

# Main Manuscript for A Bayesian Framework for Estimating the Risk Ratio of Hospitalization for People with Comorbidity Infected by the SARS- CoV-2 Virus

Xiang Gao<sup>1</sup> and Qunfeng Dong<sup>1,2\*</sup>

<sup>1</sup>Department of Medicine, <sup>2</sup>Center for Biomedical Informatics, Stritch School of Medicine, Loyola University Chicago, Maywood, Illinois 60153, U.S.A

\*Corresponding author: Qunfeng Dong

Email: [gdong@luc.edu](mailto:gdong@luc.edu)

ORCIDiDs

Xiang Gao: <https://orcid.org/0000-0002-1864-837X>

Qunfeng Dong: <https://orcid.org/0000-0002-4610-1670>

## Classification

Biological Sciences, Medical Sciences

## Keywords

COVID-19, SARS-CoV-2, Bayesian, Comorbidity, Hospitalization, Risk Ratio

## Author Contributions

X.G. contributed project conception and data analysis, Q.D. contributed Bayesian modeling and drafting the manuscript.

## This PDF file includes:

Main Text

Figures 1 to 2

Tables 1

## Abstract

Estimating the hospitalization risk for people with certain comorbidities infected by the SARS-CoV-2 virus is important for developing public health policies and guidance based on risk stratification. Traditional biostatistical methods require knowing both the number of infected people who were hospitalized and the number of infected people who were not hospitalized. However, the latter may be undercounted, as it is limited to only those who were tested for viral infection. In addition, comorbidity information for people not hospitalized may not always be readily available for traditional biostatistical analyses. To overcome these limitations, we developed a Bayesian approach that only requires the observed frequency of comorbidities in COVID-19 patients in hospitals and the prevalence of comorbidities in the general population. By applying our approach to two different large-scale datasets in the U.S., our results consistently indicated that cardiovascular diseases carried the highest hospitalization risk for COVID-19 patients, followed by diabetes, chronic respiratory disease, hypertension, and obesity, respectively.

## Significance Statement

We developed a novel Bayesian approach to estimate the hospitalization risk for people with comorbidities infected with the SARS-CoV-2 virus. Our results indicated that cardiovascular

diseases carried the highest hospitalization risk for COVID-19 patients, followed by diabetes, chronic respiratory disease, hypertension, and obesity, respectively.

## **Main Text**

### **Introduction**

Recently published data from China<sup>1</sup>, Italy<sup>2</sup>, and the U.S.<sup>3</sup> showed that a large portion of hospitalized COVID-19 patients had at least one preexisting comorbidity or medical condition. Estimating the hospitalization risk for people with a certain comorbidity condition infected by the SARS-CoV-2 virus would be important for developing public health policies and guidance (e.g., planning medical resources) based on risk stratification.

To provide such an estimation, traditional biostatistical methods (e.g., logistic regression) require knowing both the number of infected people who were hospitalized and the number of infected people who were not hospitalized. Although the number of infected people who were hospitalized might be tallied from hospital records, the number of infected people who were not hospitalized may be difficult to estimate. The reason is that the official count of infected people who were not hospitalized is limited to those tested for the virus. Since not everyone infected by the virus is tested, the number of infected people not hospitalized may be undercounted in the traditional analysis. In addition, comorbidity information for people not hospitalized may not always be readily available for traditional biostatistical analyses. For example, Petrilli et al<sup>4</sup> recently applied multiple logistic regression to identify factors associated with hospital admission using a prospective cohort, but they had stated two main limitations in their study regarding the patients who were not hospitalized: (i) many confirmed patients might not have had their detailed medical history recorded for the study, and (ii) many potential patients might have been omitted from the study since they were not tested for the virus.

To overcome these limitations, we developed a Bayesian approach to estimate the risk ratio of hospitalization for COVID-19 patients with comorbidities. The risk ratio is defined as a ratio of the probability of hospitalization for the infected people with a particular comorbidity (e.g., diabetes) versus the probability of hospitalization for the infected people without any comorbidities. In other words, people without any comorbidities are used as reference for the risk ratio estimation. Our Bayesian approach only needs (i) the number of hospitalized COVID-19 patients and their comorbidity information, which can be reliably obtained using hospital records, and (ii) the prevalence of the comorbidity of interest in the general population, which is regularly documented by public health agencies for common medical conditions (e.g., hypertension, obesity, and diabetes). By applying our approach to two different large-scale datasets in the U.S., we obtained consistent results showing that cardiovascular disease had the highest elevated risk of hospitalization, followed by diabetes, chronic respiratory diseases, hypertension, and obesity, respectively.

## **Results**

### **Estimation with the COVID-NET data**

We applied our Bayesian approach to a dataset<sup>5</sup> of hospitalized COVID-19 patients from the COVID-NET (n=2491). Table 1 provides the summary statistics (median, central 95% Bayesian credible interval) for the estimated posterior distributions of the risk ratio of hospitalization for COVID-19 patients with cardiovascular disease (6.9, 5.1-9.3), diabetes (3.6, 2.9-4.4), COPD (2.6, 2.1-3.2), asthma (2.3, 1.9-2.9), hypertension (1.7, 1.4-2.0), and obesity (1.6, 1.3-1.9). Figure 1A depicts the posterior distributions for these estimated risk ratios.

### **Estimation with the New York data**

A different dataset<sup>3</sup> from the state of New York (n=5700) was used as a comparison. The results from the New York dataset were similar to the ones from the COVID-NET dataset, showing that cardiovascular diseases (6.4, 4.9-8.5) and diabetes (4.8, 4.0-5.8) significantly

increased the risk of hospitalization for COVID-19 patients, followed by asthma (2.2, 1.8-2.6), hypertension (2.1, 1.8-2.5), COPD (1.7, 1.4-2.1), and obesity (1.4, 1.2-1.7) (Table 1, fig. 1B). The only major difference in the ranking of the risk from the COVID-NET and the New York datasets was for COPD, with a lower risk estimated from the New York dataset compared to the COVID-NET dataset.

## Discussion

Using two different large-scale datasets from COVID-NET and New York, our approach obtained similar results, which strongly indicated that COVID-19 patients with cardiovascular disease or diabetes had the highest elevated risk of hospitalization. For example, the risk ratio of hospitalization for COVID-19 patients with cardiovascular disease was estimated to have a median value greater than six, indicating that the hospitalization risk of COVID-19 patients with cardiovascular disease was six times greater than that of COVID-19 patients without any comorbidities. The hospitalization risk for COVID-19 patients also increased with chronic respiratory disease (asthma and COPD), hypertension, and obesity. These comorbidities were selected for this study since they were documented in both the COVID-NET and the New York datasets. Our preliminary exploration with the COVID-NET dataset also indicated elevated risks for people with autoimmune diseases, immune suppression conditions, and renal diseases (data not shown). We encourage other researchers to apply our Bayesian model to their own investigations.

One limitation with our current analysis is that our estimated hospitalization risk for patients with a comorbidity of interest (e.g., diabetes) may be confounded with other comorbidities in the same patient (e.g., hypertension). This limitation is attributed to our lack of access to the necessary data rather than our Bayesian approach *per se*. In this study, we relied upon two published summary statistics (i.e., COVID-NET and New York), which did not include the details of joint comorbidities in their publications<sup>3,5</sup>. For researchers who can access the complete medical records (instead of just summary statistics), they would be able to obtain the frequency of joint comorbidities (e.g., number of COVID-19 hospitalized patients with both diabetes and hypertension). Then, our Bayesian approach could be applied directly to estimate the hospitalization risk for such joint comorbidities. Specifically, instead of using the frequency of a particular comorbidity (e.g., diabetes), our model would use the frequency of the joint comorbidities (e.g., diabetes and hypertension). In addition, the joint prevalence of comorbidities (e.g., diabetes and hypertension) would be used as informative priors.

The only biological assumption in our Bayesian model is that people in the general population, regardless of the status of their comorbidities, could be equally infected by the SARS-CoV-2 virus (no assumptions on the severity of the symptoms after infection were made). The assumption of equal chance of infection by the virus is based on the rationale that SARS-CoV-2 is a newly emerged virus to the human population; thus, nobody is particularly immune to the virus. For example, it has been reported that viral loads were similar in asymptomatic and symptomatic patients<sup>6</sup>, and young children and adults were both similarly infected by the virus<sup>7</sup>. If future research shows that people with specified comorbidities do have a different chance of being infected, our Bayesian approach can be modified to accommodate that difference (e.g., using a scaling factor in our model to reflect the differential degree of infection probability).

## Materials and Methods

### Bayesian modeling

The overview of our Bayesian model is depicted in the figure 2. We classify people in the general population into three categories: (1)  $h$  - the people without any known comorbidities, (2)  $c$  - the people with a particular comorbidity of interest (e.g., diabetes), who may or may not carry other types of comorbidities, and (3)  $o$  - the people with some other types of comorbidities excluding  $c$  (e.g., any other types of comorbidities excluding hypertension). Let  $N_h$ ,  $N_c$ , and  $N_o$  denote the number of hospitalized COVID-19 patients for these three categories, respectively. Let

$N$  denote the total number of COVID-19 patients in the hospital, i.e.,  $N = N_h + N_c + N_o$ . Then the vector  $(N_h, N_c, N_o)$  is a multinomial random variable<sup>8</sup>:

$$(N_h, N_c, N_o) \sim \text{Multinomial}(\theta, N) \quad (1)$$

where  $\theta$  denotes a vector of unobserved multinomial probabilities  $(\theta_h, \theta_c, \theta_o)$  corresponding to  $N_h$ ,  $N_c$ , and  $N_o$ , respectively.

Let  $\tau_h$ ,  $\tau_c$ , and  $\tau_o$  denote the unknown probabilities of hospitalization for people infected by SARS-CoV-2 virus in the category of  $h$ ,  $c$ , and  $o$ , respectively, in the general population. For example, if 50,000 out of a total of 200,000 infected people with a particular comorbidity of interest (i.e., in the category of  $c$ ) were eventually hospitalized,  $\tau_c$  would be equal to 0.25 (i.e., 50,000/200,000). We define  $\tau_c/\tau_h$  to be the risk ratio of the probability of hospitalization for the infected people with a particular comorbidity of interest  $c$  (e.g., diabetes) versus the probability of hospitalization for the infected people without any comorbidities. It's important to note that  $\tau_h$ ,  $\tau_c$ , and  $\tau_o$  are different from  $\theta_h$ ,  $\theta_c$ , and  $\theta_o$ . However, we derived an algebraic relationship between  $\tau_c/\tau_h$  and  $\theta_c/\theta_h$ , as shown below.

Let  $\kappa_h$  and  $\kappa_c$  denote the proportion of people without any medical conditions and the people with a comorbidity of interest  $c$  (e.g., diabetes), respectively, in the general population of size  $N_{pop}$ . Let  $\rho_h$ ,  $\rho_c$ , and  $\rho_o$  denote the probabilities of being infected by SARS-CoV-2 virus for people in the categories of  $h$ ,  $c$ , and  $o$ , respectively. Then,  $\theta$  can be expressed as follows:

$$\theta_h = \tau_h \rho_h \kappa_h N_{pop} / (\tau_h \rho_h \kappa_h N_{pop} + \tau_c \rho_c \kappa_c N_{pop} + \tau_o \rho_o (1 - \kappa_h - \kappa_c) N_{pop}) \quad (2)$$

$$\theta_c = \tau_c \rho_c \kappa_c N_{pop} / (\tau_h \rho_h \kappa_h N_{pop} + \tau_c \rho_c \kappa_c N_{pop} + \tau_o \rho_o (1 - \kappa_h - \kappa_c) N_{pop}) \quad (3)$$

In Eq. (2), the numerator  $(\tau_h \rho_h \kappa_h N_{pop})$  corresponds to the expected number of hospitalized COVID-19 patients without any comorbidities. Specifically, out of the general population of size  $N_{pop}$ ,  $\kappa_h N_{pop}$  is the expected number of people without any comorbidities given the definition of  $\kappa_h$ ;  $\rho_h \kappa_h N_{pop}$  is the expected number of people without any comorbidities infected by the virus given the definition of  $\rho_h$ ;  $\tau_h \rho_h \kappa_h N_{pop}$  is the expected number of infected people without any comorbidities who were hospitalized given the definition of  $\tau_h$ . Similarly, in Eq. (3), the numerator  $(\tau_c \rho_c \kappa_c N_{pop})$  corresponds to the expected number of hospitalized COVID-19 patients with the comorbidity of interest  $c$ . In addition, the term  $\tau_o \rho_o (1 - \kappa_h - \kappa_c) N_{pop}$  in the denominator of both Eq. (2) and (3) corresponds to the expected number of hospitalized COVID-19 patients with some other types of comorbidities excluding  $c$ . Therefore, the denominator in both Eq. (2) and (3) corresponds to the expected total number of hospitalized COVID-19 patients, regardless of the status of their comorbidities.

If we assume that people in the general population, regardless of the status of their comorbidities, have an equal chance of being infected by SARS-CoV-2 virus (see Discussion), then  $\rho_h = \rho_c = \rho_o$  and the Eq. (2) and (3) can be simplified by canceling out  $\rho_h$ ,  $\rho_c$ ,  $\rho_o$ , and  $N_{pop}$  from both the numerator and denominator. The simplified Eq. (2) and (3) are shown as follows:

$$\theta_h = \tau_h \kappa_h / (\tau_h \kappa_h + \tau_c \kappa_c + \tau_o (1 - \kappa_h - \kappa_c)) \quad (4)$$

$$\theta_c = \tau_c \kappa_c / (\tau_h \kappa_h + \tau_c \kappa_c + \tau_o (1 - \kappa_h - \kappa_c)) \quad (5)$$

Dividing Eq. (5) by Eq. (4), we obtain the expression of  $\tau_c/\tau_h$  as follows:

$$\tau_c / \tau_h = \kappa_h \theta_c / \kappa_c \theta_h \quad (6)$$

To estimate the posterior probability of  $\tau_c/\tau_h$ , we need to sample from the following posterior distribution:

$$\text{Prob}(\kappa_h, \kappa_c, \theta | N, N_h, N_c, N_o) \propto \text{Multinomial}(N_h, N_c, N_o | \theta, N) \text{Prior}(\theta) \text{Prior}(\kappa_h) \text{Prior}(\kappa_c) \quad (7)$$

To specify the prior distribution for  $\theta$ , we chose a Dirichlet distribution<sup>8</sup> as it is commonly used as the conjugate prior of the multinomial likelihood described in Eq. (1).

$$(\theta_h, \theta_c, \theta_o) \sim \text{Dirichlet}(a_1, a_2, a_3) \quad (8)$$

where  $a_1, a_2, a_3$  correspond to the shape parameters of Dirichlet distribution. For this study, we set  $a_1 = a_2 = a_3 = 1$  to set a uniform prior, although those values can be adjusted to more accurately reflect the probabilities of hospitalization for each category of patients if more data becomes available in the future.

To specify the prior distributions for  $\kappa_h$  and  $\kappa_c$ , we chose beta distributions as they are commonly used to model proportions<sup>8</sup>.

$$\kappa_h \sim \text{Beta}(\alpha_h, \beta_h) \quad (9)$$

$$\kappa_c \sim \text{Beta}(\alpha_c, \beta_c) \quad (10)$$

where  $\alpha_h, \beta_h, \alpha_c$ , and  $\beta_c$  denote shape parameters of the corresponding beta distributions. Using the method of moments<sup>8</sup>, these parameters can be expressed as follows:

$$\alpha_h = \mu_h(\mu_h(1-\mu_h)/\sigma_h^2 - 1) \quad (11)$$

$$\beta_h = (1-\mu_h)(\mu_h(1-\mu_h)/\sigma_h^2 - 1) \quad (12)$$

$$\alpha_c = \mu_c(\mu_c(1-\mu_c)/\sigma_c^2 - 1) \quad (13)$$

$$\beta_c = (1-\mu_c)(\mu_c(1-\mu_c)/\sigma_c^2 - 1) \quad (14)$$

where  $\mu_h$  and  $\sigma_h^2$ , and  $\mu_c$  and  $\sigma_c^2$  represent the mean and variance of the proportions of the healthy people and people with the comorbidity of interest  $c$ , respectively, in the general population.

Then, by plugging in the priors in Eq. (7), the posterior distribution becomes the following:

$$\begin{aligned} \text{Prob}(\kappa_h, \kappa_c, \theta \mid N, N_h, N_c, N_o) &\propto \text{Multinomial}(N_h, N_c, N_o \mid \theta) \\ &\times \text{Dirichlet}(a_1, a_2, a_3) \\ &\times \text{Beta}(\alpha_h, \beta_h) \\ &\times \text{Beta}(\alpha_c, \beta_c) \end{aligned} \quad (15)$$

In summary, the foundation of our approach is based on our derived algebraic relationship (Eq. 6) between the quantity of  $\tau_c/\tau_h$  (the risk ratio) and the quantities of  $\kappa_h, \kappa_c$ , and  $\theta$ . Using a uniform Dirichlet distribution,  $\theta$  is modeled by a noninformative prior;  $\theta$  is related to the observed data ( $N, N_h, N_c$ , and  $N_o$ ) in hospitalized COVID-19 patients through the multinomial likelihood as described in Eq. (1).  $\kappa_h$  and  $\kappa_c$  are modeled by informative priors using beta distributions whose shape parameters were expressed using the published prevalence rates for comorbidities in the general population. Through sampling from the posterior distribution of  $\kappa_h, \kappa_c$ , and  $\theta$ , we were able to estimate the posterior distribution of  $\tau_c/\tau_h$  as a derived quantity of  $\kappa_h\theta_c/\kappa_c\theta_h$ .

We used WinBUGS<sup>9</sup> (version 1.4.3) to implement the above models. The posterior distributions of risk ratios for different comorbidities were estimated with the Markov Chain Monte Carlo (MCMC) sampling strategy implemented in WinBUGS<sup>10</sup> using the following parameters: the number of chains of four, the number of total iterations of 100,000, burn-in of 10,000, and thinning of 4. Convergence and autocorrelations were evaluated with trace/history and autocorrelation plots. Multiple initial values were applied for MCMC sampling.

### Comorbidity data for hospitalized COVID-19 patients

For the above Bayesian approach, the following two types of data are required: (1) the frequency of the comorbidity of interest (e.g., diabetes) in COVID-19 patients in hospitals, and (2) the prevalence of the comorbidity in the general population. For the comorbidity frequency of hospitalized COVID-19 patients, we used a large-scale dataset, available at COVID-NET<sup>5</sup>, collected from 154 acute care hospitals in 74 counties in 13 states in U.S. from March 1 to May 2, 2020. Among a total of 2491 hospitalized adult patients with laboratory-confirmed COVID-19 in

this COVID-NET dataset, 314 had asthma, 266 had COPD, 859 had cardiovascular diseases, 819 had diabetes, 1154 were obese, 1428 had hypertension, and 336 had no known medical conditions. Besides the COVID-NET dataset, we also used a published dataset from the state of New York<sup>3</sup> collected from 12 hospitals in New York City, Long Island, and Westchester County from March 1 to April 4, 2020. Among a total of 5700 hospitalized COVID-19 patients in this New York dataset, 479 had asthma, 287 had COPD, 966 had cardiovascular diseases, 1808 had diabetes, 1737 were obese, 3026 had hypertension, and 350 had no known medical conditions. In both the COVID-NET and New York datasets, cardiovascular disease referred to coronary artery disease and congestive heart failure.

For the prevalence of comorbidities in general U.S. adult population, the following estimates (mean  $\pm$  standard error) by the U.S. public health government agencies were used: asthma (7.7% $\pm$ 0.22%)<sup>11</sup>, cardiovascular disease (5.6% $\pm$ 0.14%)<sup>12</sup>, COPD (5.9% $\pm$ 0.051%)<sup>13</sup>, diabetes (13% $\pm$ 5.6%)<sup>14</sup>, obesity (42.4% $\pm$ 1.8%)<sup>15</sup>, and hypertension (49.1% $\pm$ 1.5%)<sup>16</sup>. The proportion of healthy adults in the U.S. who have no medical conditions was estimated to be 12.2% (95% CI: 10.9–13.6)<sup>17</sup>.

## Acknowledgments

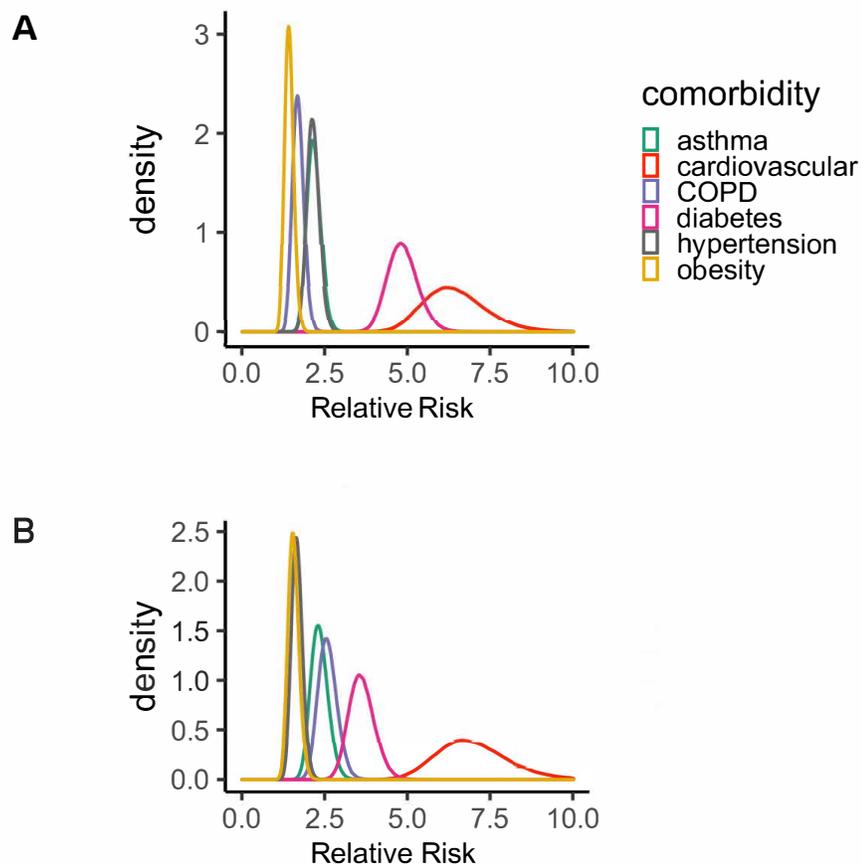
We thank Drs. Lindsay Kim and Gayle E. Langley at the U.S. Centers for Disease Control and Prevention for providing the COVID-NET dataset for our analysis.

## References

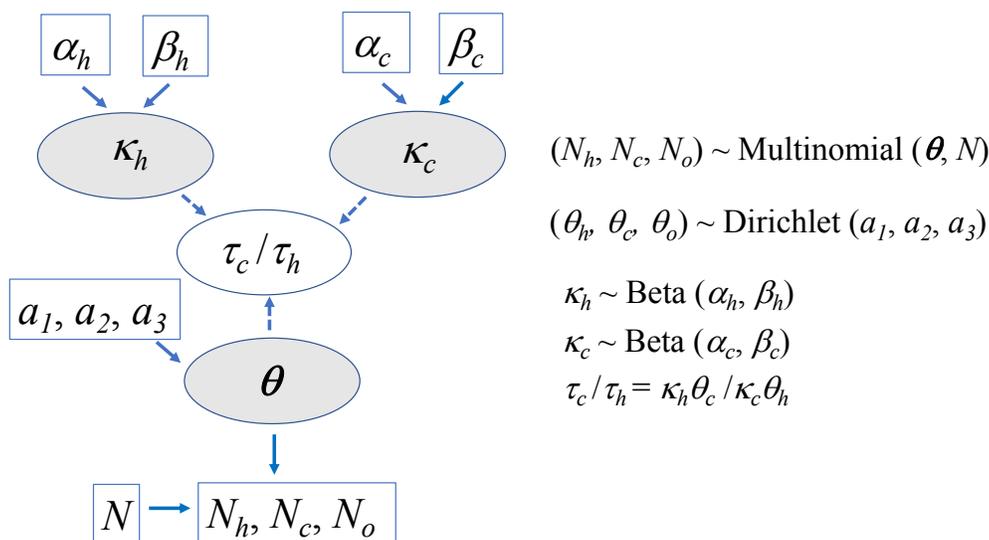
1. Zhou F., Yu T., Du R., et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet*. 2020;395(10229):1054-1062. doi:10.1016/S0140-6736(20)30566-3
2. Onder G., Rezza G., Brusaferro S. Case-Fatality Rate and Characteristics of Patients Dying in Relation to COVID-19 in Italy. *JAMA*. 2020;323(18):1775–1776. doi:10.1001/jama.2020.4683
3. Richardson S., Hirsch J. S., Narasimhan M., et al. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. *JAMA*. Published online April 22, 2020. doi:10.1001/jama.2020.6775
4. Petrilli C. M., Jones S., A, Yang J., Rajagopalan H., O'Donnell L., Chernyak Y., et al. Factors associated with hospital admission and critical illness among 5279 people with coronavirus disease 2019 in New York City: prospective cohort study. *BMJ* 2020; 369 :m1966
5. Kim L., Garg S., O'Halloran A., Whitaker M., Pham H., Anderson E. J., Armistead I., Bennett N. M., Billing L., Como-Sabetti K., Hill M., Kim S., Monroe K. L., Muse A., Reingold A., Schaffner W., Sutton M., Talbot H. K., Torres S. M., Yousey-Hindes K., Holstein R. A., Cummings C., Brammer L., Hall A., Fry A., and Langley G. E. Interim Analysis of Risk Factors for Severe Outcomes among a Cohort of Hospitalized Adults Identified through the U.S. Coronavirus Disease 2019 (COVID-19)-Associated Hospitalization Surveillance Network (COVID-NET) medRxiv 2020.05.18.20103390; doi: <https://doi.org/10.1101/2020.05.18.20103390>
6. Zou L., Ruan F., Huang M., et al. SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients. *N Engl J Med*. 2020;382(12):1177-1179. doi:10.1056/NEJMc2001737
7. [https://zoonosen.charite.de/fileadmin/user\\_upload/microsites/m\\_cc05/virologie-ccm/dateien\\_upload/Weitere\\_Dateien/analysis-of-SARS-CoV-2-viral-load-by-patient-age.pdf](https://zoonosen.charite.de/fileadmin/user_upload/microsites/m_cc05/virologie-ccm/dateien_upload/Weitere_Dateien/analysis-of-SARS-CoV-2-viral-load-by-patient-age.pdf)

8. [Gelman A., Carlin J. B., Stern, H. S., Dunson, D. B., Vehtari A., and Rubin D. B. \(2013\) \*Bayesian Data Analysis \(3rd ed.\)\* London: Chapman & Hall. ISBN 1-439-84095-4.](#)
9. Lunn, D.J., Thomas, A., Best, N., and Spiegelhalter, D. (2000) WinBUGS — a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, **10**:325–337.
10. Lunn D., Jackson C., Best N., Thomas A., and Spiegelhalter D. (2013) *The BUGS Book: A Practical Introduction to Bayesian Analysis*. Boca Raton, FL: Chapman and Hall/CRC. ISBN 978-1-58488-849-9.
11. [https://www.cdc.gov/asthma/most\\_recent\\_national\\_asthma\\_data.htm](https://www.cdc.gov/asthma/most_recent_national_asthma_data.htm)
12. [https://ftp.cdc.gov/pub/Health\\_Statistics/NCHS/NHIS/SHS/2018\\_SHS\\_Table\\_A-1.pdf](https://ftp.cdc.gov/pub/Health_Statistics/NCHS/NHIS/SHS/2018_SHS_Table_A-1.pdf)
13. [https://www.cdc.gov/mmwr/volumes/67/wr/mm6707a1.htm?s\\_cid=mm6707a1\\_w/](https://www.cdc.gov/mmwr/volumes/67/wr/mm6707a1.htm?s_cid=mm6707a1_w/)
14. <https://www.cdc.gov/diabetes/pdfs/data/statistics/national-diabetes-statistics-report.pdf>
15. <https://www.cdc.gov/nchs/data/databriefs/db360-h.pdf>
16. <https://millionhearts.hhs.gov/data-reports/hypertension-prevalence-tables.html> (Table 2)
17. Araújo J, Cai J, and Stevens J. Metabolic Syndrome and Related Disorders. Feb 2019.17(1)46-52. <http://doi.org/10.1089/met.2018.0105>.

## Figures and Tables



**Figure 1.** The posterior probability density of the risk ratio of hospitalization for each comorbidity estimated from the datasets of (A) COVID-NET and (B) New York.



**Figure 2.** Graphic representation of the Bayesian model. On the left, the boxes represent constants for either fixed parameter values ( $\alpha_h, \beta_h, \alpha_c, \beta_c, a_1, a_2, a_3$ ) for priors or observed data ( $N, N_h, N_c, N_o$ ). The grey ovals ( $\kappa_h, \kappa_c, \theta$ ) represent stochastic nodes. The white oval ( $\tau_c / \tau_h$ ) represents a deterministic node. The solid arrows represent stochastic dependence. The dashed arrows represent logical dependence. The corresponding stochastic and deterministic expressions are depicted on the right.

**Table 1.** Summary statistics of the posterior distributions of the hospitalization risk for COVID-19 patients with comorbidities

<b>Comorbidity</b>	<b>Median Risk Ratio (Central 95% Bayesian Credible Interval)</b>	
	<b>COVID-NET</b>	<b>New York</b>
<b>Asthma</b>	2.331 (1.879-2.893)	2.165 (1.793-2.610)
<b>Cardiovascular disease</b>	6.884 (5.142-9.328)	6.370 (4.829-8.498)
<b>COPD</b>	2.576 (2.079-3.187)	1.694 (1.393-2.054)
<b>Diabetes</b>	3.599 (2.924-4.427)	4.839 (4.027-5.808)
<b>Hypertension</b>	1.660 (1.368-2.016)	2.143 (1.806-2.540)
<b>Obesity</b>	1.554 (1.268-1.907)	1.425 (1.186-1.711)