Investigating the effect of drugs consumption on survival outcome of Heart Failure patients using joint models: a case study based on regional administrative data.

Indagine sull'effetto del consumo di farmaci sulla probabilità di sopravvivenza di pazienti con Scompenso Cardiaco mediante l'utilizzo di modelli congiunti: un caso studio basato su dati amministrativi regionali.

Marta Spreafico, Francesca Gasperoni, Francesca Ieva

Abstract In this work, we propose an innovative approach for investigating the effect of drugs consumption on survival outcomes of patients affected by Hearth Failure (HF), a widespread chronic heart disease. In order to achieve this goal, we consider Joint Models approach [7] on administrative dataset of Lombardia Region. In this database several information is collected about patients' pharmacological history, which can be used to recover time-dependent data concerning drug assumptions over time. Through the application of this data, we are able to study the influence of longitudinal processes given by pharmacological treatments consumptions on patients' survival outcomes.

Abstract *In questo lavoro proponiamo un approccio innovativo per indagare l'effetto del consumo di farmaci sulla probabilità di sopravvivenza dei pazienti affetti da Scompenso Cardiaco. A tal fine, consideriamo l'approccio dei modelli congiunti [7] applicandolo al dataset amministrativo di Regione Lombardia. Questo database raccoglie diverse informazioni sulla storia farmacologica dei pazienti, le quali possono essere utilizzate per risalire a variabili tempo-dipendenti rigurdanti il consumo del farmaco nel tempo. Attraverso l'uso di questi dati, siamo in grado di studiare l'influenza del processo longitudinale dato dai trattamenti farmacologici sugli esiti di sopravvivenza dei pazienti.*

Marta Spreafico
MOX, Dipartimento di Matematica, Politecnico di Milano, Piazza Leonardo da Vinci, 32, 20133 Milano, e-mail: marta1.spreafico@mail.polimi.it

Francesca Gasperoni
MOX, Dipartimento di Matematica, Politecnico di Milano, Piazza Leonardo da Vinci, 32, 20133 Milano, e-mail: francesca.gasperoni@polimi.it

Francesca Ieva
MOX, Dipartimento di Matematica, Politecnico di Milano, Piazza Leonardo da Vinci, 32, 20133 Milano, e-mail: francesca.ieva@polimi.it

Key words: Heart Failure, drug consumption, adherence, administrative data, time-varying covariate, joint models.

1 Introduction

Heart Failure (HF) is a chronic cardiac disease, widespread all over the world especially among people over 65 years. About 80,000 new cases per year are recorded [5]. In pharmacoepidemiology the concept of adherence, which generally refers to whether a patient takes a prescribed medication according to schedule [1], is a key factor in effective disease management of many chronic conditions. Drug Utilization Research (DUR) is the branch of pharmacoepidemiology that deals with the use of drugs and it has the goal of facilitating the rational use of drugs in patients populations. For these purposes administrative data allow to measure the effective drug utilization given the limitation of not being able to assert if the patient is currently consuming the dispensed drug. In fact, in order to evaluate the use of a drug we need a statistical measure of consumption. Among possible measures introduced in [9], one of the most used is the Defined Daily Dose (DDD), which is defined as *"the assumed average maintenance dose per day for a drug used for its main indication in adults"*. DDDs can be recovered using the Anatomical Therapeutic Chemical (ATC) classification system introduced by the World Health Organization (WHO) in 1976. ATC code allows to identify phramacological classes and includes information about drug's DDD and routes of administration. DDD is a unit of measure and it does not necessarily correspond to the Prescribed Daily Dose (PDD) by a doctor. In addition to evaluate the assumed drug quantity, we want also to establish if the drug is taken continuously during all the follow up period. There exist lots of different adherence measures. According to [4], we use Proportion of Days Covered (PDC), that is:

$$\text{PDC} = \frac{\text{number of distinct coverage days}}{\text{number of days in the observation period}} \in [0,1] \tag{1}$$

Finally, using PDC, we are able to determine adherent (PDC $\geq$ 0.80) and non-adherent patients or we can categorize them in four levels of PDC, which are $[0;0.25)$, $[0.25;0.5)$, $[0.5;0.75)$ and $[0.75;1]$.

2 The dataset

In the Lombardia Region dataset patients hospitalized for HF from 2000 to 2012 are considered, as described in [6]. For our work, we use a representative sample composed by 1,333,954 events related to 4,872 patients with their first HF hospitalization between 2006-2012.
Each patient, identified by its unique anonymous ID code, is followed from the

starting date (i.e. discharge from first HF hospitalization) until death or censoring. Administrative censoring date is December 31st, 2012. For each patient, age, gender, a list of comorbidities [2] and procedures he/she underwent are recorded.
Moreover, each record in the dataset is related to an event, which can be a hospitalization or a pharmacological prescription. In the first case, the dates of admission and discharge, together with the lenght of stay in hospital are given. In the second one, ATC codes, dates of prescription and coverage days are provided. We focus our work on five pharmacological classes: ACE-Inhibitors (ACE), Angiotensin Receptor Blockers (ARB), Beta-Blocking agents (BB), Anti-Aldosterone agents (AA) and Diuretics (DIU).
Finally, for each type of drug, we calculate patients' PDC and adherence level, setting an observation period of 365 days (one year), as done in [4].

3 Time-Varying covariates

When dealing with longitudinal and/or survival data, time-dependent covariates are often of interest. Since in classical survival models adherence is usually considered as a binary fixed variable, we are interested in representing pharmacological information as a time-varying covariate, which is a more realistic representation. In particular, we want to consider treatments as time-dependent internal covariates, since they are modified according to the development of the illness.
In order to do that, we compute a time-dependent variable which indicates the total days covered by the type of drug up to time t. We set an observation period of 365 days and we consider only distinct days, which means that, in case of overlapping periods between two prescriptions, we consider the first event entirely and only the days of the second one not covered by the first. Furthermore, we hypothesize that all the prescribed types of drug are assumed by patients during the whole period of hospitalization.
Each patient could potentially have five different curves, one for each pharmacological class (ACE, ARB, BB, AA and DIU) depending on which drugs he/she assumes. An example of these types of curves is given in Fig. 1.
In our analysis all types of drug lead to similar results, so in Section 4 we report only those based on ACE to avoid repetitions.

4 Joint Model of survival and drug assumption for HF patients

In 2010 Rizopoulos proposed a Joint Model (JM) for dealing with internal time-dependent covariates [7] and wrote the associated R package `JM` [8]. We use this approach in order to investigate how patients' time-to-event outcome are influenced by longitudinal data (pharmacological treatment curves).
Let $y_i(t)$ denote the value for the longitudinal outcome at time point t for the i-th

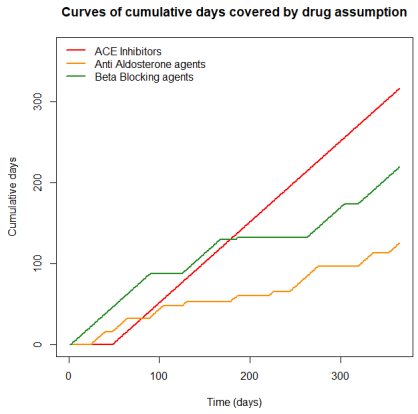

Fig. 1 Examples of time-varying curves of cumulative days covered by drug assumption for a random patient which assumes ACE (red), BB (green) and AA (orange).

subject. $y_i(t)$ is not actually observed at all time points, but only at the very specific occasions t_{ij} so the observed longitudinal data consist of the measurements $y_{ij} = \{y_i(t_{ij}), j = 1,...,n_i\}$. Let $m_i(t)$ denote the true and unobserved value of the longitudinal outcome at time t. Rizopolous supposes that there is a linear relationship between $y_i(t)$ and $m_i(t)$:

$$y_i(t) = m_i(t) + \varepsilon_i(t) = \widetilde{X}_i^T(t)\gamma + Z_i^T(t)b_i + \varepsilon_i(t) \tag{2}$$

where γ is the vector of the unknown fixed effects parameters, b_i is the vector of random effects, $\widetilde{X}_i(t)$ and $Z_i(t)$ denote row vectors of the design matrices for the fixed and random effects and $\varepsilon_i(t)$ is the normally distributed measurement error term with $\varepsilon_i(t) \sim \mathcal{N}(0, \sigma^2)$. Moreover, the random effects b_i are assumed independent of $\varepsilon_i(t)$ and normally distributed with $b_i \sim \mathcal{N}(0, D)$.
To quantify the effect of $m_i(t)$ on the risk for an event, Rizopoulos introduces the following *relative risk model*:

$$\lambda_i(t|\mathcal{M}_i(t), X_i) = \lambda_0(t)\exp\{X_i^T\beta + \alpha m_i(t)\} \tag{3}$$

where $\mathcal{M}_i(t) = \{m_i(u), 0 \leq u < t\}$ denotes the history of the true unobserved longitudinal process up to time point t, $\lambda_0(\cdot)$ denotes the baseline risk function (unspecified or approximated using step functions or spline-based approaches), X_i is a vector of baseline covariates, β is the vector of regression coefficients and α is a parameter that quantifies the effect of the undelying longitudinal outcome to the survival risk for an event.
Finally, coeffients β and α have to be estimated through maximization of the log-likelihood, that is a computationally challenging task for which some numerical approximation methods are needed. In order to do that, we take advantage of the algorithms implemented in `JM` package (see [8] for details).

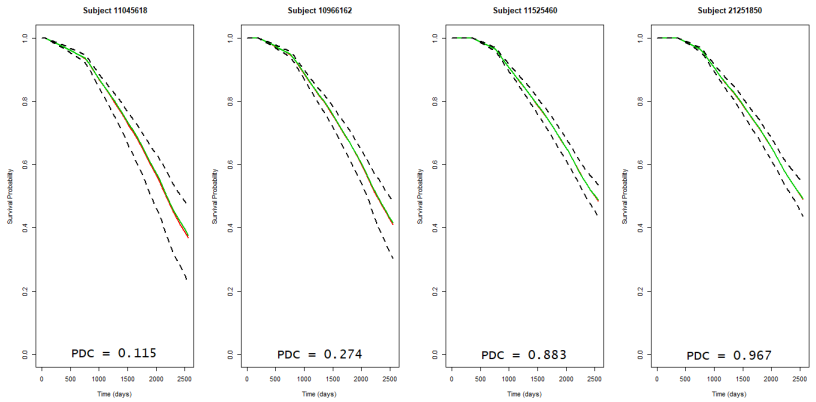

Fig. 2 Survival probability plots for female patients with 80 years old, one hospitalization and two comorbidities. From the left panel patients have a PDC of 0.115, 0.274, 0.833 and 0.967.

In our analysis the longitudinal process $m_i(t)$ is given by the square root of the value of cumulative days covered by drug assumption, whereas the survival process is adjusted according to age, gender, hospitalizations and comorbidities covariates. Moreover, we use a piecewise-constant baseline risk function and the Gauss-Hermite integration rule to approximate integrals in log-likelihood.
It results that all the covariates are significant at 5%, except for gender. In particular, being younger corresponds to a higher survival probability, whereas having a higher number of hospitalizations or of initial comorbidities corresponds to a lower survival risk, as it might be expected. We observe that the higher the value of final PDC, the higher the survival. Furthermore, having a lower PDC leads to larger confidence intervals over time, as shown in Fig. 2, so the uncertainty about the prediction of the survival outcome increases.

5 Conclusion

Modelling the drug assumption process as time-varying covariates in a joint model setting is a promising tool for exploring the effects of pharmacological treatments on survival. For example, it allows us to confirm some pharmacoepidemiological intuition as the fact that medication nonadherence is commonly associated with adverse health conditions [4] in a more suitable way.
Some improvements may be included into the model proposed in (2) in order to provide a more proper modelling of the functional covariate. Moreover, a lot of work is needed in order to include simultaneously all the treatments in a not trivial way. Limitations of administrative data should be overcome through suitable integration of administrative data with clinical registries, as proposed in [3].

References

1. Andrade, S.E., Kahler, K.H., Frech F., Chan, F.A.: Methods for evaluation of medication adherence and persistence using automated databases. In: Pharmacoepidemiology and Drug Safety, **15**, 565574 (2006).
2. Gagne, J.J., Glynn, R.J., Avorn, J., Levin, R., Schneeweiss, S. : A combined comorbidity score predicted mortality in elderly patients better than existing scores. In: Journal of Clinical Epidemiology, **64**(7), 749-59 (2011).
3. Gasperoni, F., Ieva, F., Barbati, G., Scagnetto, A., Iorio, A., Sinagra, G., Di Lenarda, A.: Multi state modelling of Heart Failure care path: a population-based investigation from Italy. In: PlosOne **12**(6): e0179176, (2017).
4. Karve, S., Cleve, M.A., Helm M., Hudson, T.J., West, D.S., Martin, B.C.: Prospective Validation of Eight Different Adherence Measures for Use with Administrative Claims Data among Patients with Schizophreniavhe. In: Value In Health, **12**(6) (2009).
5. Maggioni A.P., Spandonaro, F.: Lo scompenso cardiaco acuto in italia. Giornale italiano di cardiologia, **15**(SUPPL.2 AL N 2), 3S4S (2014)
6. Mazzali, C. 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. In: BMC Health Service Research, **16**(234) (2016).
7. Rizopolous, D.: JM: An R Package for the Joint Modelling of Longitudinal and Time-to-Event Data. In: Journal of Statistical Software, **35**(9), 1-33 (2010).
8. Rizopolous, D.: Package "JM". (June, 2017).
   URL `https://cran.r-project.org/web/packages/JM/JM.pdf`
9. World Health Organization: Introduction to Drug Utilization Research. In: WHO Library Cataloguing-in-Publication Data, (2003).