Assignment: Mortality risk factors and vaccine effectiveness for hospitalised COVID-19 patients
ACTL3141/ACTL5104, T1 2022

Due time: Tuesday 19th April 2022 11.55 am (sharp)

1 Skills developed

This assignment provides you with an opportunity to apply some of the techniques you have learned in the course lectures to a business task involving some COVID-19 data. In addition, the assignment aims to develop the course learning outcomes associated with the course aim "Understand and discuss ethical issues and implications of the modelling introduced in the course," as well as all the UNSW Business School Program Goals but in particular "5. Responsible business practice" and "3. Business communication."

2 Background

In January 2020, the World Health Organisation (WHO) declared the novel coronavirus outbreak a public health emergency of international concern, WHO's highest level of alarm.[1]

At that time there were 98 cases and no deaths in 18 countries outside China. Two years later, there have been 458,479,635 confirmed cases of COVID-19, including 6,047,653 deaths, reported to WHO.[2] Fortunately, efforts to produce a COVID-19 vaccine were unprecedented and the WHO issued the first emergency use validation for a vaccine at the end of 2021.[3] As of today, more than 60% of the world population has received at least one dose of COVID-19 vaccine.[4]

Moving forward, it is essential to study which factors influence the risk of death from COVID-19 and what is the effectiveness of vaccination against death and other severe health outcomes. This assignment aims to identify the risk factors for COVID-19 mortality and survival among hospitalised patients, assess the effectiveness of COVID-19 vaccines, and discuss the ethical implications of COVID-19 vaccination requirements.

3 Task

You work as an actuary for an international insurance company which has some businesses in Brazil. Your role is to assess the mortality and health risk of the Brazilian private health insurance portfolio.

Your boss has just emailed you a set of new Brazilian COVID-19 data along with a recent study by Hojo de Souza et al. (2021) about mortality and survival of hospitalised COVID-19 patients in Brazil. The study uses older data covering the period from February 26th to August 10th, 2020, but you have now access to

[1] Timeline: WHO's COVID-19 response link.

[2] As 16 March 2022, WHO Coronavirus (COVID-19) Dashboard link.

[3] https://www.who.int/news/item/31-12-2020-who-issues-its-first-emergency-use-validation-for-a-covid-19-vaccine-and-emphasizes-need-for-equitable-global-access

[4] https://ourworldindata.org/covid-vaccinations

data covering 1st January 2021 to 31st December 2021. Interestingly, this new dataset has information on the vaccination status of the patients, which was previously unavailable. To inform decision making your boss has asked you to analyse the new set of data. In particular, your boss has asked you to do the following tasks and write a short report to comment and summarise on your results:

1. Perform a brief descriptive analysis of the profile of hospitalised COVID-19 patients.
2. Using the methods learned in the course (e.g. KM, Cox regression), analyse the mortality and survival of hospitalised COVID-19 patients in the given dataset. In particular, but not limited to, ths study of the impact on COVID-19 mortality of age, gender, vaccination status, comorbidities and symptoms.
3. The time a patient spends in an intensive care unit (ICU) can be very costly. Thus your boss is also interested in understanding the pathway of COVID-19 patients upon admission to hospital. To model this pathway your boss is proposing to use the multistate model below:

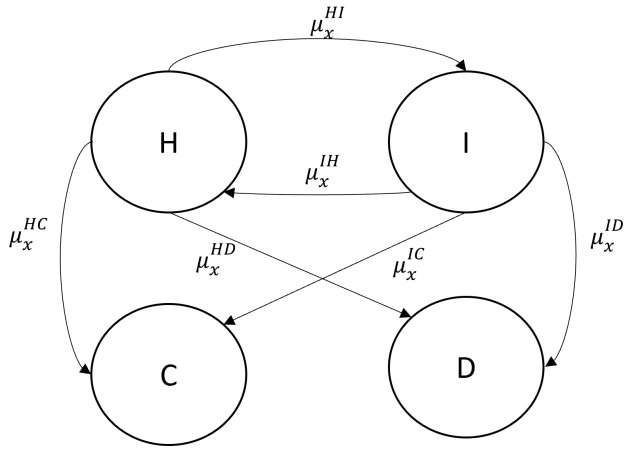

Figure 1: Hospital pathway multistate model. The states are: hospitalised but not in ICU (H), in ICU (I), death (D), and cured and discharged from hospital (C).

   Provide estimates of the age-specifc transition rates associated with the multistate model in Figure 1.

4. Your company is considering the use of COVID-19 vaccination status as a potential pricing and underwriting factor for their private health insurance business. Discuss from an ethical perspective whether the insurer should use COVID-19 vaccination status as a rating or underwriting criteria for setting their health insurance prices.

See more details on the tasks below.

4 Additional information and mark allocation

4.1 Data

For the assignment you have access to a sample of hospitalised COVID-19 patients from the Brazilian Ministry of Health Database (SIVEP-Gripe)[5] in Brazil. You have access to a pre-processed file "Covid-HospDataBrasil.csv." This file contains data on 187209 individuals who were admitted to hospital due to COVID-19 between 1st January 2021 and 31st December 2021. Among other things, for each individual

[5]Data accessed on March 14th 2022 link

the dataset contains information on their clinical symptoms (fever, cough, sore throat, etc) and pre-existing comorbidities (cardiovascular disease, asthma, diabetes, etc). The data is described below:

Variable	Description
`dateBirth`:	Date of birth of the patient.
`age`:	Age nearest birthday at the date when the patient was admitted to hospital.
`sex`:	Sex of the individual taking the values `F` and `M` for females and males, respectively.
`vaccine`:	`TRUE` if the patient had received at least one dose of COVID-19 vaccine or `FALSE` otherwise.
`dateHosp`:	Date in which the patient was hospitalised.
`dateEndObs`:	Date of the end of observation. This can be either the date of death or the date the patient was discharged from hospital.
`covidDeath`:	`TRUE` if the patient died from COVID-19 or `FALSE` if the patient recoverd from COVID and was discharged.
`icu`:	`TRUE` if the patient was admitted to an intensive care unit (ICU) or `FALSE` otherwise.
`dateAdmIcu`:	Date in which the patient was admitted to ICU. It can be greater or equal to `dateHosp`.
`dateDisIcu`:	Date in which the patient was discharged from ICU. It can be smaller or equal to `dateEndObs`.
`fever`:	`TRUE` if the patient had fever or `FALSE` otherwise.
`cough`:	`TRUE` if the patient had a cough or `FALSE` otherwise.
`sorethroat`:	`TRUE` if the patient had a sore throat or `FALSE` otherwise.
`dyspnoea`:	`TRUE` if the patient had dyspnoea or `FALSE` otherwise.
`respdistress`:	`TRUE` if the patient developed acute respiratory distress syndrome (ARDS) or `FALSE` otherwise.
`oxygensat`:	`TRUE` if the patient had blood oxygen saturation $< 95\%$ or `FALSE` otherwise.
`diarrhea`:	`TRUE` if the patient had diarrhea or `FALSE` otherwise.
`vomit`:	`TRUE` if the patient developed vomiting symptoms or `FALSE` otherwise.
`cardio`:	`TRUE` if the patient had a cardiovascular disease or `FALSE` otherwise.
`hematologic`:	`TRUE` if the patient had any hematologic diseases or `FALSE` otherwise.
`downsyn`:	`TRUE` if the patient had Down's syndrome or `FALSE` otherwise.
`hepatic`:	`TRUE` if the patient had any liver diseases or `FALSE` otherwise.
`asthma`:	`TRUE` if the patient had asthma or `FALSE` otherwise.
`diabetes`:	`TRUE` if the patient had diabetes or `FALSE` otherwise.
`neurological`:	`TRUE` if the patient had any neurological diseases or `FALSE` otherwise.
`pneumopathy`:	`TRUE` if the patient had a pneumopathy or `FALSE` otherwise.
`immuno`:	`TRUE` if the patient had any immunodeficiencies or `FALSE` otherwise.
`renal`:	`TRUE` if the patient had any renal diseases or `FALSE` otherwise.
`obesity`:	`TRUE` if the patient had obesity or `FALSE` otherwise.

4.2 Analysis, Modelling and Discussion [85 Marks]

Mark allocation for the assignment can be found in the rubric attached, refer as well to the below for more details on the tasks.

Note that there is a significant amount of research on COVID-19 mortality so you may also wish to engage in extra research beyond your analysis – please feel free to do so. Although the marks for each component of the assignment are capped, extra-research will be encouraged and will potentially offset issues if present. Note however that it is possible to attain full marks without significant extra research.

4.2.1 Descriptive analysis of the profile of hospitalised COVID-19 patients. [10 Marks]

For this part you should use several summary metrics of the characteristics of hospitalised patients, e.g., including the age profile of the patients, the common symptoms and comorbidities they present, and others metrics of your choice.

Your calculations of the different metrics should be accompanied by a discussion of the insights you get from the summary metrics.

4.2.2 Survival analysis [30 Marks]

Your survival analysis should be accompanied by a discussion of the insights you get from each task. Provide the results and analysis associated with each of the estimation and modelling tasks in the technical appendix. Note that, when applicable, you should also provide in the main report the results and analysis using your selected modelling techniques, along with justification of why a particular modelling technique was chosen.

4.2.3 Age-specific transition rates for hospital pathway of patients[20 Marks]

For this part you should compute the age-specific transition rates for the multistate hospital pathway model in Figure 1. The states in the multistate model are:

- H: Hospitalised but not in ICU
- I: Hospitalised and in ICU
- D: Death
- C: Cured and discharged from hospital

Thus, you should provide estimates for the transition rates μ_x^{HI}, μ_x^{HD}, μ_x^{HC}, μ_x^{IH}, μ_x^{ID}, and μ_x^{IC}, where x is the age of the patient.

It is important to note that patients may enter hospital directly to ICU and may also be discharged and sent home directly from ICU.

Provide in the main report graphs with your estimates of the transition rates, along with justification of any assumption you make for the calculation, and an analysis of the results. Note that you are not required to graduate the transition rates and crude (unsmoothed) estimates are enough.

4.2.4 Ethical implications of using COVID-19 vaccination status as a rating criteria [25 Marks]

In this part you should discuss from an ethical perspective whether the insurer should use COVID-19 vaccination status as a rating or underwriting criteria for their private health insurance policies. This discussion should:

1. provide pros and cons of using COVID-19 vaccination status as an underwriting and rating criterion and
2. formulate a recommendation based on the above.

To help complete this part of the assignment you should do the activities of week 8 which focus on ethical perspectives in actuarial work.

In this task, you can consider the implication of other covariates than the ones provided in the data set to make your claim. You can also assume that the health insurance provider is operating in a regulatory environment which does not have any regulatory restrictions on the choice of rating factors for private health insurance.

4.3 Presentation Format and Communcation [15 Marks]

Communication of quantitative results in a concise and easy-to-read manner is a skill that is vital in practice. As such, marks will be given for the presentation of your results. In order to maximize your marks for presentation you may wish to consider issues such as: table size/readability, figure axis/formatting, ease of reading, grammar/spelling, and report structure. You may also wish to consider the use of executive summaries and appendixes, where appropriate. Provide sufficient details to the reader so that they can judge what you are doing, using appendices for non-essential but useful results for the report as necessary.

Note that sufficient detail must be provided (in either the report body and/or appendices) so that the reviewer can follow all the steps and derivations required in your work.

Note that a **maximum page limit of 6 pages** (including tables and graphs but excluding references) is applicable to the main body of the report.[6] You should also consider the rubric for the presentation component. There is no limit to the size of the appendix. Furthermore your answer should satisfy the following formatting requirements: (i) font: Times, 12 pt or equivalent size and (ii) margins: all four of at least 2cm.

4.4 Software

You may choose which software packages to use (e.g. `R`, `Excel` or other), however, most functions you will be required to use for this task are available in `R`. Note also that most of the code enabling you to perform the calculation and analysis are in the `R` tutorials.

4.5 Assignment submission procedure

4.5.1 Turnitin submission

Your assignment report must be uploaded as a **unique document**. As long as the due date is still future, you can resubmit your work; the previous version of your assignment will be replaced by the new version.

Assignments must be submitted via the Turnitin submission box that is available on the course Moodle website. Turnitin reports on any similarities between their own cohort's assignments, and also with regard to other sources (such as the internet or all assignments submitted all around the world via Turnitin). More information is available at: [click]. Please read this page, as we will assume that you are familiar with its content.

Please **also attach any programming code and/or sample spreadsheet output** used in your analysis as a separate file in the dedicated "code_sample" Moodle assignment box on the course webpage. These will be referred to by the marker only if needed, and in particular the main assignment (with appendix) should be self contained.

4.5.2 Late submission

Please note that it is School policy that late submission of assignments will incur in a penalty.

When an assessment item had to be submitted by a pre-specified submission date and time and was submitted late, the School of Risk and Actuarial Studies will apply the following policy. Late submission will incur a penalty of 5% per day or part thereof (including weekends) from the due date and time. An assessment will not be accepted after 5 days (120 hours) of the original deadline unless special consideration has been approved. An assignment is considered late if the requested format, such as hard copy or electronic copy, has not been submitted on time or where the 'wrong' assignment has been submitted. Students who are late must submit their assignment to the LIC via e-mail. The LIC will then upload documents to the relevant

[6] Please kindly note that this is a maximum - you should feel free to use less pages if it is sufficient!

submission boxes. The date and time of reception of the e-mail determines the submission time for the purposes of calculating the penalty.

You need to check your document once it is submitted (check it on-screen). **We will not mark assignments that cannot be read on screen.**

Students are reminded of the risk that technical issues may delay or even prevent their submission (such as internet connection and/or computer breakdowns). Students should then consider either submitting their assignment from the university computer rooms or **allow enough time (at least 24 hours is recommended) between their submission and the due time.** The Turnitin module will not let you submit a late report. **No paper copy will be either accepted or graded.**

4.5.3 Plagiarism awareness

Students are reminded that the work they submit must be their own. While we have no problem with students working together on the assignment problems, the material students submit for assessment must be their own.

Students should make sure they understand what plagiarism is—cases of plagiarism have a very high probability of being discovered. For issues of collective work, having different persons marking the assignment does not decrease this probability.

References

Hojo de Souza, Fernanda Sumika, Natália Satchiko Hojo-Souza, Ben Dêivide de Oliveira Batista, Cristiano Maciel da Silva, and Daniel Ludovico Guidoni. 2021. "On the analysis of mortality risk factors for hospitalized COVID-19 patients: A data-driven study using the major Brazilian database." *PLoS ONE* 16 (3 March). https://doi.org/10.1371/journal.pone.0248580.