Anesthesia technique and mortality in patients with COVID-19 and surgery for hip fracture

Consulting report prepared for Dr. Xue Chen (Janny) Ke

Shuyi Tan and Seren Lee

October 2021

Abstract

The study proposed by Dr. Ke aims to investigate the effect the various anesthetic techniques have on the mortality rate in patients with COVID-19 and undergoing surgery for hip fracture. This report intends to provide statistical advice for this study, which includes guidelines for both exploratory data analysis and confirmatory analysis. Calculation of descriptive statistics and explicit visualization will be employed to approach preliminary results, and we will also propose feasible models including logistic models and a Cox PH model to validate the findings achieve more solid conclusions, along with their interpretations.

1 Introduction

Hip fracture is a serious injury, which can be life-threatening. During the pandemic, it is stated in the literature that patients with both hip fracture and Covid-19 have a higher death rate than those who are with only hip fracture. [Kumar et al., 2021] Patients who choose to have the hip surgery can be operated with one of the following anesthetic techniques: spinal anesthesia (SA) and general anesthesia (GA). The main objective of this study is to investigate the effect of two anesthetic techniques on patients' mortality rate, the effect of COVID-19 adjusted by anesthesia techniques, and the potential impact of the pandemic. To address the research objective, we would like to convert them into the following statistical questions along with some subsidiary inquiries:

1. Among patients with COVID-19 who have undergone surgery for hip fracture within 30 days, is spinal anesthesia (SA) associated with a lower rate of mortality postoperatively? This is our primary outcome variable.

2. Among patients with COVID-19 who have undergone surgery for hip fracture within 30 days, is spinal anesthesia (SA) associated with any changes in terms of pneumonia, mechanical ventilation, ICU admission, venous thromboembolism (VTE) (i.e., clots), myocardial infarction (MI) (i.e., heart attack), stroke, delirium, bleeding requiring transfusion or secondary procedure, readmission, discharge destination (home vs. not home), and postoperative length of stay? These are our secondary outcome variables.

3. Are COVID-19 adjusted by anesthesia technique and other confounders associated with our primary and secondary outcomes?

4. Is the pandemic associated with the primary and secondary outcomes above of patients undergoing surgery for hip fracture but without COVID-19?

5. What statistical measure can be taken to resolve the issue of data missingness in explanatory variables and outcome variables respectively?

6. Due to a large number of potential confounders, how to detect the existence of unmeasured confounders and identify their effect?

2 Data Collection and Description

Usable data for this study come from the NSQIP dataset, which is a multicentre dataset containing specific de-identified perioperative data for patients over 18 years old. Based on the data description file provided, we will mainly make use of the following columns for the proposed analysis, which do not necessarily cover all the columns that are potentially suitable for this analysis:

- **DOpertoD**: This numeric variable represents the days from operation to death, with a maximum of 30. Otherwise, the value of DOpertoD of alive patients will be -99.
- **Prep COVID-19 Diagnosis** and **New Postop COVID-10 Diagnosis**: These two variables will be used to derive a binary predictor of COVID-19 status.
- **OperYR**: It represents the year of surgery, and will be used to demonstrate temporal trends.
- **ANESTHES** and **ANESTHES_OTHER**: It is a factor variable that is to be used for deriving the predictor as anesthesia technique.

2.1 Data Missingness

Before conducting any statistical tests or building models, we recommend examining the fraction and mechanism of missing values. Calculating the fraction of missing values for every variable involved in the analysis will help to obtain a clear picture of the overall data missingness and serve as a guide to inform decisions about whether to perform imputation or not. In addition, determining the mechanism of data missingness using your professional knowledge in clinical data collection may also be needed. If the data are missing completely at random, we may remove records with the missing values, which is the complete cases method. Missing completely at random (MCAR) means that there is no relationship between the missingness of the data and any values, observed or missing. The missing data points are a random subset of the data. [Nguyen, 2021] However, in most cases, this is unlikely to be a reasonable assumption.

A weaker assumption would be missing art random (MAR). MAR means that there is a systematic relationship between the propensity of missing values and the observed data, but not the missing data. MAR requires that the cause of data missingness is unrelated to the missing values but may be related to the observed values of other variables. [Nguyen, 2021] Technically, there is no solid statistical method to confirm the mechanism of data missingness because we are not able to derive unobserved values. Instead, related knowledge in clinical data collection would be required to determine whether the data could be assumed as MAR. If the assumption approximately stands, it may be feasible to implement multiple imputation to obtain a "full" dataset. Multiple imputation is a general method to handle data missing at random, which aims to allow for the uncertainty about the missing data by creating several different plausible imputed data sets and appropriately combining results obtained from each of them. [Sterne et al., 2009] As a reminder, we should avoid imputing the missing values in the response variable. Generally speaking, we recommend doing preliminary checks to see if there exist some specific patterns among patients with unobserved values. Then starting from a simple complete-case analysis, if the results appear to have significant improvement from more complex methods (like imputation), then it might be worth exploring more advanced methods.

3 Exploratory Data Analysis

Suggested exploratory data analyses are summarized into descriptive statistics, visualization, and test for simple comparison. Calculating descriptive statistics brings up the basic idea of how the data look like. For categorical variables, they could be relative frequency and count. For continuous variables, the mean, standard deviation and five-number summary would work well.

Visualization would also be a strong tool to acquire insights. For example, we can visualize the time trend using a time series graph depicting the rate versus the month. There were multiple waves of COVID-19 during with positive cases increased steeply. [Andrew Foote, 2021] If the mortality rate also increased during the wave, it would be appropriate to investigate the association between the mortality rate of hip fracture surgery patients and COVID-19. Grouped boxplots are also suitable for comparing the distribution of continuous variables among different groups in the categorical variables. For example, we could make boxplots for *DOpertoD* and *DOptoDis* versus *ANESTHES* or *COVID-19* positive/negative.

In addition, we recommend comparing the mortality rates among groups such as General Anesthesia or Spinal Anesthesia using ANOVA / Two-sample t-test to explore if any significant differences exist. These tests are reliable when the data within each group are approximately normally distributed. But even when the data is not normally distributed, results from the test are robust due to the central limit theorem, which gives approximate normality when the sample size is large enough. Nevertheless, if the sample size is small and the normality of data is in doubt, non-parametric tests such as the Wilcoxon rank-sum test or Kruskal Wallis test would be good alternatives.

Lastly, since there are a great number of binary responses, we suggest using the chi-square test or Fisher's exact test to find some association between binary variables. If the result indicates that binary response and binary predictor are not independent, it would be reasonable to include that binary predictor in the model. For example, after conducting a chi-square test on mortality and discharge destination, if we obtain a small p-value, we infer that mortality may be partially explained by this covariate.

4 Confirmatory Data Analysis

The most intuitive model to address the research objective is the logistic regression model. Logistic regression is appropriate to use when the response is binary. Another alternative method to explore the effect of anesthetic techniques on mortality is the survival analysis.

4.1 Data Processing

Prior to analyzing the data, it is crucial to convert the original data to proper formats. We may derive the primary outcome variable *mortality* from *DOpertoD* based on determined criteria.

$$mortality = \begin{cases} 1 & DOpertoD \neq -99 \\ 0 & DOpertoD = -99 \end{cases}$$

For the variable *OperYR*, we can derive a new binary indicator, *yearPandemic*, for year cohorting in order to investigate the effect of pandemic.

$$yearPandemic = \begin{cases} 1 & year = 2020 \\ 0 & 2017 \leq year \leq 2019 \end{cases}$$

As for the core predictor *anesthes*, we may filter the patients who have received certain anesthetic techniques of interest, and transform this character variable into a categorical variable.

4.2 Logistic regression

Aside from the logistic regression model fitted to *mortality*, we suggest fitting separate linear models to different responses. There are three types of recommended models. They answer questions 1 to 4 in the introduction.

1. For questions 1 and 2, we can fit logistic models with *SA* as the explanatory variable and other potential confounders. *SA* is a binary variable with 1 representing SA and 0 otherwise. Please note that we could only use the data of patients who have been tested COVID-19 positive. If the estimated regression coefficient of *SA* is statistically significant negative, we could conclude that SA is associated with a lower probability of the response variable being 1. In other words, we could conclude that SA is related to the lower mortality rate.

2. For question 3, we can fit a logistic model with explanatory variables *SA*, *Preop COVID-19 Diagnosis*, and other potential confounders. In these models, we use all entries include positive and negative COVID-19 patients. Having a binary variable *COVID-19*, which is 0 when COVID-19 negative, if the variable *COVID-19* has a statistically significant positive regression coefficient estimate, it reveals that COVID-19 could increase the mortality rate when other variables such as *SA* and other confounders are controlled to have common values for both COVID-19 positive and negative patients.

For above two models, potential confounders are tested with normalized GVIF, $\text{GVIF}^{\frac{1}{2p}}$, to avoid the multicollinearity issue. [Fox and Monette, 1992] Since the model includes categorical variables, it is not possible to calculate VIF for those variables. Therefore, we suggest using the normalized GVIF which is a generalized version of VIF and normalized for easier comparison. A large normalized GVIF indicates that a variable might trigger a multicollinearity issue. If a variable has a normalized GVIF larger than 5, we might be worried about multicollinearity. And when the normalized GVIF is over 10, that strongly indicates the variable is problematic. In these cases, we recommend removing those confounders from the model by stepwise selection. [Fox, 2015] We could check gvif information by using the R function vif() in car package. [Fox et al., 2021] And, we could perform the stepwise selection by using stepAIC() in MASS package. [Ripley, 2021]

3. For question 4, *yearPandemic* will serve as a binary explanatory variable in the logistic regression model. We only use rows with COVID-19 negative patients. Note that *yearPandemic* is 1 if $year = 2020$ and 0 is $2017 \leq year \leq 2019$ as in 4.1. Consequently, if the variable *yearPandemic* has a statistically significant positive regression coefficient, we would conclude that the Pandemic could increase that mortality rate.

The common formula of logistic regression is

$$\log(\frac{P(y=1)}{1-P(y=1)}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$$

where there are p explanatory variables in the model, and y is one of *mortality* and secondary outcomes. When β_1 is significantly positive, we could conclude that increasing x_1 increases the probability of the response variable y being 1. For example, in model 3, where *yearPandemic* is a binary explanatory variable and *mortality* is a binary response variable, if β_1 is significantly positive, there is sufficient evidence to conclude that a patient with $yearPandemic = 1$ has a bigger chance of having $mortality = 1$.

4.3 Survival analysis

As in the previous section, *DOpertoD* is the number of days to death after the surgery. With this time-to-event variable, another regression model, the Cox proportional hazards (Cox PH) model, can be employed to help us obtain a more comprehensive understanding of the effect of our predictors on mortality. The purpose of this model is to assess simultaneously the effect of several predictors on survival. In our case, the model

can evaluate simultaneously the effect of anesthesia techniques and COVID-19 on mortality. In other words, the COX PH allows us to investigate how specific factors influence the hazard rate of the event (mortality in our case) happening at a particular time point. To brifely what a COX PH model is, the Cox PH model predicts the hazard function, which describes the risk of dying at time t. Denoted the hazard function as $h(t)$:

$$h(t) = h_0(t)exp(\beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p)$$

where t represents the survival time, and the coefficients (β_1,β_2,...,β_p measure the impact of covariates. The Cox model can be seen as a multiple linear regression of the log of the hazard on the predictors x_i, with the baseline hazard h_0 being an 'intercept' term that varies with time. In this study, let's have SA as x_1 with $x_1 = 1$ representing SA and 0 otherwise, then we can interpret that patients receiving GA have a hazard $exp(\beta_1)$ times those who received GA. In short, a non-zero β means that the corresponding predictor affects the hazard function. In R, we recommend using the coxph() function in package survival [Therneau, 2020]. As a reminder, please don't forget to check the assumptions of a Cox PH model prior to interpreting.

4.4 Sample size and power Calculations for logistic regression

We suggest calculating minimum sample size required for the specific power and the power for current sample size using R function wp.logistic() in the package WebPower. [Zhang and Mai, 2021] This function use an Wald test based algorithm introduced in [Demidenko, 2007]. The algorithm and function only work for one binary explanatory variable at a time, and requires $p0$ and $p1$ which are probabilities of y being 1 when that binary explanatory variable is 0 and 1, respectively. If the power is specified, the required sample size is returned, and power is returned when the sample size is specified.

5 Sensitivity analysis

As you may have been aware, a number of assumptions are involved in the methodology of handling data missing and constructing models. In fact, assumptions are rarely true, which is the source of uncertainty. Under the context of either data processing and modelling, it is beneficial to consider the robustness of the adopted methods/models to department from assumption. Sensitivity Analysis (SA) is a method that measures how the impact of uncertainties of one or more input variables can lead to uncertainties on the output variables. [Pichery, 2014] To what extent the conclusions are reliable will depend on the magnitude of departures. Correspondingly, it will provide important information on how much confidence to be given to the results.

5.1 Missing Values

Regarding data missingness, sensitivity analysis evaluates the impacts of different missingness mechanisms on the robustness of statistical results. It is natural to make assumptions of the mechanisms of missing data, but we will not be able to verify these assumptions based on observed data. As mentioned previously, the complete-case analysis is based on the assumption of MCAR. With incomplete predictors, unless data are MAR, the estimated statistics and regression coefficients may be biased because we cannot guarantee that the patients in the complete-case analysis have similar characteristics with those who have been removed due to unobserved data. Besides, the suggested method of multiple imputation is based on the assumption that the data are MAR. Even though MAR is a common reliable assumption, it may not always be realistic for the data at hand. Therefore, we suggest comparing the results from complete-case analysis and those generated from the dataset imputed by multiple imputation. If there is a large difference, we should carefully interpret the result and determine to what extent of faith could be put in the results. As an extension, when having models obtained from different datasets (complete case or imputed), we also suggest performing model evaluation to choose a preferred model based on specific criteria like AIC and BIC.

5.2 Definitions of COVID-19 Positive

Specific COVID-19 positive definition is used throughout the analysis. However, patients might be classified in different ways if different definitions are used. This difference could lead to a discrepancy in results because it is possible the patients in the same group of COVID-19 status, say the positive group, may not be similar across various COVID-19 definitions. If the results of different definitions agree, we may draw generalized and robust conclusions. If not, we recommend choosing a relatively more appropriate definition of COVID-19 to be the predictor based on either literature or your clinical knowledge.

6 Discussion and Conclusion

This study is initiated for investigating the association between Anesthesia technique and mortality rate and prognosis after the hip-fracture surgery for the patients who have both COVID-19 and hip-fracture. Therefore, exploring the relationship between variables and interpreting regression coefficients are to be prioritized in the analysis. For example, by looking at the boxplot, in which each box depicts the distribution of the days between surgery and death, we would be able to get some preliminary ideas on the effect of anesthetic techniques and COVID-19.

In the following confirmatory analysis, the estimates of regression coefficients can help to verify your findings from the EDA. It is always important to keep in mind that we cannot infer that the association between a predictor and the outcome variable is negligible despite no statistical significance. In addition, we do not recommend treating $p = 0.05$ as a solid cutoff for deciding statistical significance but just for reference. As a complementary analysis, the result of the survival analysis suggested in section 4.3 should be interpreted carefully. Unlike a common linear regression model, the Cox PH model compares the survival distribution of time. Also, there is a core assumption of proportional hazard. Therefore, violation of the assumption needs to be checked before interpreting. We recommend using a Kaplan-Meier plot for validating the assumption of proportional hazard.

In the sensitivity analysis, it provides essential assistance for assessing the validity of models and generalization of conclusions, especially the impact of uncertainties. Comparing the models from complete-case analysis and imputed-case analysis will help to identify selection bias and validate the performance of models. The E-value and Durbin-Wu-Hausman test helps to check the effect of unmeasured confounders. Based on the test results and the magnitude of E-value, if there is significant residual confounding, it suggests that more effort should be made to improve the control for confounders. For the sensitivity analysis on the definitions of COVID-19 positive, it is expected to help confirm the difference or similarity among these definitions. If the results do not agree using different definitions of COVID-19 positive, we suggest acknowledging the difference and choosing a "preferred" definition to be the COVID-19 predictor based on actual needs.

7 Further reading

We recommend reading *Sample size determination for logistic regression revisited* [Demidenko, 2007] and *Sample size and optimal design for logistic regression with binary interaction* [Demidenko, 2008] for further information on sample size and power calculation for the logistic regression model.

References

[Andrew Foote, 2021] Andrew Foote, J. L. (2021). 15 months, 3 waves: How covid-19 coursed through the capital.

[Cheng et al., 2014] Cheng, J., Lorch, S. A., Small, D. S., and Guo, Z. (2014). Using an instrumental variable to test for unmeasured confounding. *Statistics in Medicine*, 33(22).

[Demidenko, 2007] Demidenko, E. (2007). Sample size determination for logistic regression revisited. *Statistics in medicine*, 26(18):3385–3397.

[Demidenko, 2008] Demidenko, E. (2008). Sample size and optimal design for logistic regression with binary interaction. *Statistics in medicine*, 27(1):36–46.

[Fox, 2015] Fox, J. (2015). *Applied regression analysis and generalized linear models*. Sage Publications.

[Fox and Monette, 1992] Fox, J. and Monette, G. (1992). Generalized collinearity diagnostics. *Journal of the American Statistical Association*, 87(417):178–183.

[Fox et al., 2021] Fox, J., Weisberg, S., and Price, B. (2021). *car: Companion to Applied Regression*. R package version 3.0-11.

[Kumar et al., 2021] Kumar, P., Jindal, K., Aggarwal, S., Kumar, V., and Rajnish, R. K. (2021). 30-day mortality rate in hip fractures among the elderly with coexistent covid-19 infection: A systematic review. *Indian journal of orthopaedics*, pages 1–11.

[Nguyen, 2021] Nguyen, M. (2021). Assumptions.

[Pichery, 2014] Pichery, C. (2014). Sensitivity analysis. In Wexler, P., editor, *Encyclopedia of Toxicology (Third Edition)*, pages 236–237. Academic Press, Oxford, third edition edition.

[Ripley, 2021] Ripley, B. (2021). *MASS: Support Functions and Datasets for Venables and Ripley's MASS*. R package version 7.3-54.

[Smith and McKenna, 2013] Smith, T. J. and McKenna, C. M. (2013). A comparison of logistic regression pseudo r2 indices. *Multiple Linear Regression Viewpoints*, 39(2):17–26.

[Sterne et al., 2009] Sterne, J. A. C., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., Wood, A. M., and Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*, 338.

[Therneau, 2020] Therneau, T. M. (2020). *A Package for Survival Analysis in R*. R package version 3.1-12.

[VanderWeele and Ding, 2017] VanderWeele, T. J. and Ding, P. (2017). Sensitivity analysis in observational research: introducing the e-value. *Annals of internal medicine*, 167(4):268–274.

[Zhang and Mai, 2021] Zhang, Z. and Mai, Y. (2021). *WebPower: Basic and Advanced Statistical Power Analysis*. R package version 0.6.

Appendix

Unmeasured confounders

In the regression models, potential confounders are included in the model as explanatory variables. Due to the lack of a complete list of necessary confounders, unmeasured confounders may result in biased effect estimates. One of the most intuitive reflections of unmeasured confounders will be R^2. R^2 is a useful metric for multiple linear regression, but does not have the equivalent meaning in logistic regression models. Instead, pseudo R^2 is used in logistic regression models, such as the McFadden's pseudo R^2. These pseudo R^2s have been developed that are intended as logistic regression analogs of R^2 as used in ordinary least-squares regression. [Smith and McKenna, 2013] A poor fit revealed by extremely low value of the pseudo R^2 indicates that a model does not explain the variation of responses well. Consequently, there may exist unmeasured confounders. Alternative methods to detect unmeasured confounders include E-value and the Durbin-Wu-Hausman test. For E-value, it is defined to be "the minimum strength of association, on the risk ratio scale, that an unmeasured confounder would need to have with both the treatment and the outcome to fully explain away a specific treatment-outcome association, conditional on the measured covariates, to explain away a treatment-outcome association." [VanderWeele and Ding, 2017] A large E-value implies big unmeasured confounding would be needed to explain away an effect estimate. Conversely, A small E-value implies little unmeasured confounding would be needed to explain away an effect estimate. [VanderWeele and Ding, 2017] For the Durbin-Wu-Hausman test, it is a popular test for endogeneity in Econometrics. Endogeneity occurs when there is a correlation between the explanatory variables and the error term in the statistical model. In a broad sense, endogeneity can be caused by either omitted variables, or unobserved heterogeneity. In addition, it is found that the Durbin–Wu–Hausman test can have inflated type I error rates when there is treatment effect heterogeneity. [Cheng et al., 2014] Therefore, we recommend using Durbin-Wu-Hausman test as a preliminary exploration, and using E-value as the main measure of unmeasured confounders.

However, it is important to note that unmeasured confounders are inevitable issues even though we could somehow notice their existence. They will always subsist and result in irreducible variance in regression models. Hence, though we might be able to reduce to some extent, we could not get rid of all.