Semi-Markov multistate models to analyze the disease progression of hospitalized COVID-19 patients during the first three waves in the Barcelona metropolitan area

Klaus Langohr[1], *Xavier Piulachs*[1], *Natàlia Pallarès*[2], *Carlota Gudiol*[3], *Cristian Tebé*[2], *Guadalupe Gómez Melis*[1]

[1]klaus.langohr@upc.edu, xavier.piulachs@upc.edu, lupe.gomez@upc.edu, Department of Statistics and Operations Research, Universitat Politècnica de Catalunya · BarcelonaTech, Barcelona, Spain

[2]npallares@idibell.cat, ctebe@idibell.cat, Institut d'Investigació Biomèdica de Bellvitge, Barcelona, Spain

[3]cgudiol@bellvitgehospital.cat, Hospital Universitari de Bellvitge, Barcelona, Spain

Data of more than 4000 hospitalized COVID-19 patients from the Barcelona metropolitan area during the first three waves of the corona virus pandemic are the basis of the present work, which is part of the DIVINE project (***D****ynamic evaluat****I****on of COV****I****D-19 cli****N****ical stat****E****s and their prognostic factors to improve the intrahospital patient management*). The work focuses on properly addressing two specific features in the setting of multistate models. First, the traditionally assumed Markov property is subject under examination by using a recently-published Cox-model-based procedure for testing the Markov assumption in a given transition. This test departs from previous approaches in that it provides assessment for virtually any Markov-free multistate modeling context. Second, the multicohort nature of the data set is modelled in two different ways: on the one hand, cohort-specific fixed effects are included in the transition-specific Cox model. On the other hand, stratified Cox models with cohort-specific baseline hazard functions are fitted to the data. This cohort effect, considered either in isolation or interacting with any covariate, is expected to capture underlying disease patterns that are not explicitly collected by other explanatory covariates. The main objective of our proposed models is to provide a general procedure to analyze complex disease processes that may sequentially affect different population cohorts.

The multistate model under study considers two initial states (no severe pneumonia and severe pneumonia), three transient states (recovery; noninvasive mechanical ventilation; and invasive mechanical ventilation), and two absorbing states (discharge and death), and a total of 14 transitions between two subsequent states. The transition-specific hazards are modeled with semi-parametric Cox models that not only account for the effect of transition-specific baseline covariates, but also for the potential impact of the sojourn time at a previous state what converts them in semi-Markov models.

We discuss the strengths and inconveniences of the semi-Markov models and provide graphical tools to illustrate the cohort effect. These added features within our models allow for a better understanding of the biology underlying any pathological process presented in form of sequential cohorts.

Keywords: Stratified Cox model, semi-Markov multistate model.