**Supplementary**

Appendix S1. Implementation and estimates of Fine-Gray model

Appendix S2. Internal validation by bootstrap

Figure S1. Survival curves for derivation and validation cohort (One patient in the validation cohort was excluded due to missing LOS).

Table S1. “Baseline” survival probability (Wuhan, China)

Table S2. Thresholds and corresponding proportion and death toll included in each risk group

Table S3. Basic characteristics used in entropy balancing in Derivation cohort, New York cohort and Lombardy cohort

Table S4. “Baseline” survival probability (New York, USA)

Table S5. “Baseline” survival probability (Lombardy, Italy)

Table S6 Methodology quality assessment based on PROBAST risk of bias assessment tool

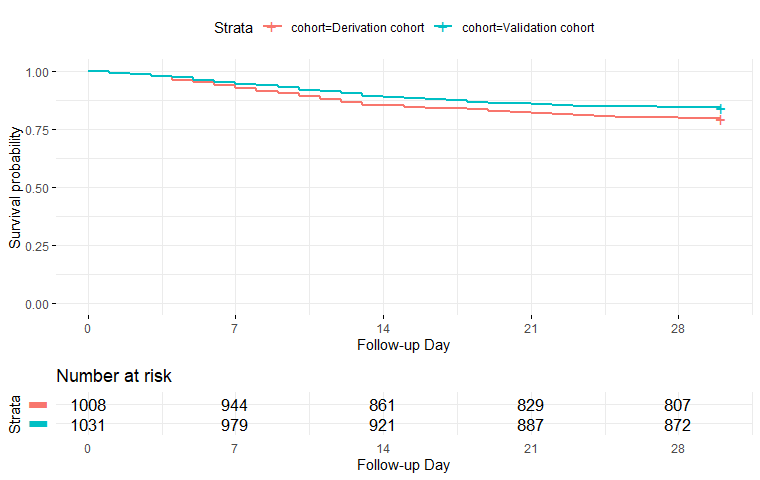
**Appendix S1. Implementation and estimates of Fine-Gray model**

Since there was no patients lost to follow-up in our study, conventional Cox models gives the same estimate as Fine-Gray models by treating discharged patients as being right censored at maximum follow-up time (defined as 30 days after hospital admission) irrespective of the actual length of hospital stay.1 A valid nonparametric estimates of the cumulative incidence function could also be obtained by Kaplan-Meier method using this data coding approach.1

**Appendix S2. Internal validation by bootstrap**

We performed internal validation to estimate the optimism (the level of model overfitting) and adjusted measures of C-index and calibration slope by bootstrapping 1000 samples of the original data. The model derivation process was repeated in each bootstrap dataset and then created 1000 bootstrap models (1000 apparent C-indice). We applied these 1000 bootstrap models to the original dataset (1000 validated C-indice). Optimism in C-index was determined through the difference between the average apparent C-index and the average validated C-index. Optimism-adjusted C-index was obtained by subtracting the optimism from the original C-index. Calibration slope was calculated by estimating the regression coefficient of the PI from the 1000 bootstrap models in the original dataset. Average calibration slope was obtained to be a uniform shrinkage factor.

1. Austin PC, Lee DS, D'Agostino RB, et al. Developing points-based risk-scoring systems in the presence of competing risks. *Stat Med* 2016;35(22):4056-72. doi: 10.1002/sim.6994 [published Online First: 2016/05/21]



**Figure S1. Survival curves for derivation and validation cohort**

**Table S1. “Baseline”† survival probability (Wuhan, China)**

|  |  |  |  |
| --- | --- | --- | --- |
| Day | Survival | Day | Survival |
| 1 | 99.68% | 16 | 91.62% |
| 2 | 99.35% | 17 | 91.42% |
| 3 | 98.96% | 18 | 91.21% |
| 4 | 98.33% | 19 | 90.71% |
| 5 | 97.82% | 20 | 90.26% |
| 6 | 97.15% | 21 | 89.88% |
| 7 | 96.50% | 22 | 89.57% |
| 8 | 95.88% | 23 | 89.18% |
| 9 | 95.45% | 24 | 88.78% |
| 10 | 94.73% | 25 | 88.69% |
| 11 | 93.98% | 26 | 88.61% |
| 12 | 93.19% | 27 | 88.53% |
| 13 | 92.49% | 28 | 88.37% |
| 14 | 92.29% | 29 | 88.20% |
| 15 | 91.76% | 30 | 88.03% |

† this baseline refers to an “average” patient with PI 0.5662.

**Table S2. Thresholds and corresponding proportion and death toll included in each risk group**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Risk group | Range in PI | Derivation | | Validation | |
| Proportion | Death | Proportion | Death |
| 1. Low risk | <= -0.81 | 16.0% | 3 | 15.6% | 0 |
| 2. Low-Intermediate risk | -0.81 to 0.50 | 34.0% | 17 | 32.5% | 8 |
| 3. Moderate risk | 0.50 to 2.03 | 34.0% | 70 | 38.6% | 65 |
| 4. High risk | > 2.03 | 16.0% | 117 | 13.8% | 89 |

**Table S3. Basic characteristics used in entropy balancing in Derivation cohort, New York cohort and Lombardy cohort**

|  | Derivation cohort  (n=1008) | New York cohort  (n=5700) |  | Lombardy cohort (n=1591) |
| --- | --- | --- | --- | --- |
| Age, years | 55 (44-65) | 63 (52-70) |  | 63 (56-70) |
| Sex, female | 439 (43.6%) | 2263 (39.7%) |  | 287 (18%) |
| Hypertension | 232 (23.0%) | 3026 (56.6%) |  | 509 (49%) |
| Coronary heart disease | 32 (3.2%) | 595 (11.1%) |  | 223 (21%) |
| Diabetes | 110 (10.9%) | 1808 (33.8%) |  | 180 (17%) |
| Malignancy | 31 (3.1%) | 320 (6%) |  | 81 (8%) |

**Table S4. “Baseline”† survival probability (New York, USA)**

|  |  |  |  |
| --- | --- | --- | --- |
| Day | Survival | Day | Survival |
| 1 | 99.67% | 16 | 91.89% |
| 2 | 99.05% | 17 | 91.81% |
| 3 | 98.67% | 18 | 91.66% |
| 4 | 97.96% | 19 | 90.98% |
| 5 | 97.58% | 20 | 90.52% |
| 6 | 96.69% | 21 | 90.06% |
| 7 | 95.99% | 22 | 89.55% |
| 8 | 95.57% | 23 | 89.32% |
| 9 | 95.18% | 24 | 88.80% |
| 10 | 94.60% | 25 | 88.76% |
| 11 | 93.96% | 26 | 88.74% |
| 12 | 93.48% | 27 | 88.71% |
| 13 | 92.81% | 28 | 88.64% |
| 14 | 92.64% | 29 | 88.28% |
| 15 | 91.99% | 30 | 88.06% |

† this baseline refers to an “average” patient with PI 0.5662.

**Table S5. “Baseline”† survival probability (Lombardy, Italy)**

|  |  |  |  |
| --- | --- | --- | --- |
| Day | Survival | Day | Survival |
| 1 | 99.64% | 16 | 92.52% |
| 2 | 99.20% | 17 | 92.41% |
| 3 | 98.82% | 18 | 92.28% |
| 4 | 98.18% | 19 | 91.66% |
| 5 | 97.84% | 20 | 91.40% |
| 6 | 97.23% | 21 | 91.28% |
| 7 | 96.75% | 22 | 90.49% |
| 8 | 96.27% | 23 | 90.18% |
| 9 | 95.98% | 24 | 89.89% |
| 10 | 95.41% | 25 | 89.87% |
| 11 | 94.55% | 26 | 89.84% |
| 12 | 94.06% | 27 | 89.79% |
| 13 | 93.61% | 28 | 89.68% |
| 14 | 93.18% | 29 | 89.56% |
| 15 | 92.64% | 30 | 89.28% |

† this baseline refers to an “average” patient with PI 0.5662.

**Table S6 Methodology quality assessment based on PROBAST risk of bias assessment tool**

|  |  |  |
| --- | --- | --- |
| **Question** | **Answer** | **Rationale** |
| **Domain 1: Participants** |  |  |
| 1·1 Were appropriate data sources used, e.g., cohort, RCT or nested case-control study data? | Yes | Cohort design was used. |
| 1·2 Were all inclusions and exclusions of participants appropriate? | Probably yes | The patients had no event (either death or discharge) at the end of follow-up date were excluded from the model development cohort. This is not model developers’ choice, but due to data availability. Given the time between end of accrual and end of follow-up was 40 days, the actual proportion of excluded patients was very small, and the potential risk of bias can be mitigated. |
| **Overall risk of bias of Domain 1** | **Low risk of bias** |  |
| **Domain 2: Predictors** |  |  |
| 2·1 Were predictors defined and assessed in a similar way for all participants? | Probably yes | Most of the predictors were patient characteristics and laboratory tests, which had standard definition. The scope of chronic disease history may be slightly different between the two hospitals. In general, all predictors can be considered as assessed in a similar way. |
| 2·2 Were predictor assessments made without knowledge of outcome data? | Yes | All predictors were collected at patients admission. |
| 2·3 Are all predictors available at the time the model is intended to be used? | Yes | All predictors included in the final model were patient characteristics and laboratory tests, which were easy to access· |
| **Overall risk of bias of Domain 2** | **Low risk of bias** |  |
| **Domain 3: Outcome** |  |  |
| 3·1 Was the outcome determined appropriately? | Yes | In-hospital death was an event of interest for admitted patients with COVID-19. Discharge was properly considered as a competing risk event of in-hospital death, this is an advantage of this study compared with other studies (in which discharge was censored). |
| 3·2 Was a pre-specified or standard outcome definition used? | Yes | In-hospital death was a hard end-point. |
| 3·3 Were predictors excluded from the outcome definition? | Yes | In-hospital death was objective. |
| 3·4 Was the outcome defined and determined in a similar way for all participants? | Yes | In-hospital death was objective. |
| 3·5 Was the outcome determined without knowledge of predictor information? | Yes | In-hospital death was objective. |
| 3·6 Was the time interval between predictor assessment and outcome determination appropriate? | Yes | All patients were staying in the hospitals and continuously followed up till discharge, death or the end of follow-up (which was 40 days after end of accrual). The time interval between predictor assessment and outcome determination is sufficient. |
| **Overall risk of bias of Domain 3** | **Low risk of bias** |  |
| **Domain 4: Analysis** |  |  |
| 4·1 Were there a reasonable number of participants with the outcome? | Yes | The total number of candidate variables was 11, and the number of events was 211, so the number of events per variable (EPV) =211/11≈19, which was reasonable. |
| 4·2 Were continuous and categorical predictors handled appropriately? | Yes | Functional form of the relation between continuous variables and outcome was investigated, and suitable variable transformations were performed in case the linear assumption did not hold. No categorization was done for continuous variables. |
| 4·3 Were all enrolled participants included in the analysis? | Yes | All patients met the inclusion criteria were included in the analysis. |
| 4·4 Were participants with missing data handled appropriately? | Probably yes | Predictors with high percentage of missing value were excluded from the final model. |
| 4·5 Was selection of predictors based on univariable analysis avoided? | Yes | Variable selection was based on multivariable analysis, and univariable was performed for investigating the non-linear relation. |
| 4·6 Were complexities in the data (e.g., censoring, competing risks, sampling of controls) accounted for appropriately? | Yes | Discharge was properly considered as a competing risk event of in-hospital death in the analysis. |
| 4·7 Were relevant model performance measures evaluated appropriately? | Yes | Discrimination was assessed by C-statistics, and calibration curve was used to assess calibration. |
| 4·8 Were model overfitting and optimism in model performance accounted for? | Yes | Optimism in model performance was adjusted with bootstrapping internal validation, a shrinkage factor was multiplied to the linear predictor to prevent overfitting. |
| 4·9 Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis? | Yes | Coefficients of the multivariable Cox model was directly used in the calculation of the prognostic index (PI). |
| **Overall risk of bias of Domain 4** | **Low risk of bias** |  |