Semi-Markov multistate model with interval-censored transition times

Xavier Piulachs, Klaus Langohr, and Guadalupe Gómez

IX Retreat GRBIO
Vilanova i la Geltrú (Barcelona)

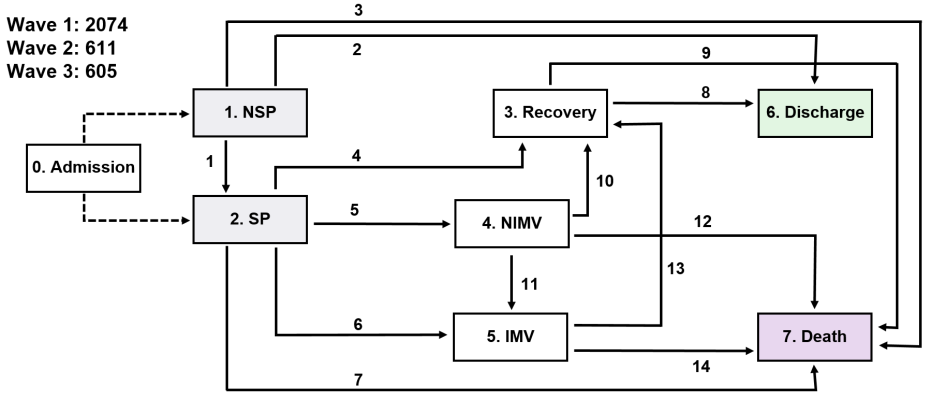
July 11, 2024

Table of Contents

INTRODUCTION (I)

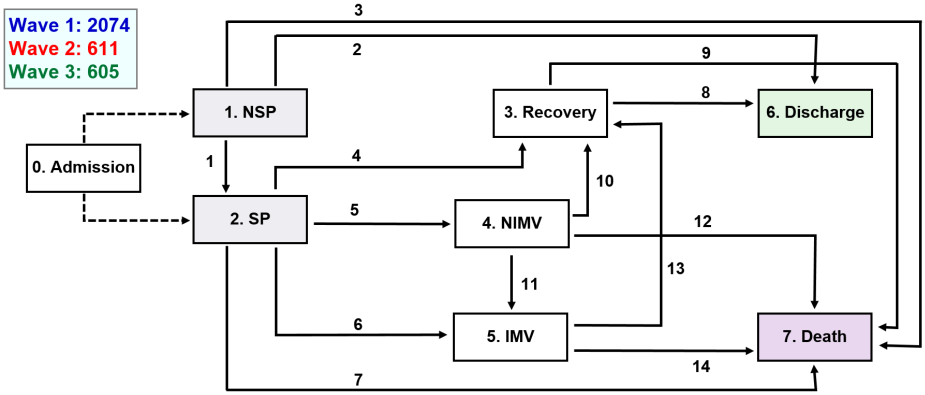
- **General study**
  Dynamic evaluation of COVID-19 clinical states and their prognostic factors to improve the intra-hospital patient management (DIVINE).
- **Team**
  Biostatisticians and clinicians from UPC, UB, and IDIBELL.
- **General research**
  Analyzing multistate event history data from multiple cohorts.
- **Motivating dataset**
  3290 COVID-19-hospitalized adults in the southern Barcelona metropolitan area during the first three pandemic waves.

INTRODUCTION (II)



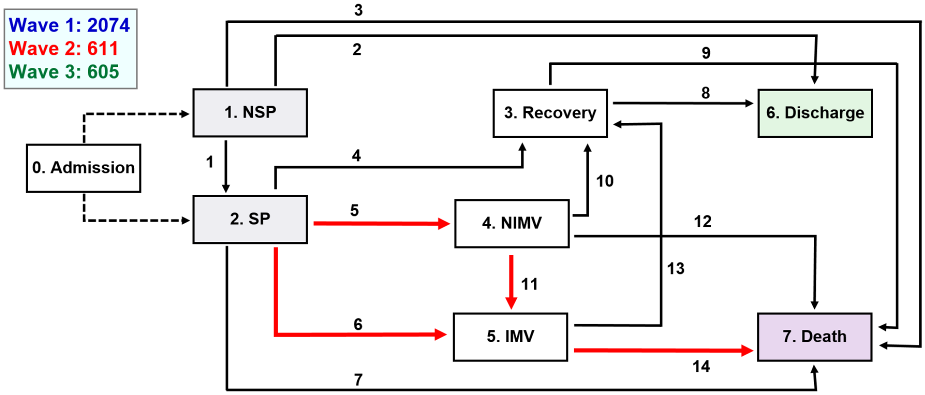
The analysis of the multistate process involves distinct challenges, some of which have already been undertaken.

INTRODUCTION (II)



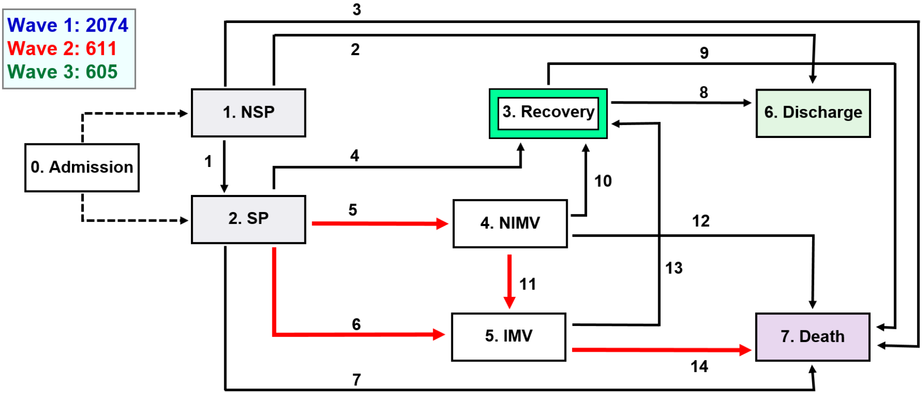
1) Accommodation of the cohort effect: As a fixed covariate or as a stratum variable.

INTRODUCTION (II)



2) Analysis of the key medical transitions, while testing the validity of the Markov property.

INTRODUCTION (II)



3) We know the exact value of all transition times except for the Recovery state. Based on medical criteria, a two-day stay was assumed.

Table of Contents

RESEARCH QUESTION (I)

- We turn our attention to this last point, where unobserved transition times to the $S \equiv$ Recovery state, say T_S^*, were deterministically replaced.
- This assignment allows for adjusting the multistate model by using standard software **(Putter, 2007)**.
- However, in each case we are summarizing all uncertainty with a unique value, which could lead to inaccurate estimates.

RESEARCH QUESTION (I)

- We turn our attention to this last point, where unobserved transition times to the $S \equiv$ Recovery state, say T_S^*, were deterministically replaced.
- This assignment allows for adjusting the multistate model by using standard software **(Putter, 2007)**.
- However, in each case we are summarizing all uncertainty with a unique value, which could lead to inaccurate estimates.

Alternative treatment for these unobserved times?

RESEARCH QUESTION (II)

Ideas:

- We can take advantage of the known ICU release time for some of those transitioning through $S \equiv$ Recovery, treating this time as T_S^*.
- When T_S^* is unobserved, this is not missing but partially known, as it lies between the transition times of the states immediately preceding and following Recovery $\Longrightarrow$ **interval censoring**.
- The Cox partial likelihood function is no longer valid, but alternative ways to continue using standard procedures can be found.

RESEARCH QUESTION (II)

Ideas:

- We can take advantage of the known ICU release time for some of those transitioning through $S \equiv$ Recovery, treating this time as T_S^*.
- When T_S^* is unobserved, this is not missing but partially known, as it lies between the transition times of the states immediately preceding and following Recovery $\Longrightarrow$ **interval censoring**.
- The Cox partial likelihood function is no longer valid, but alternative ways to continue using standard procedures can be found.

Multiple imputation techniques within the scope of multistate models

Table of Contents

MODELING PROCEDURE (I)

General framework

- Consider the states $\mathcal{R} = \{1, \ldots, R\}$, and the stochastic process $\{Y_i(t),\, t \geq 0\}, i = 1, \ldots, n$.
- Heterogeneous population with $G > 1$ cohorts. Each subject's cohort, $g = 1, \ldots, G$, is controlled by $\{c_{gi},\ g \geq 2\}$, so $c_{gi} = 1$ if subject is in cohort $g \in \{2, \ldots, G\}$ and $c_{gi} = 0$ otherwise.
- Cox-based transition hazards between consecutive states $\{\ell, m\} \in \mathcal{R}$, including cohort as a stratum variable.
- Test Markov rule conditional on covariates **(Titman and Putter, 2022)**, accommodating semi-Markov processes by including the time of entry into the current state ℓ.

MODELING PROCEDURE (II)

Stratum-cohort Cox regression model

$$h_i^{\ell m}(t \mid \mathbf{x}_i, t_{\ell i}, c_{gi} = 1) = h_{0g}^{\ell m}(t) \exp\left\{ (\boldsymbol{\beta}^{\ell m})^\top \mathbf{x}_i + \gamma^{\ell m} t_{\ell i} + \eta_{g \times x_q}^{\ell m} x_{qi} \right\}.$$

- $h_{0g}^{\ell m}(t)$: baseline hazard function for the gth cohort;
- $\boldsymbol{\beta}^{\ell m}$: vector of p regression coefficients corresponding to $\mathbf{x}_i$;
- $\gamma^{\ell m}$: regression coefficient for time $t_{\ell i}$ at which state ℓ is reached;
- $\eta_{g \times x_q}^{\ell m}$: interaction term that relates the gth cohort to the covariate x_{qi}.

When rejecting $\mathrm{H}_0: \eta_{2 \times x_q}^{\ell m} = \ldots = \eta_{G \times x_q}^{\ell m} = 0$, the change in $h_i^{\ell m}(\cdot)$ for a Δ-unit increase in qth covariate is quantified by

$$\mathrm{HR}_{x_q \mid g}^{\ell m} = \exp\left\{ \left(\beta_{x_q}^{\ell m} + \eta_{g \times x_q}^{\ell m} \right) \Delta \right\}.$$

MODELING PROCEDURE (III)

Interval-censored transition times

- Consider $\widetilde{n} \leq n$ subjects undertaking the $L \longrightarrow S \longrightarrow U$ path, where the jth subject, $j = 1, \ldots, \widetilde{n}$, has known times T^*_{Lj} and T^*_{Uj}.
- Among the $\widetilde{n}$ subjects transitioning to the S state, $\widetilde{n}_{obs}$ have observed times and $\widetilde{n}_{ic}$ interval-censored times: $\widetilde{n} = \widetilde{n}_{obs} + \widetilde{n}_{ic}$. The focus is on the latter, where only $T^*_{Sj} \in (T^*_{Lj}, T^*_{Uj})$ is known.
- The process consists of three computational steps.

MODELING PROCEDURE (IV)

STEP 1

Assume $T^*_{Sj} \sim$ Weibull $\{\lambda(\mathbf{x}_j),\ \kappa\}$ **(Alarcón et al., 2019)**. This yields the log-linear regression

$$\log T^*_{Sj} = \gamma_0 + \boldsymbol{\gamma}^\top \mathbf{x}_j + \sigma\, \varepsilon_j, \quad \varepsilon_j \sim \text{Gumbel}(0, 1),$$

where the ML estimates for $\{\gamma_0, \boldsymbol{\gamma}, \sigma\}$ in the $L \longrightarrow S \longrightarrow U$ path are derived **(Therneau, 2024)**. Thus, $\lambda(\mathbf{x}_j) = \exp[-\{\gamma_0 + \boldsymbol{\gamma}^\top \mathbf{x}_j\} \,/\, \sigma]$ and $\kappa = 1/\sigma$ are obtained.

STEP 2

The D imputations for each T^*_{Sj} are drawn from the Weibull distribution, confined to $(T^*_{Lj},\ T^*_{Uj})$. The dth dataset, $d = 1, \ldots, D$, has now either known or right-censored transition times, so a multistate model can be fitted. This provides D sets of estimates on θ: $\{\hat{\theta}_d,\ \widehat{\text{SE}}(\hat{\theta}_d)\}$.

STEP 3

The derived estimates are properly averaged **(Rubin, 2004)**:

$$\hat{\theta}_{av} = \frac{1}{D}\sum_{d=1}^{D}\hat{\theta}_d$$

$$\widehat{\mathsf{SE}}_{av}(\hat{\theta}) = \left\{\frac{1}{D}\sum_{d=1}^{D}\widehat{\mathsf{SE}}^2(\hat{\theta}_d) + \frac{(D+1)}{D}\frac{1}{(D-1)}\sum_{d=1}^{D}(\hat{\theta}_d - \hat{\theta}_{av})^2\right\}^{1/2}.$$

MODELING PROCEDURE (V)

STEP 3

The derived estimates are properly averaged **(Rubin, 2004)**:

$$\hat{\theta}_{av} = \frac{1}{D} \sum_{d=1}^{D} \hat{\theta}_d$$

$$\widehat{\mathsf{SE}}_{av}(\hat{\theta}) = \left\{ \frac{1}{D} \sum_{d=1}^{D} \widehat{\mathsf{SE}}^2(\hat{\theta}_d) + \frac{(D+1)}{D} \frac{1}{(D-1)} \sum_{d=1}^{D} (\hat{\theta}_d - \hat{\theta}_{av})^2 \right\}^{1/2}.$$

Table of Contents

ANALYSIS OF EMPIRICAL DATASET (I)

- Prognostic covariates included at a given transition are sex, age, and the percentage of inspired oxygen (21% at room air).
- The time of entry into the SP state for the non-Markovian transitions: SP $\to$ NIMV and SP $\to$ IMV.
- To exemplify multiple imputation procedure, consider a new subject profile k from our target population. We focus on computing the conditional probability of occupying state m at time t, given: state ℓ at time $s < t$, $\mathbf{x}_k$, $t_{\ell k}$, and the gth wave ($g \in \{1, 2, 3\}$):

$$\theta(t) \equiv \Pr\{Y_k(t) = m \mid Y_k(s) = \ell,\, \mathbf{x}_k,\, t_{\ell k},\, c_{gk} = 1\}.$$

EXAMPLE

A low-risk and a high-risk 50-year-old male who are in the $\ell =$ SP state at time $s = 0$, with different health forecasts. We focus on the conditional probability of occupying either the Discharge or Death states within the gth wave, $g \in \{1, 2, 3\}$, at times $t \in \{10, 20, 30, 40, 50\}$, but some transition times to $S \equiv$ Recovery, T^{*}_{Sj}, are interval censored.

Path 1: $\widetilde{n}_1 = 409$ and $\widetilde{n}_{ic1} = 9$

$L1$ (SP) $\longrightarrow$ $S1$ (Recovery) $\longrightarrow$ $U1$ (Discharge)

Path 2: $\widetilde{n}_2 = 263$ and $\widetilde{n}_{ic2} = 150$

$L2$ (NIMV) $\longrightarrow$ $S2$ (Recovery) $\longrightarrow$ $U2$ (Discharge).

ANALYSIS OF EMPIRICAL DATASET (III)

STEP 1

Assume a Weibull distribution for each path, $T^*_{S1j} \sim$ Weibull $\{\lambda_1(\mathbf{x}_{j1}),\ \kappa_1\}$ and $T^*_{S2j} \sim$ Weibull $\{\lambda_2(\mathbf{x}_{j2}),\ \kappa_2\}$ and estimate the parameters.

STEP 2

Impute $D = 25$ values for each $T^*_{Sj} \Longrightarrow$ 25 completed datasets fitted by the multistate model. Compute transition probabilities for low- and high-risk profiles: $\{\hat{\theta}_d(10), \ldots, \hat{\theta}_d(50)\}_g$ and $\{\widehat{\mathsf{SE}}(\hat{\theta}_d(10)), \ldots, \widehat{\mathsf{SE}}(\hat{\theta}_d(50))\}_g$.

STEP 3

The 25 estimates for $\theta(t)$ are averaged within gth wave at time t:

$$\hat{\theta}_{av}(t) = \frac{1}{25}\sum_{d=1}^{25}\hat{\theta}_d(t)$$

$$\widehat{\mathsf{SE}}_{av}\{\hat{\theta}(t)\} = \left[\frac{1}{25}\sum_{d=1}^{25}\widehat{\mathsf{SE}}^2\{\hat{\theta}_d(t)\} + \frac{26}{25}\frac{1}{24}\sum_{d=1}^{25}\{\hat{\theta}_d(t) - \hat{\theta}_{av}(t)\}^2\right]^{1/2}.$$

ANALYSIS OF EMPIRICAL DATASET (IV)

- Using the estimates from each prespecified time point, the evolution of conditional survival probabilities for low-risk and high-risk profiles in each wave can be reconstructed.
- For a given profile and wave, the trajectory is smoothed using B-splines.

ANALYSIS OF EMPIRICAL DATASET (V)

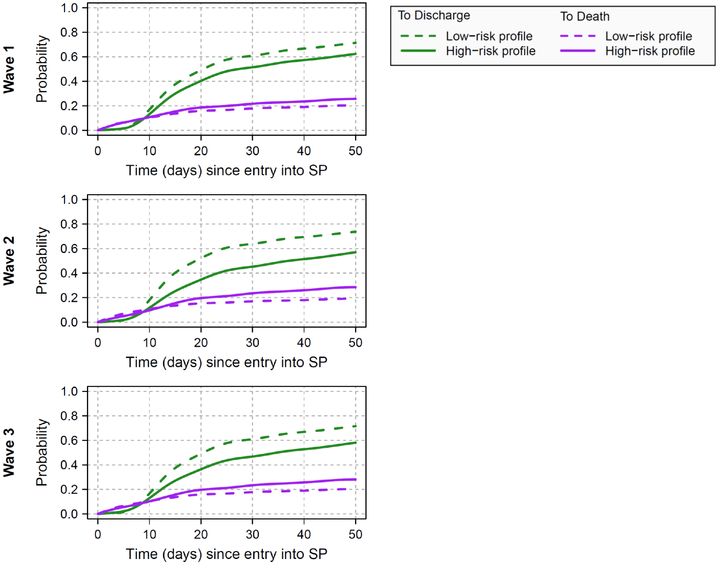


Table of Contents

FINAL REMARKS

- Current work involves a preliminary approach to multiple imputation methods within the scope of a Cox-based multistate modeling framework.
- It would be advisable to analyze the performance of the method across different numbers of imputations, as well as its dependence on the specification of a particular parametric model for interval-censored transition times.
- The introduction of subject-specific random effects could be considered.

BIBLIOGRAPHY

- Titman, A.C., Putter, H. (2022). General tests of the Markov property in multi-state models. Biostatistics, 23, 380–396.
- Alarcón-Soto, Y., Langohr, K., Fehér, C., García, F., and Gómez, G. (2019). Multiple imputation approach for interval-censored time to HIV RNA viral rebound within a mixed effects Cox model. Biometrical Journal, 61, 299–318.
- Therneau, T.M. (2024). A Package for Survival Analysis in R. Package version 3.7-0.
- Putter, H., Fiocco, M., Geskus, R.B. (2007). Tutorial in biostatistics: Competing risks and multi-state models. Statisatics in Medicine, 26, 2389–2430.
- Rubin, D. B. (2004). Multiple imputation for nonresponse in surveys. John Wiley & Sons, Hoboken, New Jersey, USA.

Thank you for your attention