Multi-State model with nonparametric discrete frailty

Modelli multi-stato con termine di frailty discreto nonparametrico

Francesca Gasperoni, Francesca Ieva, Anna Maria Paganoni, Chris Jackson and Linda Sharples

Abstract In this work, we propose a novel semi-Markov multi-state model with a nonparametric discrete frailty and an application to an administrative clinical database about heart failure patients from a Northern Region of Italy. In particular, we investigate a illness-death model with recovery in which the states space is composed by hospital admission, hospital discharge and death, as unique absorbing state. The available data are grouped longitudinal time-to-event data, indeed for each patient we know the times of admission and discharge of all hospitalizations (2005-2012), the time of death (if it occurs) and the healthcare provider (grouping factor). Thanks to this model, we can investigate the effect of covariates, detect the presence and a pattern of latent populations of healthcare providers across transitions.

Abstract *In questo lavoro proponiamo un modello multi-stato di tipo semi-Markov con una termine random discreto noparametrico e un'applicazione ad un database amministrativo clinico riguardante pazienti affetti da scompenso cardiaco in una regione del Nord Italia. In particolare, ci concentriamo su un modello di tipo illness-death con guarigione, in cui lo spazio degli stati è composto da ammissione in ospedale, dimissione dall'ospedale e morte, unico stato assorbente. I dati che abbi-*

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

Anna Maria Paganoni
MOX, Dipartimento di Matematica, Politecnico di Milano, Piazza Leonardo da Vinci, 32, 20133 Milano, e-mail: anna.paganoni@polimi.it

Chris Jackson
MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge (UK), CB2 0SR, e-mail: chris.jackson@mrc-bsu.cam.ac.uk

Linda Sharples Department of Medical Statistics, London School of Hygiene & Tropical Medicine, Keppel Street, London, WC1E 7HT, e-mail: Linda.Sharples@lshtm.ac.uk

amo a disposizione sono dati longitudinali di tipo tempo all'evento e raggruppati, infatti per ogni paziente siamo a conoscenza dei tempi di ammissione, dimissione (nell'arco di tempo 2005-2012) e momento del decesso, se avvenuto, e la struttura ospedaliera (fattore di raggruppamento). Tramite questo modello possiamo investigare l'effetto delle covariate, individuare la presenza di possibili popolazioni latenti di strutture ospedaliere su ciascuna transizione e possibili pattern fra le transizioni.

Key words: Multi-state model, nonparametric discrete frailty, clinical administrative database

1 Multi-State models with frailty

Multi-State models are mathematical models through which we deal with both competing and progressive events. In multi-state framework, we model the history of a statistical unit (i.e., the clinical history of a patient) through a multi-state process (X_t) $t \geq 0$, where X_t denotes the state occupied by the statistical unit at time t. We define the set of the possible states as $S = \{0, 1, 2, ..., J\}$ and $X_t \in S$. The mathematical quantities that define a multi-state process are the transition probability matrix, **P**(s,t) and the initial state of the process X_0. The elements of the transition probability matrix are defined as:

$$P_{lj}(s,t) := P(X_t = j | X_s = l, H_s), \quad s \leq t \quad j, l \in S; \tag{1}$$

which means that the element (l, j) of **P**(s,t) is the probability of moving from state l to state j between times s and t, given H_s, which is the history of the process up to time s. More formally, H_s is defined as the filtration generated by the process itself, $H_s = \mathscr{F}_{s^-}$. A typical assumption consists in the Markovianity of the process, which states that there is a complete independence between the future and the past:

$$P_{lj}(s,t) := P(X_t = j | X_s = l), \quad s \leq t \quad j, l \in S. \tag{2}$$

Eq (2) states that the next state to be visited and the transition time depends only on the present state. Putter et al. [13] referred to Markov multi-state models as 'clock forward' models, while mentioned 'clock reset' models as semi-Markov multi-state models. According to semi-Markov assumption, the evolution of the process depends on the last occupied state and on the sojourn time. Another important quantity in multi-state models is the transition intensity, which is the instantaneous probability of executing a transition:

$$\lambda_{lj}(t) := \lim_{\Delta t \to 0} \frac{P(X_{t+\Delta t} = j | X_t = l)}{\Delta t}, \quad j, l \in S. \tag{3}$$

If the transition intensity is time-dependent we have time-inhomogenous Markov process, if it is constant we have time-homogenous Markov process.

Several authors explored the theory behind multi-state processes [1] [6] [8] [13], and few packages have been published in R [2] [7]. However, only few authors dealt with multi-state models with frailty. Frailty terms are random terms that are usually introduced in modelling time-to-event data with a twofold aim: on one hand, for taking into account the unexplained heterogeneity of the data and, on the other hand, for taking into account a grouped structure of data. The greatest part of them proposed semi-markov Markov multi-state models [3] [10] [11]. The proposed frailties were generally transition-specific and group-specific (shared frailty) [10] [11], or subject-specific [3] [15]. Ripatti et al. [14] proposed mixed frailty terms, some of which shared across transitions, while Liquet et al. [10] proposed a joint frailty to link two transitions. However, all of them included parametric frailties, such as Gamma, Normal or Compound-Poisson.

2 Semi-Markov multi-state model with a nonparametric discrete frailty term

In this work, we propose a novel multi-state model for grouped and longitudinal data. In particular, we introduce a transition-specific nonparametric discrete frailty term. This kind of frailty has been proposed in the simpler framework of time-to-event data [5] and it allows to avoid any a priori specification of the shape of the frailty and to detect possible clusters of groups (latent populations of known groups, such as hospitals).

The hazard function for individual i in group j in transition l is:

$$\lambda^l(t; X^l_{ij}, w^l_k, z^l_{jk}) = \prod_{k=1}^{K^l} \left[\lambda_0(t)^l w^l_k \exp((X^l_{ij})^T \beta^l) \right]^{z^l_{jk}}. \tag{4}$$

Mathematical assumptions and notations are the same as the ones in time-to-event framework [5]. In Eq.(4), we can recognize a Cox model (a nonparametric part made by an unspecified baseline function $\lambda_0(t)$, and a parametric exponential part made by a vector of patient-specific covariates X_{ij} and the associated regression parameters β) with a nonparametric frailty term $\mathbf{w}^l$. This term is modeled through a random variable with discrete distribution, with an unknown number of points in the support. In particular, we assume that each group j can belong to one latent population k, $k = 1, ..., K^l$, with probability π^l_k. In this case, $[w_1, ..., w_K]^l$ are the points in the support of $\mathbf{w}^l$, K^l is the support's cardinality and $\mathbf{P}\{w^l = w^l_k\} = \pi^l_k$. Also the number of latent population K depends on l, which means that the number of latent population detected is transition-specific. In order to build the model, we introduce an auxiliary indicator random variable z^l_{jk} which is equal to 1 if the j-th group belongs to the k-th population in the l-th transition, so $z^l_{jk} \overset{i.i.d}{\sim} Bern(\pi^l_k)$. The requirement $\sum_{k=1}^{K} z^l_{jk} = 1$, for each j and l, is equivalent to the assumption that each group belongs to only one population in each transition.

3 Application to a clinical administrative database

The proposed semi-Markov multi-state model with nonparametric discrete frailty has been applied to clinical administrative data related to patients with a diagnosis of heart failure and treated in the Lombardia Region, Italy. Despite of the fact that multi-state models are perfect tools for studying chronic conditions, they have been modestly used for investigating heart failure [4].

We propose a illness-death multi-state model with recovery, see Fig. 1, in which the state space S is composed by: hospital admission, hospital discharge and death (as unique absorbing state). We model the hazard function for each possible tran-

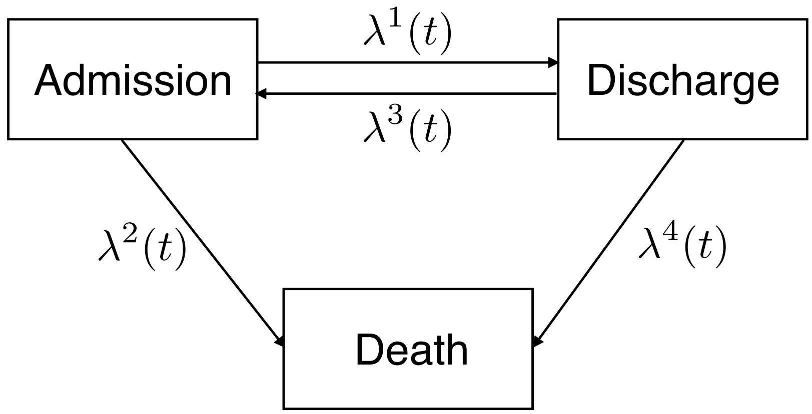

Fig. 1 Multi-state model proposed for the application to clinical administrative data.

sition, $\lambda^l(t)$ ($l \in \{1,2,3,4\}$), according to Eq (4) and we set the same regression parameters for all transitions: age, gender, a binary variable which is 1 if the patient has more than three comorbidities and the number of procedures.

The observed time window in the clinical administrative database that we studied is 2005 - 2012. In order to observe the complete clinical history of patients, we selected a cohort with only those patients whose first discharge was recorded between 2006 and 2007. Moreover, we selected those healthcare providers with 20-1,500 patients, in order to have more robust results. Than, the initial cohort is composed by $40,048$ patients.

For the sake of brevity, we report here only the results related to the first transition. In the first transition, we detected 7 latent populations (see Fig. 2). We observed that a patient hospitalized in a healthcare provider that belongs to latent population 7 has a higher risk (of 3.79 points) with respect to a patient with the same characteristics and hospitalized in a healthcare provider that belongs to latent population 1. The effect of covariates is coherent to what has been observed in a previous work on the same pathology [4]. Aging, having more than three comorbidities and having undergone a procedure lead to a lower risk of transition, which means a longer length of stay. Males have a higher risk of being discharged with respect to women.

The same evaluation can be done for the other transitions. For all of them, we decided to select the number of latent populations according to BIC index, which is, together with AIC, a very popular index in finite mixture model literature [12]. The other index that we show is the one proposed by Laird [9], which tends to overestimate the number of latent populations.

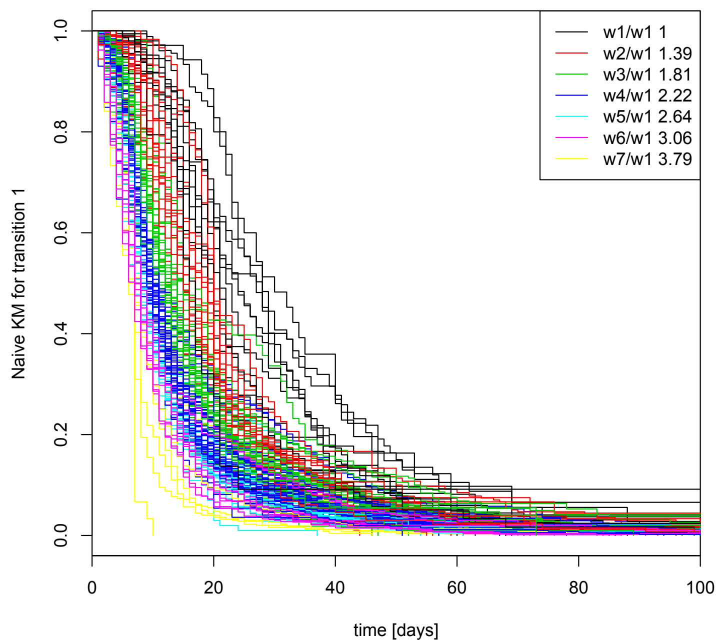

Fig. 2 Naive Kaplan-Meier for transition 1, a zoom in the first 100 days. The KM are naive because the considered censoring is not independent from the realization of the event of interest [13].

4 Conclusions

In this work, there are two sources of novelty: the inclusion of nonparametric discrete frailty in a semi-Markov model and the detection of latent populations of healthcare providers across transitions. The obtained results can be easily read and exploited by healthcare managers in decision making process, but also by clinicians in explorative analysis of clinical administrative databases.

Table 1 Estimates obtained through the proposed procedure in the first transition.

Transition	Latent populations	π	$\mathbf{w}/w_1$	β
1	K_{BIC} = 7 K_{AIC} =10 K_{Laird} = 11	$\pi_1 = 0.09$ $\pi_2 = 0.14$ $\pi_3 = 0.25$ $\pi_4 = 0.28$ $\pi_5 = 0.12$ $\pi_6 = 0.08$ $\pi_7 = 0.04$	$w_1/w_1 = 1$ $w_2/w_1 = 1.39$ $w_3/w_1 = 1.81$ $w_4/w_1 = 2.22$ $w_5/w_1 = 2.63$ $w_6/w_1 = 3.06$ $w_7/w_1 = 3.79$	β_{AGE} = -0.006 β_{SEX} = 0.069 β_{3COM} = -0.267 β_{NPRO} = -0.387

References

1. Andersen, P. K., Keiding, N.: Multi-state models for event history analysis. Statistical methods in medical research, **11**, 91–115. (2002)
2. de Wreede, L. C., Fiocco, M., Putter, H.: mstate: an R package for the analysis of competing risks and multi-state models. Journal of Statistical Software, 38, 1–30. (2011)
3. Foucher, Y., Saint-Pierre, P., Daures, J., Durand, J.: A semi-Markov frailty model for multi-state and clustered survival data. Far East Journal of Theoretical Statistics, **19**, 185. (2006)
4. 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. PloS one. **12**, e0179176. (2017)
5. Gasperoni, F.; Ieva, F.; Paganoni, A.M.; Jackson C.H.; Sharples L.D.: Nonparametric frailty Cox models for hierarchical time-to-event data. MOX report. **45**. (2017)
6. Hougaard, P.: Multi-state models: a review. Lifetime data analysis, **5**, 239–264. (1999)
7. Jackson, C. H.: Multi-state models for panel data: the msm package for R. Journal of Statistical Software, 38(8), 1–29. (2011)
8. Kalbfleisch, J. D., Lawless, J. F.: Likelihood analysis of multi?state models for disease incidence and mortality. Statistics in medicine, **7**, 149–160.(1988)
9. Laird, N.: Nonparametric maximum likelihood estimation of a mixing distribution. Journal of the American Statistical Association, **73**, 805–811. (1978)
10. Liquet, B., Timsit, J. F., Rondeau, V.: Investigating hospital heterogeneity with a multi-state frailty model: application to nosocomial pneumonia disease in intensive care units. BMC medical research methodology, **12**, 79. (2012)
11. Ma, T. Y., Joly, I., Raux, C.: A shared frailty semi-parametric markov renewal model for travel and activity time-use pattern analysis. (2010)
12. McLachlan, G., Peel, D.: Finite mixture models. John Wiley & Sons. (2004)
13. Putter, H., Fiocco, M., Geskus, R. B.: Tutorial in biostatistics: competing risks and multi?state models. Statistics in medicine. **26**, 2389–2430. (2007)
14. Ripatti, S., Gatz, M., Pedersen, N. L., Palmgren, J.: Three-state frailty model for age at onset of dementia and death in Swedish twins. Genetic epidemiology, **24**, 139–149. (2003)
15. Yen, A. M., Chen, T. H., Duffy, S. W., Chen, C. D.: Incorporating frailty in a multi-state model: application to disease natural history modelling of adenoma-carcinoma in the large bowel. Statistical methods in medical research, **19**, 529–546. (2010)