Multi-level multi-state modelling applied to hospital admission in mexican patients with COVID-19

Juan Pablo Diaz Martinez[1], Karen Janik Orozco Becerril, Marco Antonio Gallegos Herrada, Mayra Alejandra Gutierrez Garcia, Osvaldo Espin-Garcia[2], Ruth Fuentes Garcia[3]*

* Corresponding author: rfuentes@ciencias.unam.mx

Abstract

Since the beginning of the SARS-CoV 2 pandemic, healthcare authorities have made clear that it is crucial to track and identify COVID-19 symptoms and seek medical attention in the presence of the first warning signs, as immediate medical attention can improve the patient's prognosis. Therefore the present work aims to analyze the risks associated with the time between the patient's first symptoms and hospitalization followed by death. A cross-sectional study was performed among Mexican population diagnosed with COVID-19 and hospitalized from March to January 2021. Four different multi-level multi-state Bayesian models were developed to asses risk factors associated with different patient trajectories: symptoms-hospitalization and hospitalization-death. Comorbidities that could worsen the patient outcome were included as linear predictions; these analyses were further broken down to the different states of the Mexican Republic and the healthcare providers within. Model III was chosen as it demonstrated the best performance in terms of leave one out (LOO) validation. Increased risk for hospitalization was observed at the global population level for chronic renal disease, whereas for death such was the case for Chronic Obstructive Pulmonary Disease (COPD) and the three-way interaction between diabetes, hypertension and obesity. Our results show that there are differences in mortality among the states without accounting for institution and it is related to the prompt time of death or viceversa. Risk differences among the studied six healthcare providers were also found while patients from state-managed hospitals and private sector showed lower risks, patients from the Mexican Institute of Social Security (IMSS) seem to be at highest risk. The proposed modelling can be helpful to improve healthcare assistance at a regional level , additionally it could inform statistical parameter inference in epidemiological models.

Introduction

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic was declared a Public Health Emergency of International Concern on January 30, 2020 by the World Health Organization.The Mexican Health Authorities declared the first lockdown on March 26 with 585 cases and 8 deaths reported for COVID-19 [1]; by the end of the first lockdown (June 5th, 2020) total number of cases and deaths were 110,026 and 13,170, respectively. By November 1, Mexico became the fourth country in number

---

[1]Institute of Health, Policy & Management, University of Toronto
[2]Dalla Lana School of Public Health, University of Toronto
[3]Facultad de Ciencias, UNAM

of deaths of SARS-CoV-19 (106,765 deaths), with 1,122,362 incident cases [2]; by April 15th 2021 the number of deaths had raised to 214,372 with 2,309,099 incident cases [3].

Over time it has become clear that comorbidity factors such as hypertension, type 2 diabetes mellitus, obesity and smoking increase the seriousness of the disease, lead to a higher rate of hospitalizations with an additional 25% of the cases requiring intensive care unit (ICU) admission and ultimately, intubation and death [4,5].

Mexico ranks second in obesity among OECD countries, with an obesity rate of 72.5% among the adult population, which is associated with the high prevalence of type 2 diabetes mellitus, estimated at 13% of the adult population in 2017, the highest rate among OECD countries; the rate of hypertension is also one of the highest chronic diseases among adult population with 30% [6]. The high prevalence of these comorbidities together with the precarious healthcare system could be among the main reasons of the elevated severity of the number of cases and deaths rates in the country [7,8].

Mexican, healthcare providers are divided in public and private services. There are different public institutions which provide care to different populations: federal government employees (ISSSTE), the army (SEDENA) and navy personnel (SEMAR), workers from the state-owned oil company (PEMEX) and private companies employees (IMSS). There are also public hospitals for population with no health service coverage (SSA). In general, the care within different healthcare providers cannot be considered homogeneous, therefore it is likely relevant and informative for the final outcome of a COVID-19 patient.

There have been many efforts using local data to understand how patients' comorbidities are affected by COVID-19; the work by Bello-Chavolla et al. [9] proposed a clinical score to predict COVID-19 lethality, including different factors like type 2 diabetes mellitus and obesity among confirmed and negative COVID-19 cases in Mexico. This work lead to believe that obesity mediates 49.5% of the effect of diabetes on COVID-19 lethality. Also, early-onset diabetes conferred an increased risk of hospitalization while obesity increased the risk ICU for admission and intubation. Moreover, Olivas-Martínezet al. [5] found that main risk factors associated with in-hospital death were male sex, obesity and oxygen saturation < 80% on admission using data from a SARS-CoV-2 referral center in Mexico City.

After onset of infection there is a period of time between symptom detection and hospitalization. The time elapsed before patients approach hospitals could be excessively long. Once patients are admitted to hospital, there is also a period of time between the admission and death. Estimation of these lengths of time through a multilevel model could enable a better information system to estimate incidence and transmission rates, particularly at regional level where differences can be apparent. In addition, these times are useful for estimating hospitalizations and deaths in COVID-19 epidemiological models [10,11].

This work considers a multi-state model under a Bayesian framework to estimate times between symptom detection and hospitalization and between hospitalization and death. We used data of confirmed and negative COVID-19 cases and their demographic and health characteristics from the General Directorate of Epidemiology of the Mexican Ministry of Health; the analysis provides of general overview of these times in each state of the country and the different health institutions within. Variables believed to affect the patient's final outcome such as the aforementioned comorbidites are included in the model as fixed effects. Additionally, regional heterogeneity is accounted for as random effects through nested models that consider the regional contribution and the healthcare service provider. Other efforts in recent literature [12] have considered more states (hospitalization-ICU, ICU-death, ICU-discharged), which allows researchers to asses whether improvements in patient outcomes have been sustained, finding evidence that

median hospital stays have lengthened. Unfortunately, data available for Mexico lacks the necessary granularity to determine such states. Nevertheless, we believe this model could better inform the estimation of the incidence and transmission rates, which is particularly important while new variants and increased transmission rates are present.

Methods and materials

Data Source and Study Population

We conducted an observation study using the official database from the Mexican Ministry of Health, these data provide an overview of hospital admissions, deaths and the period of time between hospitalizations and first COVID-19 symptoms between March and January 2020. The data analyzed included confirmed individuals with a positive test for SARS-CoV-2 [3]; in late 2020, the Mexican Ministry of Health change its confirmed definition with the objective of including postmortem cases. The information recorded on every individual includes: sex, age, nationality, place of residence, migratory status, and different comorbidites. Data registered on the COVID-19 event includes: type of first contact medical unit, management received (either hospitalization or outpatient), and dates of onset of COVID-19 symptoms, admission to hospitalization, and death. Data on the evolution during the stay in the medical units were not released for public use, such as date of recovery. Exclusion criteria were the observations with incomplete data about hospital admission, symptoms or comorbidities. Additionally, patients whose time of initial symptoms was captured as the day they were admitted to hospital were removed, since this time was likely to be unknown. Finally, we only included patients who experienced either hospitalization or death due to the lack of date of recovery in the dataset.

The following variables were included as linear predictors for modeling time from symptoms to hospitalization: presence of chronic obstructive pulmonary disease (COPD), obesity, chronic kidney disease (CKD), asthma and immune-suppression. For time from hospitalization to death, we included the following variables: presence of type 2 diabetes mellitus, COPD, obesity, hypertension, CKD and the interaction between obesity, diabetes and hypertension. Both times also included age and sex as predictors.

About 87% of the population in Mexico is affiliated to some healthcare provider, but during this pandemic the mexican government established a list of hospitals designated to treat COVID-19 patients without any affiliation distinction. In this study we identified 6 different healthcare providers which were classified according to their sectors IMSS, ISSSTE, SEDENA/SEMAR/PEMEX, SSA, ESTATALES (healthcare provider within each state) these five constitute public care providers while the sixth sector is private hospitals.

Modelling

We developed four different models for the trajectories of interest, *Symptoms-Hospitalization* and *Hospitalization-Death*.

Figure 1 goes here.

We used a $\mathbf{QR}$ reparameterization for the predictor matrix $\mathbf{X}$, i.e. $\mathbf{X} = \mathbf{QR}$, where $\mathbf{Q}$ is an orthogonal matrix and $\mathbf{R}$ is an upper triangular matrix. This parameterization is recommended when no prior information is available on the location of the predictors'

coefficients [13]. Moreover, we used a noncentered parameterization [14] by shifting the data's correlation with the parameters to the hyperparameters.

Each model captures different levels of information, as more levels were included it was possible to differentiate the results according to the added information.

Model I: One level

Let M and H correspond to survival times for deaths and hospitalizations, respectively. We assumed that these times are observations from two independent Weibull distributions, such that,

$$M \sim Weibull(\alpha, \eta)$$
$$H \sim Weibull(\alpha, \upsilon)$$
$$\eta = \exp\left(-\frac{\mu_m + \mathbf{Q}^*\vartheta}{\alpha}\right)$$
$$\upsilon = \exp\left(-\frac{\mu_h + \mathbf{Q}^{**}\theta}{\alpha}\right)$$
$$\alpha = \exp(\alpha_r * 10)$$
$$\alpha_r \sim N(0, 1)$$
$$\mu_m, \mu_h \sim N(0, 10)$$
$$\vartheta, \theta \sim U(-\infty, \infty)$$

where $\mathbf{Q}^*$ and $\mathbf{Q}^{**}$ are the orthogonal matrices from the $\mathbf{QR}$ reparameterization, θ and ϑ are the coefficient vectors for deaths and hospitalizations, μ_m and μ_h represent the global intercepts for deaths and hospitalizations, $alpha$ denotes the shape of the Weibull distribution, and α_r is an extra parameter for the noncentered parameterization This part of the model is described in red in Figure 1.

Model II: Two levels

The second model adds an additional level to account for each state in Mexico as a random effect to explain deaths, such that,

$$M_l \sim Weibull(\alpha, \eta)$$
$$H \sim Weibull(\alpha, \upsilon)$$
$$\eta = \exp\left(-\frac{\mu_m + \mu_l^r + \mathbf{Q}^*\vartheta}{\alpha}\right), l = 1, ..., 32$$
$$\upsilon = \exp\left(-\frac{\mu_h + \mathbf{Q}^{**}\theta}{\alpha}\right)$$
$$\mu_l = \sigma * \mu_l^r$$
$$\alpha = \exp(\alpha^r * 10)$$
$$\alpha^r \sim N(0, 1)$$
$$\mu_l^r \sim N(0, 1)$$
$$\sigma \sim t_3^+(0, 1)$$
$$\mu_m, \mu_h \sim N(0, 10)$$
$$\vartheta, \theta \sim U(-\infty, \infty)$$

where $\mu_l, l = 1, ..., 32$ represents a local intercept per state (random effect), μ_l^r denotes the extra paramateres for the noncentered parameterization, and σ is the dispersion

parameter of the local intercepts. This part of the model is described in light red in
Figure 1.

Model III: Three levels Based on Model II, we considered an extra random effect
$\mu_k, l = 1, ..., 6$ which takes into account the healthcare provider previously described.
Because of this, an extra dispersion parameter σ is added. This part of the model is
described in light pink in Figure 1.

$$
\begin{aligned}
M &\sim Weibull(\alpha, \eta) \\
H &\sim Weibull(\alpha, \upsilon) \\
\eta &= \exp\left(-\frac{\mu_m + \mu_l^r + \mu_k^r + \mathbf{Q}^*\vartheta}{\alpha}\right), l = 1, ..., 32, k = 1, ..., 5 \\
\upsilon &= \exp\left(-\frac{\mu_h + \mathbf{Q}^{**}\theta}{\alpha}\right) \\
\mu_l &= \sigma_l * \mu_l^r, l = 1, 2 \\
\mu_k &= \sigma_l * \mu_k^r \\
\alpha &= \exp(\alpha^r * 10) \\
\alpha^r &\sim N(0, 1) \\
\mu_l^r, \mu_k^r &\sim N(0, 1) \\
\sigma_l &\sim t_3^+(0, 1) \\
\mu_m, \mu_h &\sim N(0, 10) \\
\vartheta, \theta &\sim U(-\infty, \infty)
\end{aligned}
$$

Model IV:

The final model uses the same structure as Model III but nests healthcare provider
within each state.

Model fit

As mentioned before, we considered a Bayesian approach to fit the model to the
observed data corresponding to times between symptoms and hospitalization, H, and
time between hospitalization and death, M. The associated model parameters are given
by $\boldsymbol{\theta} = \{\alpha^r, \eta, \upsilon, \mu_m, \mu_h, \theta, \vartheta, \mu_l^r, \mu_k^r\}$. We form the joint posterior distribution over both
the model parameters and auxiliary variables, given the observed data:

$$
\pi(\boldsymbol{\theta}|H, M) \propto f(H, M|\boldsymbol{\theta})p(\boldsymbol{\theta})
$$

Inference is carried out using Markov chain Monte Carlo (MCMC), obtaining a
sample from the joint posterior distribution over the parameters given the observed data.
All Markov chains were generated with `CmdStan` [15], using the No-U-Turn sampler [16].
A check on the posterior predictive distributions is essential for validating results.
Additionaly, to choose the model that best fits the data we considered the leave-one-out
cross-validation (LOO) proposed by Vehtari et al. [17], which estimates pointwise
out-of-sample prediction accuracy, using the log-likelihood evaluated at the posterior
simulations of the parameter values.

Model	elpd leave one out	p leave one out
Model I	-75473.3 (187.8)	26.9 (2.2)
Model II	-75352.2 (186.3)	84.1 (4.7)
Model III	-75284.9 (186.6)	92.5 (4.9)
Model IV	-75612.0(194.4)	239.4(15.0)

Table 1. Expected log-pointwise predictive density (ELPD) for a new data set and effective number of parameters (standard deviation).

Results

After applying exclusion criteria a total sample of 1200 registers of adult patients belonging to any healthcare provider, either private or public in the 32 states of Mexico was selected, preserving the population characteristics.

We show results for model III which performed better in terms of the likelihood and showed good convergence of all parameteres. The posterior 0.95 credibility intervals for parameters of interest at different levels of the model are shown in Figures 2 and 3. It is worth pointing out that we are displaying the log hazard ratio, hence positive values for parameters will point to increasing risks for the corresponding transition and level.

Figure 2 goes here.

Figure 3 goes here.

Increased risk for hospitalization was observed at the global population level for chronic renal disease, whereas for death such was the case for COPD and the interaction of diabetes, hypertension, and obesity.

Figure 4 goes here.

Our results show that there are differences in mortality between the states without accounting for institution and it is related to the prompt time of death or viceversa. Figure 2 shows the states in which the overall rate of mortality is higher such as Campeche, Colima, Guanajuato, Hidalgo, Jalisco, Morelos, Nayarit, Oaxaca, Puebla, Tabasco and Veracruz. The difference might be linked to the late hospitalization of patients.

Figure 4 displays evidence that there is a peak on mortality rate 5 days after hospitalization, which could be related to the late hospitalization of patients with mild symptoms who developed "happy hypoxemia," that is extremely low blood oxygenation, but without sensation of dyspnea [18]. In Wuhan, within a cohort of patients infected with (SARS-COV-2) who did not present dyspnea 62% showed severe disease and 46% ended up intubated, ventilated or dead [19].

Regarding the 6 healthcare providers included in the analysis differences were also found. State-managed hospitals and private sector showed lower risks compared to IMSS which seems to be the one with the highest risk (Figure 3). Although it is worth mentioning that following the national hospital transformation plan [20], IMSS alone has transformed about 260 medical units to treat COVID-19 patients [21]. Additionally it has the largest number of affiliations and they are liklely to have higher risk exposure.

Disscusion

Multiple sources have shown that the presence of comorbidities such as diabetes, hypertension, obesity, COPD, asthma, inmmuno-suppression and chronic kidney disease are associated to a worse outcome for the patients diagnosed with COVID-19. Particularly for those who are hospitalized in the ICU due intubation. To our knowledge this is the first study in Mexico which analyzed the time elapsed between the patient´s first symptoms, hospitalization and death; these analyses were further broken down to the different states of the republic and the healthcare providers within them. Thus it is possible to identify providers and states with an increased risk of hospitalization and death. Mexico is the third place in obesity among the OCDE countries, which could be the main reason of the high number of severe cases of COVID-19.

One problem that aggravates the situation is the precarious state of the public healthcare system which universal coverage is estimated about 87% of the mexican population. It is clear that mexican healthcare has overrun during this pandemic and has appealed to private health providers to cope with the treatment of COVID-19 patients. Among all health care providers the IMSS is the one with the highest risk, however it is also the largest healthcare provider across Mexico with hospitals from level 2-level 4 of which "Siglo XXI" is a country-leader in research and innovative treatments and procedures. In March, 2020 38 hospitals were "converted" [20] to exclusively treat COVID-19 among which 18 were IMSS hospitals only; after one year 960 hospitals across the country were converted to treat patients with COVID-19 of these 289 (30%) belong to IMSS [21].

The proposed modelling can be helpful to a regional level to improve healthcare assistance, it could additionally be useful to inform statistical estimation of parameters for an epidemiological model.

This study has shown that there are differences in mortality between the different states of the republic; there are regions in which the overall rate of mortality is higher due the late hospitalization of patients such as Veracruz, Nuevo Leon, San Luis Potosí, Guanajuato, Chiapas and Mexico City. Breaking down this analysis to regional levels we found a higher risk of hospitalization, specifically in Veracruz which has been historically unsteady regarding the public healthcare system in sectors like IMSS, ISSSTE, SEDENA/SEMAR/PEMEX. The fact that different final outcomes could be related to patient's late hospitalization, hence suggesting that the average patient waits until the symptoms are severe to seek professional healthcare, needs to be further investigated.

One of the limitations of the study is the reduced number of states we were able to include in the models due to the lack of information regarding dates of discharged of recovered patient´s after hospitalization.

References

1. Carrillo-Vega MF, Salinas-Escudero G, García-Peña C, Gutiérrez-Robledo LM, Parra-Rodríguez L. Early estimation of the risk factors for hospitalization and mortality by COVID-19 in Mexico. PLOS ONE. Public Library of Science; 2020;15: e0238905. Available: https://doi.org/10.1371/journal.pone.0238905

2. Organización Panamericana de la Salud O. Documentos técnicos de la OPS - Enfermedad por el Coronavirus (COVID-19) [Internet]. 2021. Available: https://www.paho.org/es/documentos-tecnicos-ops-enfermedad-por-coronavirus-covid-19

3. [Internet]. Sitio oficial COVID-19 México Dirección General de Epidemiología - Gobierno Federal. Available: https://covid19.sinave.gob.mx/

4. SEDESA. Base del Sistema Nacional de Vigilancia Epidemiologica para el seguimiento a posibles casos de COVID-19 en la Ciudad de México [Internet]. 2021.

Available: http://sinave.gob.mx/ 226

5. Olivas-Martínez A, Cárdenas-Fragoso JL, Jiménez JV, Lozano-Cruz OA, 227
Ortiz-Brizuela E, Tovar-Méndez VH, et al. In-hospital mortality from severe covid-19 in 228
a tertiary care center in mexico city; causes of death, risk factors and the impact of 229
hospital saturation. Plos one. Public Library of Science San Francisco, CA USA; 230
2021;16: e0245772. 231

6. OCDE. Health at a Glance 2019: OECD Indicators. [Internet]. OCDE; 2019 p. 4. 232
Available: https://www.oecd.org/mexico/health-at-a-glance-mexico-ES.pdf 233

7. Burki T. COVID-19 in latin america. The Lancet Infectious Diseases. Elsevier; 234
2020;20: 547–548. 235

8. Ji Y, Ma Z, Peppelenbosch MP, Pan Q. Potential association between covid-19 236
mortality and health-care resource availability. The Lancet Global Health. Elsevier; 237
2020;8: e480. 238

9. Bello-Chavolla OY, Bahena-López JP, Antonio-Villa NE, Vargas-Vázquez A, 239
González-Díaz A, Márquez-Salinas A, et al. Predicting Mortality Due to SARS-CoV-2: 240
A Mechanistic Score Relating Obesity and Diabetes to COVID-19 Outcomes in Mexico. 241
Journal of Clinical Endocrinology and Metabolism. 2020;105: 2752–2761. 242
doi:10.1210/clinem/dgaa346 243

10. Flaxman S, Mishra S, Gandy A, Unwin HJT, Mellan TA, Coupland H, et al. 244
Estimating the effects of non-pharmaceutical interventions on covid-19 in europe. 245
Nature. Nature Publishing Group; 2020;584: 257–261. 246

11. Unwin HJT, Mishra S, Bradley VC, Gandy A, Mellan TA, Coupland H, et al. 247
State-level tracking of covid-19 in the united states. Nature communications. Nature 248
Publishing Group; 2020;11: 1–9. 249

12. Kirwan PD, Elgohari S, Jackson CH, Tom BDM, Angelis DD, Presanis AM. 250
Trends in risks of severe events and lengths of stay for COVID- 19 hospitalisations in 251
England over the pre-vaccination era : results from the Public Health England 252
SARI-Watch surveillance scheme.: 1–45. Available: 253
https://arxiv.org/ftp/arxiv/papers/2103/2103.04867.pdf 254

13. Team SD. Stan user's guide [Internet]. Stan. Available: 255
https://mc-stan.org/docs/2_26/stan-users-guide/index.html 256

14. Papaspiliopoulos O, Roberts GO, Sköld M. A general framework for the 257
parametrization of hierarchical models. Statistical Science. JSTOR; 2007; 59–73. 258

15. Team SD. CmdStan: The command-line interface to stan [Internet]. CmdStan. 259
Available: https://mc-stan.org/users/interfaces/cmdstan 260

16. Hoffman MD, Gelman A. The no-u-turn sampler: Adaptively setting path 261
lengths in hamiltonian monte carlo. J Mach Learn Res. 2014;15: 1593–1623. 262

17. Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using 263
leave-one-out cross-validation and WAIC. Statistics and Computing. 2017;27: 264
1413–1432. doi:10.1007/s11222-016-9696-4 265

18. González-Duarte A, Norcliffe-Kaufmann L. Is 'happy hypoxia' in COVID-19 a 266
disorder of autonomic interoception? A hypothesis. Clinical Autonomic Research. 267
Springer Berlin Heidelberg; 2020;30: 331–333. doi:10.1007/s10286-020-00715-z 268

19. Guan W-j, Ni Z-y, Hu Y, Liang W-h, Ou C-q, He J-x, et al. Clinical 269
Characteristics of Coronavirus Disease 2019 in China. New England Journal of 270
Medicine. 2020;382: 1708–1720. doi:10.1056/nejmoa2002032 271

20. Mendoza-Popoca CÚ, Suárez-Morales M. Hospital reconversion in response to 272
the COVID-19 pandemic. Revista Mexicana de Anestesiologia. 2020;43: 151–156. 273
doi:10.35366/92875 274

21. UNAM LI de T e IE(L del I de G de la. Sistema de Información de la Red IRAG 275
[Internet]. 2021. Available: 276
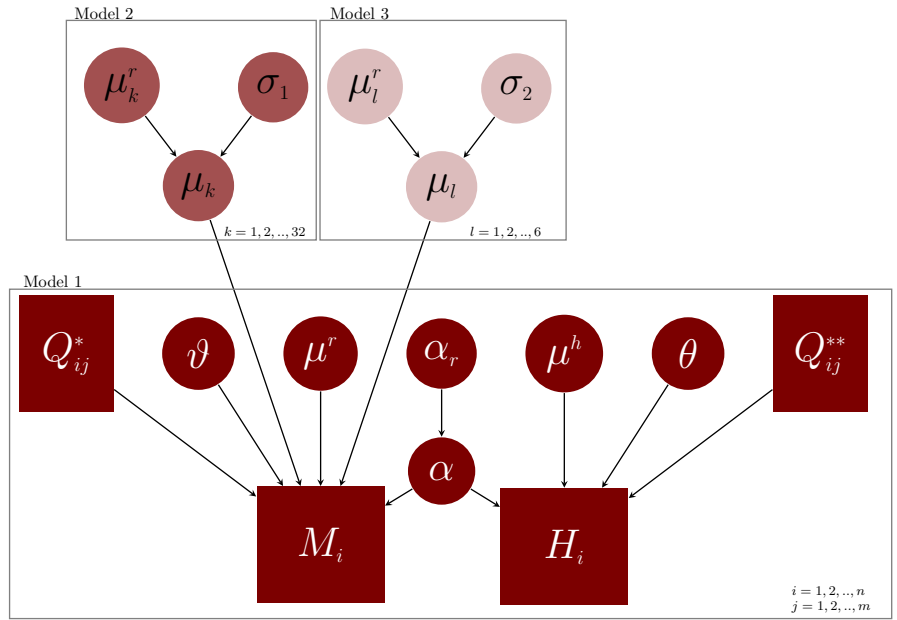https://www.gits.igg.unam.mx/red-irag-dashboard/reviewHome 277

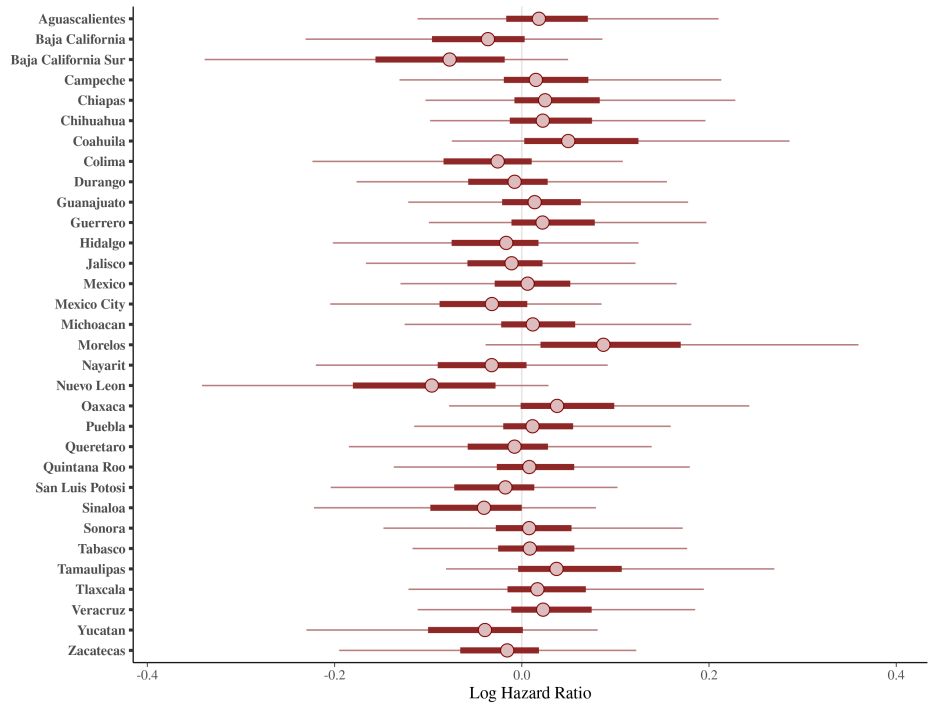**Figure 1**. Graphical representation of the model: directed acyclic grph (DAG)

Figure 2. Log hazard rate for state random effect (95% Credible Interval)
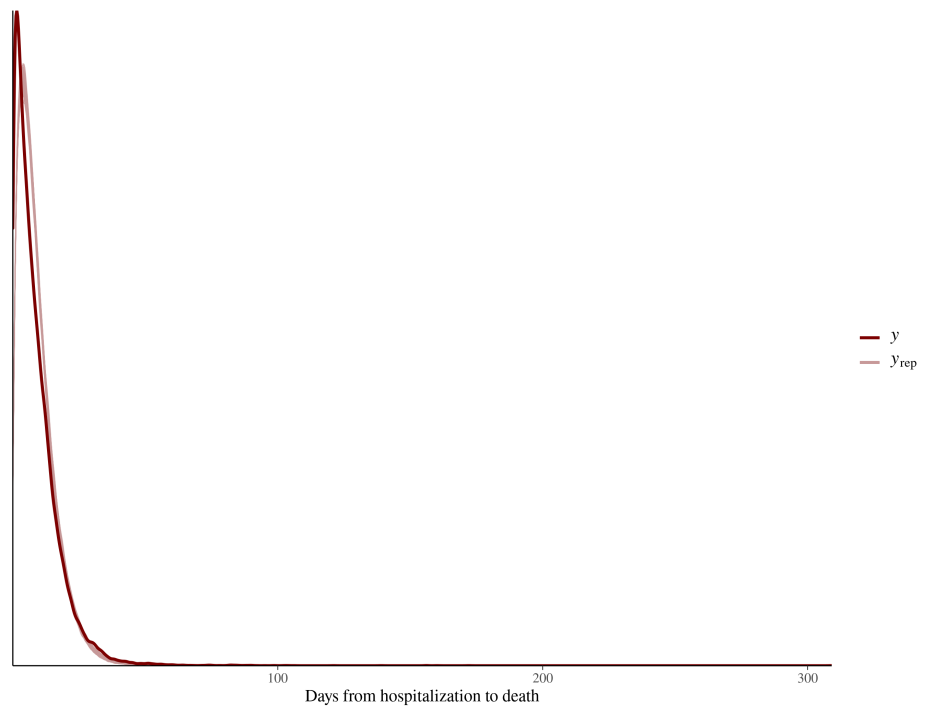
Figure 3. Posterior predictive distribution density plot for deaths
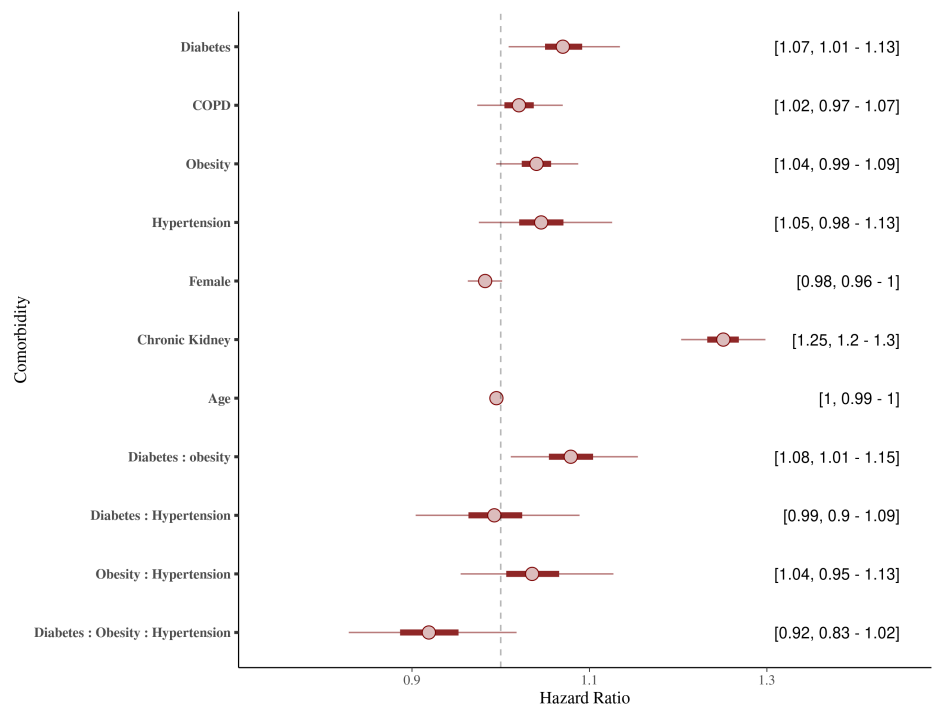
Figure 4. Effect on time from hospitalization to death by covariates included in the final model.

```r
#table_Comorbidites = read.csv("../ArtículoAbril2021/tableOne_comorbidities&Tim
table_Comorbidities = read.csv("../ArtículoAbril2021/tableOne_comorbidities&Time
# tibble::print.tbl_df(table)
#knitr::kable(table_Comorbidities,
#             #row.names = T,
#             caption = "Comorbidites")
```