joint model M_2 (left panel) and M_4 (right panel). The values of the time-varying covariates $\texttt{cum\_months}_E(t)$ and $\texttt{adherence}_E(t)$ 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 point $t = 4$ months, central panels 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 satisfactory 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 modells, 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

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.