Mortality and health risk factors of hospitalised COVID-19 patients: a statistical analysis of the Brazilian private health insurance portfolio

Jason Khu, z5254974

**Actuarial Models and Statistics (ACTL3141)**

Monday, 18th April 2022

**Abstract**

An analysis of the mortality and health risk factors of hospitalised COVID-19 patients was performed using data on Brazil’s private health insurance portfolio – ranging from 1st January 2021 to 31st December 2021.

Based on the descriptive analysis, there was indication of a positive relationship between (old) age and COVID-19 mortality and health risk, which was then supported by the age-specific transition rate estimates derived for the hospital pathway multistate model. Based on the survival analysis, many covariates like dyspnoea, ARDS, cardiovascular disease, liver disease, diabetes and neurological disease were found to be significant factors that drove higher mortality risk. One unexpected result, which was eventually examined further, was that COVID-19 vaccination also drove higher mortality risk. Finally, it was concluded that vaccination status should not be used as a pricing and under-writing factor for the private health insurance business.

**Introduction**

With 22.29 million cumulative confirmed cases recorded at the end of 2021 (Our World in Data, 2022), the COVID-19 pandemic has brought about new and greater risk to the mortality and health of the Brazilian population.

In this report, descriptive analysis and survival analysis were performed using data on Brazil’s private health insurance portfolio. Furthermore, the report discusses the appropriateness, from an ethical perspective, of using COVID-19 vaccination status as a rating and under-writing criteria for setting private health insurance prices.

**Body**

**Descriptive Analysis**

The dataset contains information on 187,209 patients who have been hospitalised with COVID-19 throughout the 2021 calendar year. Variables that were examined in the descriptive analysis included ICU status, death status, vaccination status, age, and the presence of any comorbidities and symptoms. The distribution of age (nearest birthday) of the patients when admitted to hospital is shown in Fig 1. The distribution of age spans between age 0 and 114 and is close to symmetrical, with the median (61.0), mean (60.9) and mean (58.0) being relatively close together.

There is an indication that age may be a confounding factor of COVID-19 mortality and health risk. In Fig 2, the distribution of age is segmented by ICU admission status: patients admitted to ICU are shown in red while patients not admitted to ICU are shown in blue. The median (62.0) and mean (61.2) age of patients admitted to ICU are slightly higher than the median (61.0) and mean (60.7) of those who were not admitted to ICU.

The relationship between age and COVID-19 mortality and health risk is more pronounced from a mortality perspective. Overall, for this dataset, the death rate is 40.4% and the ICU admission rate is 40.9%. In Fig 3, the distribution of age is segmented by the ‘death’ variable, where patients who have died from COVID-19 are shown in blue while patients who have recovered (and were discharged) are shown in red. The differential between the medians and means of these two groups are significantly high (9.0 and 8.0 years respectively), indicating that (old) age may be a significant confounding factor of mortality risk; this will be explored further in the Survival Analysis section.

The most frequent symptoms and comorbidities are shown near the top of Fig 4 and Fig 5. The symptoms that were experienced by most patients (more than 50% of the population) were dyspnoea

Chart, histogram

Description automatically generated Chart, histogram

Description automatically generated

***Fig 1.*** *Distribution of age (when admitted to hospital)*

***Fig 2.*** *Distribution of age (when admitted to hospital) for admitted to ICU vs not admitted to ICU*

Chart, histogram

Description automatically generated Chart, funnel chart

Description automatically generated

***Fig 4.*** *Most frequent symptoms*

***Fig 3.*** *Distribution of age (when admitted to hospital) for cured/discharged outcome vs death outcome*

Chart

Description automatically generated

***Fig 5.*** *Most frequent comorbidities*

(78.1%), low oxygen saturation (77.7%), coughing (71.5%), respiratory distress (63.8%), and fever (53.6%). On the other hand, the comorbidity that was present in most patients is cardiovascular disease

(52.3%). Diabetes (35.3%) and obesity (19.2%) follow cardiovascular disease as the second and third

most common comorbidities respectively.

There is an indication that COVID-19 vaccination reduces the presence of symptoms, the probability of being admitted to ICU, and the probability of dying from COVID-19. Overall, for this dataset, the proportion of patients who were vaccinated is 39.9%. Table 1 shows the proportion of patients with each symptom – divided by vaccination status, with computed differentials. By looking at the differentials, it appears that COVID-19 vaccination generally reduces the likelihood of patients showing symptoms. Fig 6 shows a split of the population by ICU admission status and vaccination status. From Fig 6, a smaller percentage of vaccinated patients (39.9%, n=29793) were being admitted to ICU compared to unvaccinated patients (41.6%, n=46775). Fig 7 shows a similar split to Fig 6, but instead it is split by the outcome at the end of the observation period (i.e., death vs cured/discharged) and vaccination status. From Fig 7, a slightly smaller percentage of vaccinated patients (40.1%, n=29957) died from COVID-19 compared to unvaccinated patients (40.6%, n=45652).

***Table 1.*** *Frequency of symptoms in vaccinated vs unvaccinated population*

|  |  |  |  |
| --- | --- | --- | --- |
| Symptom | Proportion of vaccinated (%) | Proportion of unvaccinated (%) | Differential (%) |
| Dyspnoea | 76.2 | 79.3 | -3.1 |
| Low oxygen saturation | 77.7 | 77.7 | 0 |
| Cough | 71.2 | 71.8 | -0.6 |
| Respiratory distress | 61.7 | 65.2 | -3.5 |
| Fever | 50.2 | 55.9 | -5.7 |
| Sore throat | 16.3 | 19.0 | -2.7 |
| Diarrhea | 14.2 | 15.7 | -1.5 |
| Vomit | 9.0 | 9.5 | -0.5 |

Chart

Description automatically generated Chart

Description automatically generated

***Fig 7.*** *Population split by outcome and vaccination status*

***Fig 6.*** *Population split by ICU admission status and vaccination status*

**Survival Analysis**

To aid the survival analysis, multiple variables were engineered using existing columns in the dataset (see Appendix A for the definitions of these new variables).

Plots of the Kaplan Meier (KM) survival curves were used to examine the relative effect of each covariate on the survival estimates (see Appendix B for plots). For the data, it is assumed there is non-informative censoring and that the lives were independent. Table 2 shows a comparison between the two curves for each covariate, including p-values from the log-rank and peto-peto test. A “-” was recorded whenever the KM estimates appeared to be similar or whenever there was a clear intersection between the two curves based on the plots.

From Table 2: old age, admission to hospital during autumn, COVID-19 vaccination, dyspnoea, ARDS, low oxygen saturation, cardiovascular disease, liver disease, diabetes, neurological disease, pneumopathy, immunodeficiency and renal disease appear to be significant factors that increase mortality risk. At a significance level of 5%, these covariates had covariate-influenced groups with generally lower survival curves that were significantly different to their baseline.

A Cox proportional hazards model was used to perform unadjusted regression analysis, considering the time between hospitalisation and the outcome at the end of the observation period (i.e., whether they died from COVID-19 or were discharged). Table 3 records the unadjusted hazard rates, confidence intervals and p-values of all covariates. The blue and orange highlights in the table indicate whether there was any violation of the assumptions of the Cox proportional hazards model.

***Table 2.*** *Covariates and their relative effect on the KM survival estimates (with their p-values from the log-rank and peto-peto test).*

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Higher/lower KM survival estimates | p-value  (log-rank test) | p-value  (peto-peto test) |
| Age ≥ 40 | Lower | <0.001 | <0.001 |
| Age ≥ 60 | Lower | <0.001 | <0.001 |
| Age ≥ 80 | Lower | <0.001 | <0.001 |
| Male | - | 0.005 | 0.1 |
| Winter season | - | <0.001 | <0.001 |
| Spring season | - | <0.001 | <0.001 |
| Summer season | Higher | <0.001 | <0.001 |
| Autumn season | Lower | <0.001 | <0.001 |
| Vaccine | Lower | <0.001 | <0.001 |
| Fever | - | <0.001 | <0.001 |
| Cough | Higher | <0.001 | <0.001 |
| Sore throat | - | <0.001 | <0.001 |
| Dyspnoea | Lower | <0.001 | <0.001 |
| ARDS (Acute Respiratory Distress Syndrome) | Lower | <0.001 | <0.001 |
| Oxygen saturation < 95% | Lower | <0.001 | <0.001 |
| Diarrhea | - | <0.001 | <0.001 |
| Vomit | - | <0.001 | <0.001 |
| Cardiovascular disease | Lower | <0.001 | <0.001 |
| Hematologic disease | - | 0.020 | 0.010 |
| Down syndrome | - | 0.800 | 0.800 |
| Liver disease | Lower | <0.001 | <0.001 |
| Asthma | - | <0.001 | <0.001 |
| Diabetes | Lower | <0.001 | <0.001 |
| Neurological disease | Lower | <0.001 | <0.001 |
| Pneumopathy | Lower | <0.001 | <0.001 |
| Immunodeficiency | Lower | <0.001 | <0.001 |
| Renal disease | Lower | <0.001 | <0.001 |
| Obesity | - | <0.001 | <0.001 |

***Table 3.*** *Covariates and their hazard rates (with their 95% confidence interval and p-value) as per the Cox proportional hazards model.*

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Hazard rate | 95% Confidence interval | p-value |
| Age ≥ 40 | 1.688 | (1.637, 1.740) | <0.001 |
| Age ≥ 60 | 1.591 | (1.567, 1.616) | <0.001 |
| Age ≥ 80 | 1.824 | (1.792, 1.857) | <0.001 |
| Male | 1.020 | (1.006, 1.035) | 0.006 |
| Winter season | 0.890 | (0.875, 0.904) | <0.001 |
| Spring season | 0.909 | (0.885, 0.933) | <0.001 |
| Summer season | 0.885 | (0.862, 0.909) | <0.001 |
| Autumn season | 1.177 | (1.160, 1.194) | <0.001 |
| Vaccine | 1.042 | (1.027, 1.058) | <0.001 |
| Fever | 0.875 | (0.862, 0.887) | <0.001 |
| Cough | 0.885 | (0.871, 0.898) | <0.001 |
| Sore throat | 0.929 | (0.911, 0.947) | <0.001 |
| Dyspnoea | 1.227 | (1.205, 1.251) | <0.001 |
| ARDS (Acute Respiratory Distress Syndrome) | 1.263 | (1.243, 1.282) | <0.001 |
| Oxygen saturation < 95% | 1.253 | (1.229, 1.277) | <0.001 |
| Diarrhea | 0.916 | (0.897, 0.935) | <0.001 |
| Vomit | 0.948 | (0.924, 0.973) | <0.001 |
| Cardiovascular disease | 1.146 | (1.130, 1.163) | <0.001 |
| Hematologic disease | 1.088 | (1.015, 1.165) | 0.017 |
| Down syndrome | 0.985 | (0.895, 1.084) | 0.756 |
| Liver disease | 1.224 | (1.156, 1.296) | <0.001 |
| Asthma | 0.829 | (0.796, 0.863) | <0.001 |
| Diabetes | 1.166 | (1.149, 1.183) | <0.001 |
| Neurological disease | 1.337 | (1.300, 1.374) | <0.001 |
| Pneumopathy | 1.302 | (1.264, 1.342) | <0.001 |
| Immunodeficiency | 1.124 | (1.084, 1.166) | <0.001 |
| Renal disease | 1.284 | (1.248, 1.320) | <0.001 |
| Obesity | 0.910 | (0.894, 0.927) | <0.001 |

For the proportional hazards assumption to hold true, the (log) cumulative hazard rates of the baseline and covariate-influenced group need to be parallel (i.e., should not intersect) (see Appendix C for the graphical diagnostics).

From Table 3, out of the covariates that did not violate the proportional hazards assumption, the same variables, as those inferred from Table 2, along with being male and having hematologic disease appear to drive higher mortality risk. These covariates have a 95% confidence interval greater than 1, and a p-value less than 5%.

Despite the lower proportion of people who have died from COVID-19 in the vaccinated population (as per the descriptive analysis), the results from the survival analysis indicate that the probability of a vaccinated patient dying from COVID-19 was higher. This may be because vaccinated patients also had a higher rate of recovery, thereby having a higher propensity to leave the hospital and hence the population at risk (see Appendix D for plots of the transition rates split by vaccination status).

**Transition rates for hospital pathway of patients**

Discrete estimates of the transition rates for this hospital pathway model were derived using maximum likelihood estimates. For the computation of the estimates, the waiting time in each state was expressed in days.

The following assumptions about the records of data were used for the computation of the rates:

1. Age-specific rates are based on the age nearest birthday at the date when the patient was admitted to hospital (i.e., age changes were not considered).
2. A patient who has been admitted and discharged from ICU was not admitted to ICU again.
3. Whenever the ICU admission date was equal to the hospitalisation date, it was assumed that the patient entered the hospital directly to ICU. Otherwise, patients enter the hospital without starting in ICU.
4. Whenever the ICU discharge date was equal to the end-of-observation date, it was assumed that the patient stayed in ICU until the outcome at the end of observation. This means that:
5. Whenever the ICU discharge date was equal to the hospital discharge date, it was assumed that the patient was cured directly from the ICU. This is represented by a transition from I to C.
6. Whenever the ICU discharge date was equal to the death date, this is represented by a transition from I directly to D.
7. Whenever a patient, at any point, gets admitted to the ICU and the hospital discharge date was not equal to the ICU discharge date, it was assumed that the patient remained hospitalised after spending time in ICU. This is represented by a transition from I to H.

The computation of discrete estimates, across all ages, are shown below:

For the age-specific transition rates, the estimates were recalculated for every single age and plotted on a line chart (see Appendix E). To clearly see trends in the transition rates, Fig 8 shows a breakdown of the transition rate estimates by different age intervals, where each age interval spans 10 years – starting from age 0.

In general, based on the graphs, patients of older age have a lower propensity to transition to the cured/discharged state (C) and a higher propensity to transition to the death state (D). This inference is consistent with the findings presented in the survival analysis, where it was found that (old) age was a significant factor influencing mortality risk. This is also consistent with the WHO advice, which suggests that older people face significant risk of developing severe illness if they contract the disease due to the physiological changes that come with ageing (World Health Organisation, 2020).

Chart, line chart

Description automatically generated Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated Chart, line chart

Description automatically generated

***Fig 8.*** *Discrete estimates of the transition rates graphed across age intervals of ten years*

Furthermore, it can be seen that older patients have a lower propensity to transition from the ICU state (I) to the hospitalised (but not in ICU) state (H). Nevertheless, very young (0-20 years old) and very old (80-120 years old) patients have a lower propensity to transition from the hospitalised (but not in ICU) state (H) to the ICU state (I).

**Ethical implications of using COVID-19 vaccination status as a rating criteria**

The ethical implications of using COVID-19 vaccination status as a pricing factor were assessed with consideration of the two primary stakeholders involved: the private health insurer and policyholders. Table 4 lists pros and cons of using vaccination status as a pricing factor.

Overall, COVID-19 vaccination should not be used as an underwriting factor. While it is a significant factor that can contribute to lower mortality and health risk, it discriminates against policyholders who choose not to or are unable to be fully vaccinated. Nevertheless, there are also issues that come up regarding data security and the unforeseen impact of more long-term vaccine side effects.

***Table 4.*** *Pros and cons of using vaccination status as a pricing factor.*

|  |
| --- |
| Pros |
| * **Incentive to get vaccinated.** Using COVID-19 vaccination as a pricing factor can incentivise policyholders to getvaccinated to reduce their likelihood of getting infected and thereby ensure their own health. Based on the results from the descriptive analysis and analysis on the transition rates, vaccination does reduce the probability of dying and being admitted to ICU, and generally improves the rate of recovery in hospital. * **Health/mortality risk prediction accuracy.** Assuming that vaccination is a significant factor driving lower health and mortality risk, using vaccination as an under-writing factor would increase the accuracy of risk prediction. |
| Cons |
| * **Discrimination.** It is discriminatory against policyholders who do not take the vaccine for personal or other health reasons. Based on a 2021 study, 10.5% of people in Brazil were vaccine-hesitant because of concerns regarding vaccine efficacy and vaccine side effects (Moore, 2021). Nevertheless, policyholders who have had adverse side-effects after vaccination may not plan to progress towards being fully vaccinated. * **Sensitive information and data security.** Policyholders’ vaccination status is sensitive information. Policyholders may have concerns regarding data security when sharing this information with the insurer. * **On-going examination of vaccination side effects.** While vaccination may be a significant factor in predicting health and mortality risk, there is still on-going examination of the side effects of vaccination. These side effects may present a conflicting view on vaccination and its effect on mortality and health risk and can thereby add a cost for the insurer in the future. As an example, earlier in 2022, there has been a study discovering significantly increased risk of myocarditis after vaccination (Oster, 2022). Myocarditis is a cause of cardiovascular disease. |

**Conclusion**

The results of the survival analysis, suggest that COVID-19 vaccination increases mortality risk. However, based on the descriptive analysis, combined with the computation of the transition rates, COVID-19 vaccination reduces the probability of dying and being admitted to ICU, and generally improves the rate of recovery in hospital.

It was recommended that COVID-19 vaccination should not be used as an underwriting factor due to discrimination, sharing and protection of sensitive information, and unforeseen vaccine side effects that may change health and mortality predictions.

**References**

[1] Moore, DCBC 2021, ‘Low COVID-19 vaccine hesitancy in Brazil’, *Vaccine*, vol. 39, issue 42, accessed 18 April 2022, <<https://www.sciencedirect.com/science/article/pii/S0264410X21011920>>

[2] Oster, ME 2022, ‘Myocarditis Cases Reported After mRNA-Based COVID-19 Vaccination in the US From December 2020 to January 2021’, JAMA, vol. 327, accessed 18 April 2022, <<https://jamanetwork.com/journals/jama/fullarticle/2788346>>

[3] Our World in Data, 2022, accessed 18 April 2022, <<https://ourworldindata.org/coronavirus/country/brazil>>

[4] World Health Organisation, 2020, accessed 18April 2022, <<https://www.who.int/news-room/feature-stories/detail/who-delivers-advice-and-support-for-older-people-during-covid-19#:~:text=Although%20all%20age%20groups%20are,potential%20underlying%20health%20conditions>>

**Appendices**

**Appendix A: Data dictionary for engineered variables**

|  |  |
| --- | --- |
| Variable | Definition |
| Age ≥ 40 | TRUE if the patient’s age (nearest birthday when the patient was admitted to hospital) was greater than or equal to 40 or FALSE otherwise. |
| Age ≥ 60 | TRUE if the patient’s age (nearest birthday when the patient was admitted to hospital) was greater than or equal to 60 or FALSE otherwise. |
| Age ≥ 80 | TRUE if the patient’s age (nearest birthday when the patient was admitted to hospital) was greater than or equal to 80 or FALSE otherwise. |
| Winter season | TRUE if the patient’s hospitalisation date was during the Winter season (i.e, the month was either June, July or August) or FALSE otherwise. |
| Spring season | TRUE if the patient’s hospitalisation date was during the Spring season (i.e, the month was either September, October or November) or FALSE otherwise. |
| Summer season | TRUE if the patient’s hospitalisation date was during the Summer season (i.e, the month was either December, January or February) or FALSE otherwise. |
| Autumn season | TRUE if the patient’s hospitalisation date was during the Winter season (i.e, the month was either March, April or May) or FALSE otherwise. |

**Appendix B: Kaplan-Meier curves for all variables**

Chart, histogram

Description automatically generated Chart, histogram

Description automatically generated

Chart, histogram

Description automatically generated Chart

Description automatically generated

Chart

Description automatically generated Chart, line chart

Description automatically generated

Chart

Description automatically generated Chart, histogram

Description automatically generated

Chart

Description automatically generated Chart

Description automatically generated

Chart

Description automatically generated Chart, line chart

Description automatically generated

Chart, histogram

Description automatically generated Chart, histogram

Description automatically generated

Chart, histogram

Description automatically generated Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated Chart

Description automatically generated

Chart

Description automatically generated Chart, histogram

Description automatically generated

Chart

Description automatically generated Chart

Description automatically generated

Chart

Description automatically generated Chart

Description automatically generated

Chart, histogram

Description automatically generated Chart

Description automatically generated

Chart, histogram

Description automatically generated Chart

Description automatically generated

**Appendix C: Graphical diagnostics for all variables – comparing cumulative hazard rates**

Chart

Description automatically generated Chart

Description automatically generated

Chart

Description automatically generated Chart

Description automatically generated

Chart

Description automatically generated Chart

Description automatically generated

Chart

Description automatically generated Chart

Description automatically generated

Chart

Description automatically generated Chart

Description automatically generated

Chart

Description automatically generated Chart

Description automatically generated

Chart

Description automatically generated Chart

Description automatically generated

Chart

Description automatically generated Chart

Description automatically generated

Chart

Description automatically generated Chart

Description automatically generated

Chart

Description automatically generated Chart

Description automatically generated

Chart

Description automatically generated Chart

Description automatically generated

Chart

Description automatically generated Chart

Description automatically generated

Chart

Description automatically generated Chart

Description automatically generated

Chart

Description automatically generated Chart

Description automatically generated

**Appendix D: Transition rate estimates split by vaccination status**

Chart, line chart

Description automatically generated Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated Chart, line chart

Description automatically generated

**Appendix E: Transition rate estimates across singular ages**

Chart, histogram

Description automatically generated Chart, histogram

Description automatically generated Chart, line chart

Description automatically generated Chart, line chart, histogram

Description automatically generated

Chart, line chart

Description automatically generated Chart, histogram

Description automatically generated