

SRD-related clinical characteristics and outcomes in covid-19 patients



Weill Cornell
Medicine

Group 6 – Project 3:

Menghan Zhang, Ziwen Zhang,

Wentao Li, Jiayue Cheng

Contents

- Background
- Methods
- Results
- Conclusions & Limitations
- References



Background

Introduction of Project 3:

- ❖ Systemic rheumatic diseases (SRD) are chronic, inflammatory autoimmune disorders
- ❖ There are more than 100 kinds of SRD, including osteoarthritis, rheumatoid arthritis, systemic lupus erythematosus, etc.
- ❖ SRD can cause damage to patients' vital organs, such as lungs, heart, kidney or nervous system
- ❖ Sometimes, SRD may result in serious morbidity or mortality if not properly managed or combined with other diseases, such as Covid-19

Study Aim

- ❑ Investigate the relationship between SRD status and Covid-19 symptoms
- ❑ Investigate the association between SRD related variables and hospital length of stay along with evaluating the predictor of largest impact on outcome we can manipulate

Variable Categories

- **Socio-demographics** : age, sex, race, etc.
- **Disease History**: hypertension, diabetes, stroke, etc.
- **Medication**: NSAIDs, Statins, Hydroxychloroquine, Antibiotics, etc.
- **Lab Results**: Peak/Lowest White Blood Cell, Peak/Lowest Lymphocytes, Peak/ Lowest Lymphocytes, etc.
- **Outcome**: Hospital length of stay, etc.

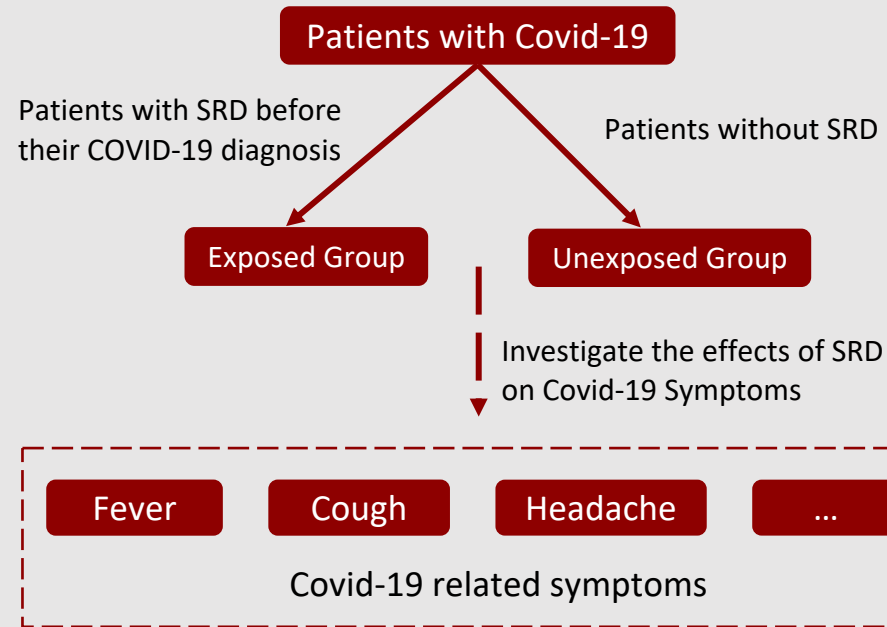


Methods

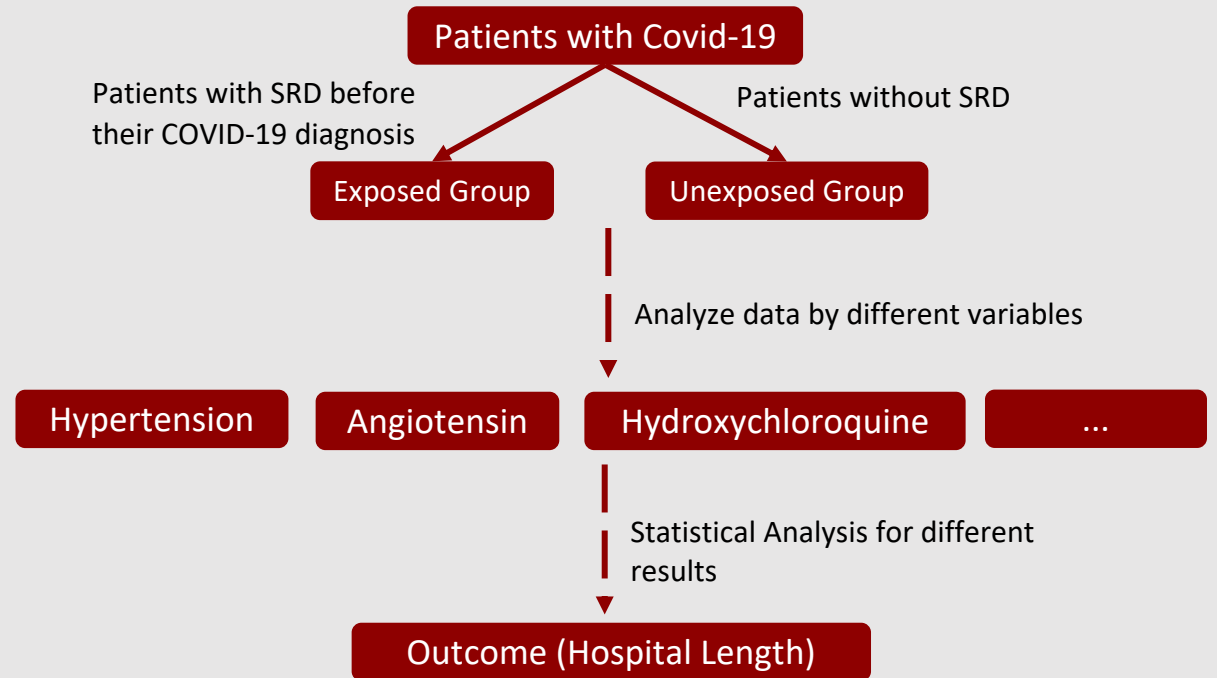
Study Design

Retrospective Observational Cohort Study

Aim 1



Aim 2



Methods

Modeling

| Differences | Multiple Linear Regression | Survival Analysis | Logistic Regression |
|---------------------------|--|--|---|
| Goal | Investigate the linear relationship between several variables and the outcome . | Describe the length of time between a start point (origin) and an endpoint. | Finds an equation to predict the outcome of a categorical variable. |
| Requirements of Variables | <ul style="list-style-type: none">At least 2 independent variablesDependent variable is continuous | <ul style="list-style-type: none">Time-dependent variablesDependent variables: time and the event | <ul style="list-style-type: none">Independent VariablesDependent variable is categorical |
| Output | A range of values | A time until an event occurs | 2 (0 or 1) |

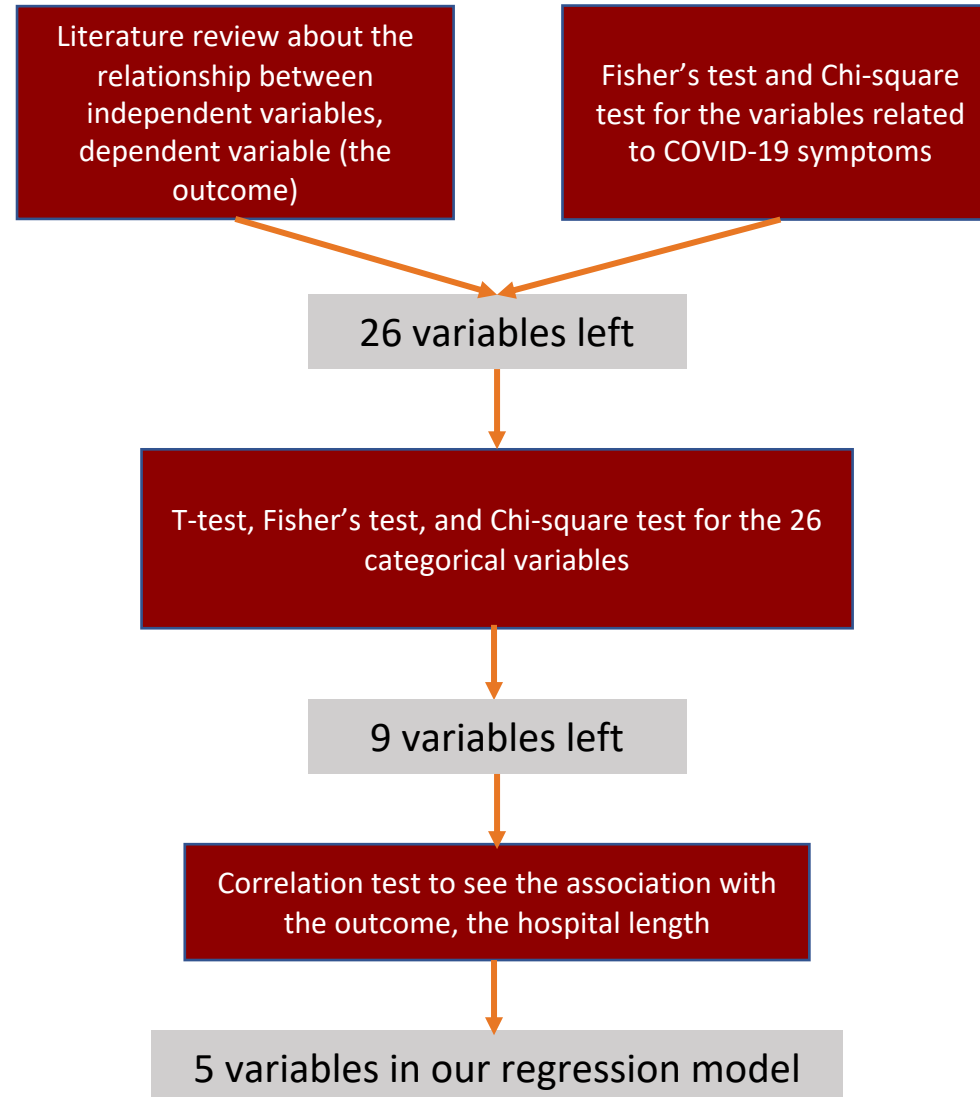
Reasons for choosing MLR for our model

- The dependent variable (length of stay) is continuous, so multiple linear regression is better than logistic regression
- There is no baseline results at the start point, and the variables we investigated are not time-dependent variables, so survival analysis is not suitable



Methods

Data Dimension Reduction



Results Analysis – Aim1

Investigate the relationship between SRD status and Covid-19 symptoms

| Characteristic | No SRD, N = 164 | SRD, N = 92 | p-value |
|-----------------------|-----------------|-------------|---------|
| Fever | 112 (68%) | 61 (66%) | 0.7 |
| Cough | 121 (74%) | 64 (70%) | 0.5 |
| Dyspnea | 106 (65%) | 62 (67%) | 0.7 |
| Sore Throat | 11 (6.7%) | 9 (9.8%) | 0.4 |
| Rhinorrhea | 4 (2.4%) | 6 (6.5%) | 0.2 |
| Headache | 7 (4.3%) | 7 (7.6%) | 0.3 |
| Myalgias | 31 (19%) | 16 (17%) | 0.8 |
| Nausea | 30 (18%) | 18 (20%) | 0.8 |
| Diarrhea | 40 (24%) | 24 (26%) | 0.8 |
| Sputum | 9 (5.5%) | 9 (9.8%) | 0.2 |
| Presyncope | 4 (2.4%) | 4 (4.3%) | 0.5 |
| Chest Pain | 22 (13%) | 15 (16%) | 0.5 |
| Abdominal Pain | 7 (4.3%) | 3 (3.3%) | >0.9 |
| Altered Mental Status | 22 (13%) | 15 (16%) | 0.5 |
| Anosmia | 5 (3.0%) | 1 (1.1%) | 0.4 |
| Ageusia | 3 (1.8%) | 1 (1.1%) | >0.9 |
| Other Symptom | 40 (24%) | 21 (23%) | 0.8 |

Results Analysis:

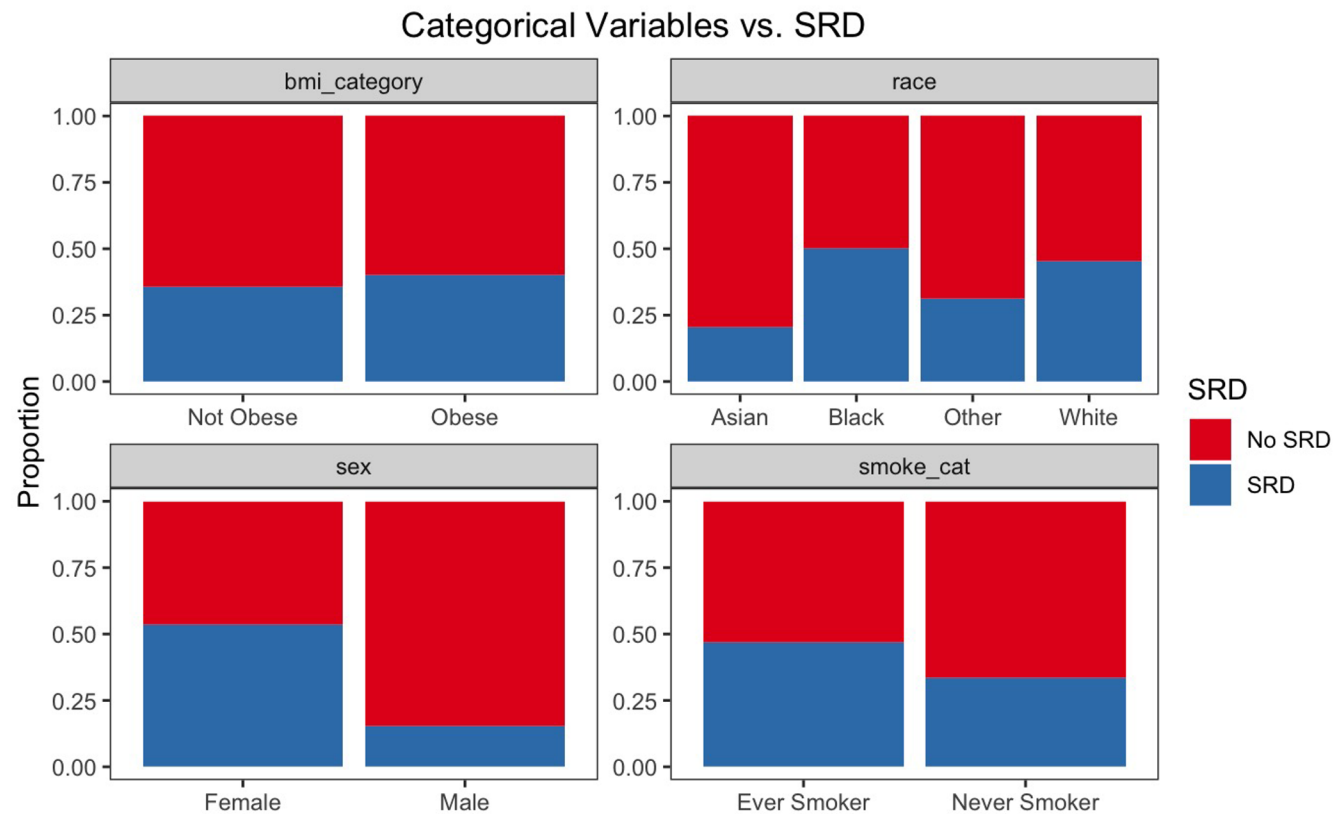
- A table shows the number and percentage of different Covid-19 symptoms (20) between two groups along with the associated p.values
- H0: no relationship; H1: relationship
- Categorical variables: Chi-squared test and fisher test
- P.values are too big, eg. Fever
- Conclusion: no relationship
- Exclude the Covid-19 symptom variables from our future statistical analysis



Weill Cornell
Medicine

Results Analysis – Aim2

Descriptive Analysis



Results Analysis:

- No obvious differences regarding SRD patients' distribution in different obesity status groups
- Regarding different races, the proportion of SRD varies
- Regarding gender, the proportion of SRD patients are the most obvious
- Regarding smoking history, the proportion of SRD patients in smoker groups is higher than the percentage in the non-smoker group



Results Analysis

Investigate the relationships between variables and SRD

| Characteristic | No SRD, N = 164 | SRD, N = 92 | p-value |
|---------------------------|----------------------|----------------------|------------------|
| Age | 65 (49, 78) | 67 (59, 80) | 0.4 |
| Sex | | | <0.001 |
| Female | 73 (45%) | 75 (82%) | |
| Male | 91 (55%) | 17 (18%) | |
| BMI | 28 (24, 32) | 27 (24, 32) | 0.8 |
| Smoking | | | 0.054 |
| Ever Smoker | 34 (21%) | 29 (32%) | |
| Never Smoker | 130 (79%) | 63 (68%) | |
| CAD | 23 (14%) | 20 (22%) | 0.11 |
| Diabetes | 57 (35%) | 27 (29%) | 0.4 |
| Stroke | 11 (6.7%) | 6 (6.5%) | >0.9 |
| Renal Disease | 14 (8.5%) | 10 (11%) | 0.5 |
| Active Cancer | 4 (2.4%) | 2 (2.2%) | >0.9 |
| Pulmonary | 31 (19%) | 28 (30%) | 0.036 |
| Hypertension | 97 (59%) | 66 (72%) | 0.044 |
| Transplant | 6 (3.7%) | 4 (4.3%) | 0.7 |
| Pregnancy | 3 (1.8%) | 0 (0%) | 0.6 |
| Angiotensin | 35 (21%) | 33 (36%) | 0.012 |
| Statin | 52 (32%) | 34 (37%) | 0.4 |
| Hydroxychloroquine | 3 (1.8%) | 21 (23%) | <0.001 |
| Antibiotics | 105 (64%) | 57 (62%) | 0.7 |
| Steroids | 26 (16%) | 47 (51%) | <0.001 |
| IVIG | 0 (0%) | 1 (1.1%) | 0.4 |
| Tocilizumab | 7 (4.3%) | 0 (0%) | 0.052 |
| Sarilumab | 0 (0%) | 1 (1.1%) | 0.4 |
| Peak.White.Blood.Cell | 10 (7, 17) | 11 (7, 17) | 0.8 |
| Peak.Hemoglobin | 14.40 (12.90, 15.50) | 12.95 (12.12, 15.30) | 0.019 |
| Peak.Lymphocytes | 2 (1, 8) | 2 (1, 3) | 0.015 |
| Lowest.Hemoglobin | 11.60 (8.95, 13.05) | 10.25 (7.93, 11.40) | <0.001 |
| Lowest.Lymphocytes | 0.71 (0.50, 1.05) | 0.70 (0.42, 1.00) | 0.2 |
| Lowest.White.Blood.Cell | 5.36 (4.20, 6.90) | 5.20 (3.70, 7.60) | >0.9 |

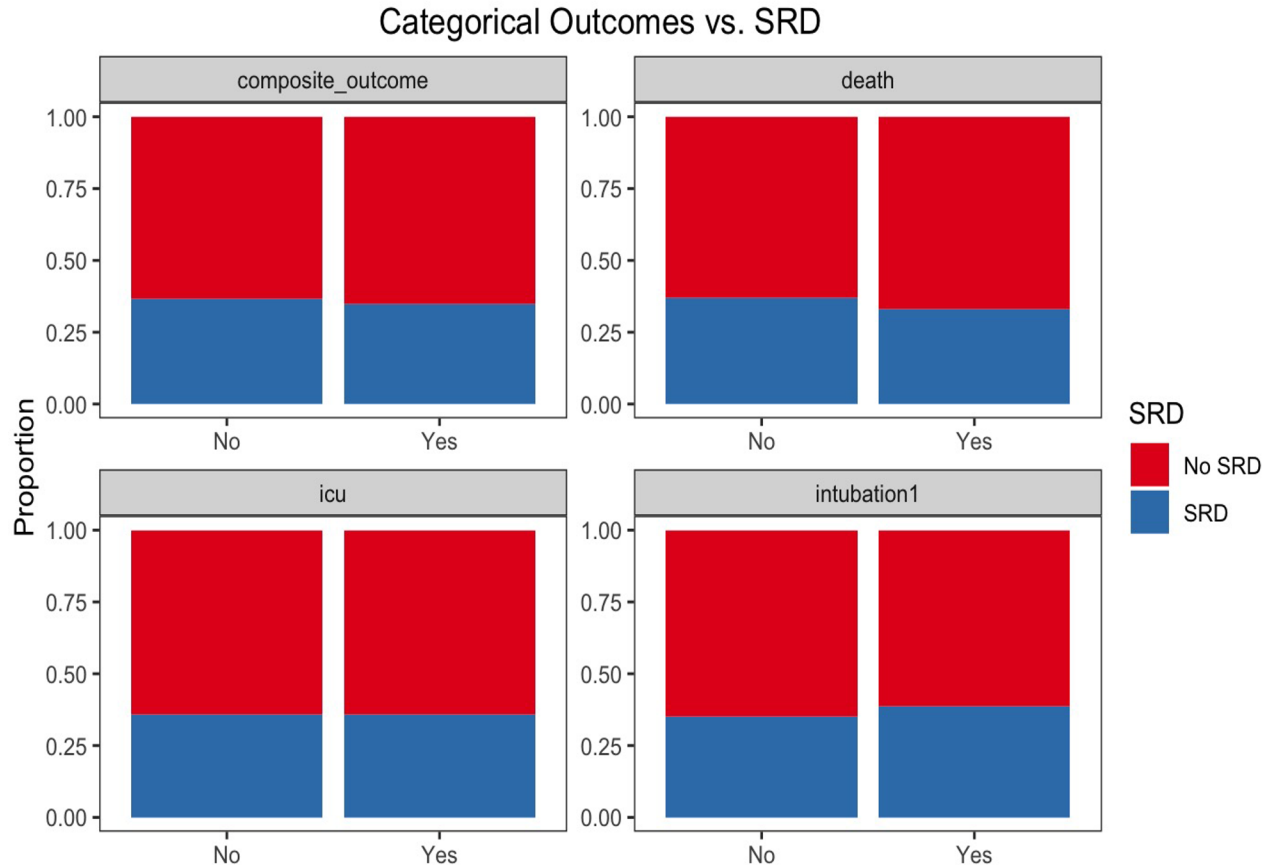
Results Analysis:

- There are 26 variables left
- To investigate whether there is any differences in variables between the two groups
- A table shows median and IQR values, count and percentage, and corresponding p.values
- A mix of numeric variables and categorical variables
- Set significance level at 0.05
- There are 9 pairs of significantly different values, eg. steroids
- Filter out 9 variables that have relationships with SRD status
- Confirm 9 variables for further analysis



Results Analysis

Investigate the relationship between SRD and clinical outcomes



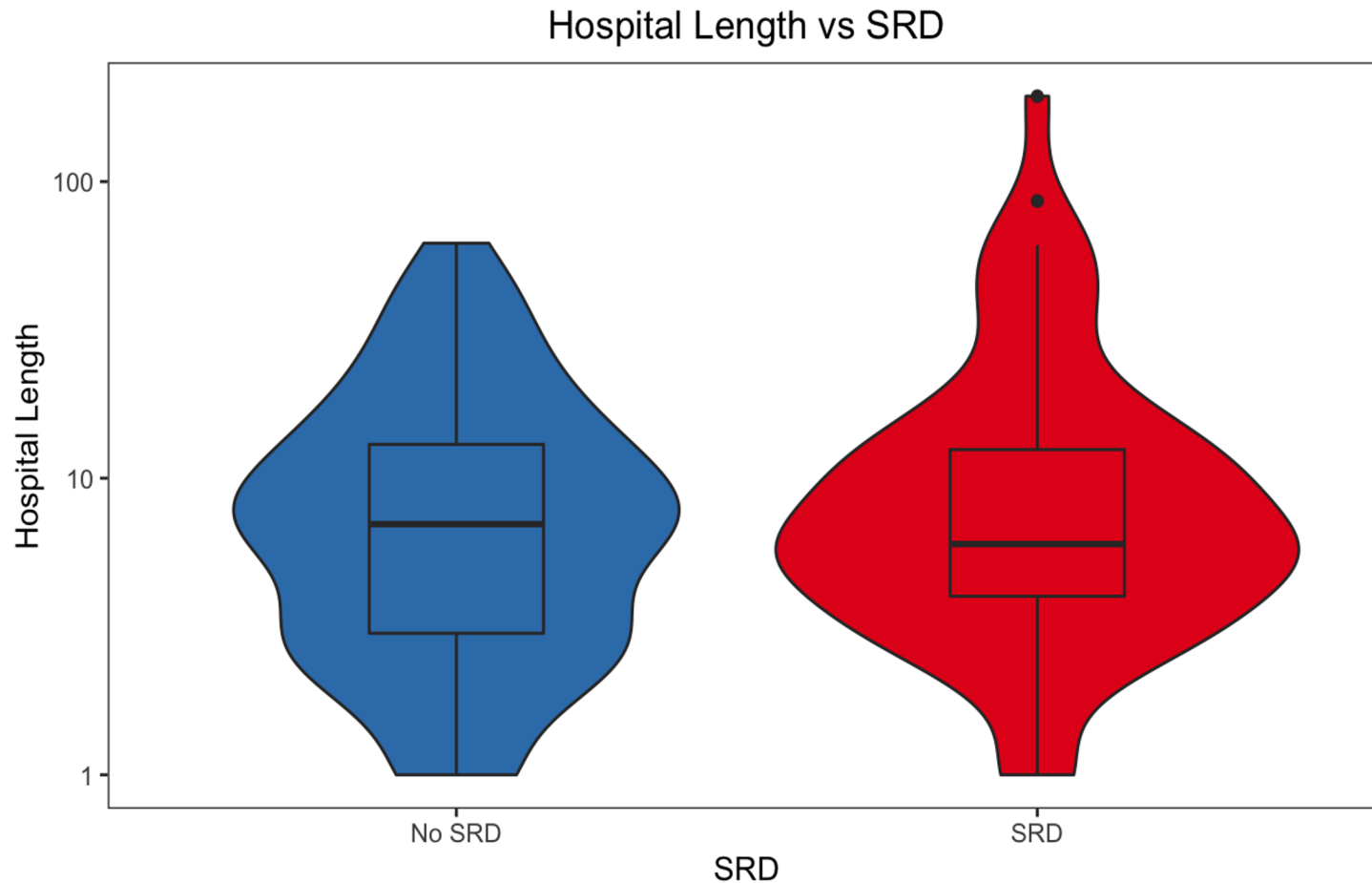
Reasons for choosing “Hospital Length of Stay” as our outcome

- 6 outcomes in our dataset, 4 for categorical variables and 2 for continuous variables
- The reasons for excluding the outcomes:
 - 1) No obvious differences in proportion between the SRD status of each categorical outcome
 - 2) We are not interested in exploring the outcome of death, because:
 - ☐ The sample size of death is very small and can not provide strong evidence
 - ☐ The causes of death are diverse and complex, so it is hard to remove other confounding factors.
 - ☐ Our goal is to investigate the most useful therapy to treat patients and how to rehabilitate them. Thus, we will focus on hospital length of stay which is the direct outcome rather than death



Results Analysis

Investigate the relationship between SRD and the hospital length



Results Analysis:

- Hospital length is a key variable to measure the treatment effects of Covid-19
- Based on our violin plot along with box plot, the median of hospital length in SRD group is smaller than the No-SRD group
- The range of hospital length in SRD group is narrower
- The distribution is skewed, outliers



Results Analysis – Univariate Analysis

Investigate the relationships between SRD-related variables and the length of stay

| Variable | Corr | P.values |
|--------------------|-------|----------|
| Sex | -0.13 | 0.0710 |
| Pulmonary | 0.04 | 0.5592 |
| Hypertension | 0.06 | 0.4200 |
| Angiotensin | 0.04 | 0.5637 |
| Hydroxychloroquine | 0.02 | 0.7950 |
| Steroids | 0.29 | 0.0001 |
| Peak Lymphocytes | 0.06 | 0.4456 |
| Peak Hemoglobin | 0.24 | 0.0014 |
| Lowest Hemoglobin | -0.56 | 0.0000 |

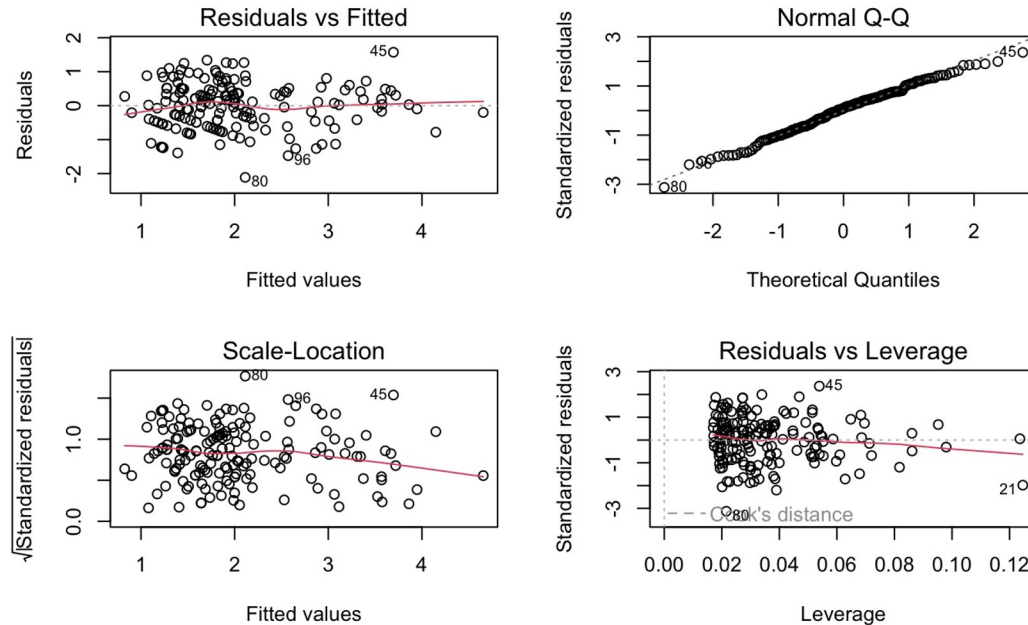
Results Analysis:

- Based on previous variable screening, there are 9 SRD-related variables left
- Aim to filter again to find final variables to fit in our model
- A table showing Pearson correlation coefficients and p.values between these variables and our outcome variable, hospital length
- Set the significance level at 0.2
- According to p.values, find 4 potential predictors for our model



Results Analysis – Multivariable Analysis

Statistical Analysis



Results Analysis

- With the increase of Peak Hemoglobin level ($\beta = 0.089$, $p = 0.005$) and Lowest Hemoglobin level ($\beta = 0.27$, $p < 0.0001$), the length of hospital stay will increase and decrease respectively, after the adjustment of sex
- We can conclude that sex is the most important predictor, and this implies that the length of hospital stay is longer among the male patients than the female patients after adjustment for the other predictors

- Firstly, we found that the residuals versus fitted plot did not show a fairly random pattern after conducting residual analysis – the scatter points showed a fan shape (heteroskedasticity).
- Therefore, we transformed the Hospital length to $\log(\text{Hospital Length})$, which fit our multivariable linear regression (MLR) well:

$$\log(\widehat{\text{Hospital LOS}}) = -0.369 * \text{sex} + 0.089 * \text{Peak Hemoglobin} - 0.27 * \text{Lowest Hemoglobin} + 3.874$$



Conclusions & Limitations:

Conclusion:

- ❑ The best predictor of the outcome (hospital length) is sex, and males have higher risk of severe outcome, which means longer hospital length
- ❑ No evidence to believe that SRD has a correlation with the hospital length based on our analysis although we hypothesized that SRD may have effects on the outcome based on some literature review
- ❑ There is no significant Racial difference between SRD and non-SRD groups

Limitations:

- Small sample size
- Length of stay is a numeric variable with a right-skewed distribution



References

- [1] Tiffany Y., Kristin M. D., Naomi J. P., Xiaoqing F., et al. Incident systemic rheumatic disease following COVID-19. *The Lancet Rheumatology*. 2021 June 01; 3(6) : E402 – E404.

- [2] Siegel C. H., Choi J., D'Angelo D., et al. Outcomes of COVID-19 and Factors Associated With Its Severity Among Hospitalized Patients With and Without Systemic Rheumatic Disease During the First Wave of the Pandemic in New York City. *Journal of Clinical Rheumatology*. 2022 July 29.

- [3] Bournia, V. K., Fragoulis, G. E., Mitrou, P., et al. Different Covid-19 Outcomes Among Systemic Rheumatic Diseases: A Nation-wide Cohort Study. *Rheumatology*. 2022 Aug 03.

- [4] Neira, D. P., Watts, A., Seashore, J., Polychronopoulou, E., Kuo, Y. F., & Sharma, G. Smoking and risk of COVID-19 hospitalization. *Respiratory medicine*. 2021 Apr 17; 182: 106414.

- [5] Ogedegbe G., Ravenell J., Adhikari S., et al. Assessment of Racial/Ethnic Disparities in Hospitalization and Mortality in Patients With COVID-19 in New York City. *JAMA Netw Open*. 2020;3(12): e2026881.