



# Modelado estadístico para la toma de decisiones sobre la pandemia causada por el COVID19

Coordina: Héctor Benítez Pérez, IIMAS, UNAM  
Primer Reporte: 13 de abril 2020

## **Contenido:**

Reporte de datos, análisis y modelos al 13 de abril 2020

## **Anexos:**

Reporte análisis covid jorge l. Perez

Informe\_logistica

Subgrupogeo

Iniciativa unam covid19

Reporte\_vulnerabilidad\_1

Informe de actividades-análisis de twitter



## Reporte al 13 de abril de 2020

### Modelado estadístico para la toma de decisiones sobre la pandemia causada por el COVID19

Este documento reporta de manera resumida los avances de un grupo de trabajo conformados por académicos de la UNAM para coadyuvar en la toma de decisiones ante la contingencia creada por el COVID19. El grupo de académicos que comenzó a trabajar el pasado viernes 3 de abril de 2020 esta integrado por:

- |                                  |                              |                                 |
|----------------------------------|------------------------------|---------------------------------|
| 1. Luis Álvarez Icaza (IIng)     | 16. Fabián García (IIMAS)    | 31. Antonio Neme (IIMAS)        |
| 2. Igor Barahona (IM - UC)       | 17. Nancy Gonzalez (IM-UJ)   | 32. Cecilia Noguez (IF)         |
| 3. Luis Benet (ICF)              | 18. Rosa Gutierrez (IBT)     | 33. Miguel Núñez Kant (UdG)     |
| 4. Denis Boyer (IF)              | 19. Noe Hernández (IIMAS)    | 34. Eugenia O'Reilly (IM)       |
| 5. Gilberto Calvillo (IM-UC)     | 20. Carlos I. Herrera (UNAM) | 35. Diego Palomares (IBT)       |
| 6. Ruth Corona (IM-UJ)           | 21. Claudia Juárez (IIMAS)   | 36. Isaac Pérez Castillo (IF)   |
| 7. Ricardo Cruz (IIMAS)          | 22. María Lárraga (IIng)     | 37. Luis Pineda (IIMAS)         |
| 8. Marco Herrera (FC)            | 23. Carlos López (IF)        | 38. Carlos E. Rodríguez (IIMAS) |
| 9. Antonio Capella (IM)          | 24. Jesus Lopez (IM-UC)      | 39. Ernesto Rubio (IIMAS)       |
| 10. Fabio De Colle (ICN)         | 25. Sergio I. Lopez (FC)     | 40. Erick Treviño (IM - UC)     |
| 11. Carlos Del Burgo (INAOE)     | 26. José Antonio Loza (IBT)  | 41. Emmanuel Torres (IIng)      |
| 12. Jesús del Río (IIng)         | 27. Natalia Mantilla (FC)    | 42. Marco Tilio Ángulo (IM)     |
| 13. Celia Escamilla Rivera (ICN) | 28. Ramsés H. Mena (IIMAS)   | 43. Octavio Valenzuela (IA)     |
| 14. Zian Fanti Gutiérrez (IIMAS) | 29. Carlos Nataren (IF)      | 44. Blanca H. Vázquez (IIMAS)   |
| 15. Gibran Fuentes (IIMAS)       | 30. Lukas Nellen Filla (ICN) | 45. Jorge X. Velasco (IM)       |

Acordamos seguir tres etapas:

**Etapa 1** Articular preguntas y temas en los que podemos trabajar y aportar algo.

**Etapa 2** Identificar tareas específicas, maximizando nuestras áreas de dominio

**Etapa 3** Conformar un reporte preliminar.

Al momento el grupo se encuentra en las primeras dos etapas y se han explorado los temas que a continuación resumimos en las siguientes sub-secciones.

# Índice

1. Integración y depuración de datos disponibles	3
2. Análisis estadísticos	4
3. Modelos epidemiológicos	7
3.1. Incertidumbre y planeación . . . . .	7
3.2. Evaluación de políticas de control . . . . .	10
4. Modelado de la pandemia por medio de un análisis estadístico de datos	10
5. Estimaciones y modelaje estocástico de dinámica de propagación en una población para distintos grupos de riesgo	12
6. Análisis de marcadores de riesgo	13
7. Estimation of hidden COVID-19 cases in Mexico City	15
8. Modelado matemático en pacientes con COVID-19	16
9. Estudio de viabilidad para la estimación de escenarios en México para la mitigación de muerte por desbalance entre demanda y capacidad por la epidemia causada por el virus SARS-CoV-2	17
10. Detectando el impacto del transporte público sobre la transmisión del COVID-19 en la Ciudad de México.	23
11. COVID-19 forecasting	25
12. Report on the consequences of the mitigation measures announced on March 23 and March 30.	30

# 1. Integración y depuración de datos disponibles

(-Equipo transversal-)

**Integrantes:** Ricardo Cruz (IIMAS), Rosa María Gutiérrez (IBT), Zian Fanti Gutiérrez (IIMAS), Noe Hernández (IIMAS), José Antonio Loza (IBT), Diego Palomares (IBT), Luis A. Pineda (IIMAS).

**Objetivo general:** Desarrollar e implementar de una base de datos para el seguimiento a pacientes COVID y análisis de la evolución de la epidemia a nivel nacional.

**Avance (15 %):**

Usando ingeniería reversa y tomando como sustrato las tablas proporcionadas por el gobierno de la CDMX y que contienen el vaciado del cuestionario de ingreso de los pacientes a las unidades médicas receptoras (datos personales, catálogo de hospitales receptores, catálogo de pruebas y laboratorios, validación de pruebas, sintomatología y comorbilidades, evolución del paciente e itinerarios de viaje). Contamos con fechas de ingreso, evolución de los pacientes y fechas de realización de pruebas. Esto tomando en cuenta los datos proporcionados entre el 02 y 05 de abril.

La tabla proporcionada fue organizada en nuevas tablas para construir un modelo de base de datos entidad relación que tiene como objetivos específicos los siguientes.

**Objetivos específicos:**

1. Seguimiento de pacientes COVID atendidos en hospitales del sector público y privado.
2. Organización de los datos proporcionados en una base de datos relacional.
3. Generación y visualización de consultas.
4. Acceso remoto de la información y consultas por los miembros del grupo.
5. Incorporación de resultados de los grupos de modelación.
6. Proveer al sector salud con datos y análisis que permitan tomar acciones pertinentes en función de escenarios observados y proyectados.

Se construyó el modelo de base de datos al 10 de Abril del 2020, que está sujeto a cambios tomando en cuenta cambios en la información proporcionada y necesidades de los grupos de investigación. En la Figura 1, se muestran las tablas creadas y sus propiedades.

El sistema a la fecha de este informe es capaz de proporcionar “queries” usando como herramienta la linea de comandos. No obstante, el grupo de trabajo esta avanzando en el desarrollo de una interfase web que permita a los colaboradores de del grupo de modelado obtener la información para el desarrollo de modelos.

**Métodos desarrollados hasta el momento.**

- Se analizaron los datos contenidos en tablas proporcionadas por el gobierno de la CDMX.
- Se reclasificaron los datos de acuerdo a atributos, que permitieron agruparlos en tablas.

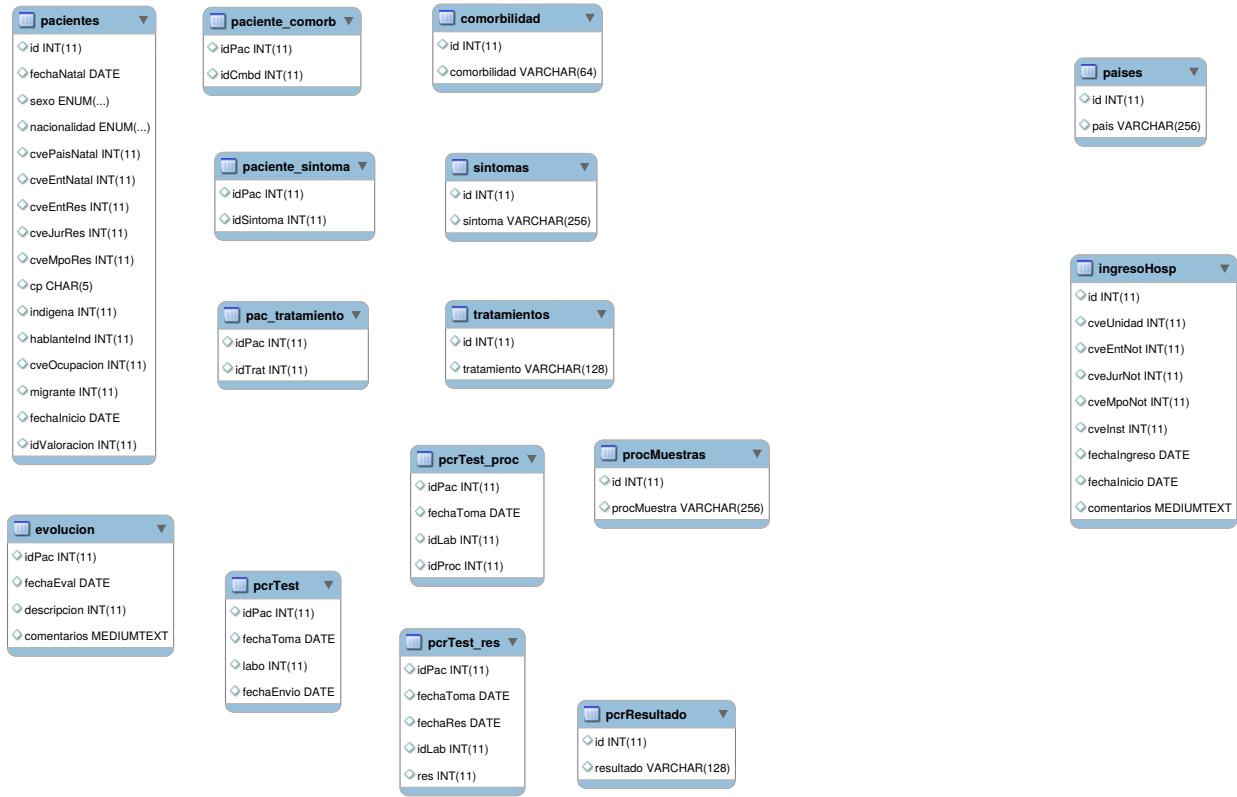


Figura 1: Estructura de las tablas y sus propiedades

- Se dividió y organizó la información en tablas usando scripts desarrollados por el grupo de trabajo.
- Se programó una base de datos usando el motor para bases de datos MySQL.

### Métodos a desarrollar.

- Sistema de transacción de datos que permita compartir la información entre IIMAS e IBT
- Desarrollo de consultas
- Desarrollo de interface de consultas

La base de datos será almacenada en 2 servidores. IBT en Cuernavaca, IIMAS Ciudad de México.

## 2. Análisis estadísticos

**Integrantes:** Igor Barahona (IM-Cuernavaca), Carlos Erwin Rodríguez (IIMAS) y Claudia Juárez (IIMAS).

**Objetivos:** Utilizar datos de la CDMX para:

- Construir un modelo que permita predecir el número de pacientes infectados por COVID-19 distinguiendo: unidad de salud, sector (IMSS, ISSSTE, PRIVADOS, etc) y delegación, proporcionando los intervalos de confianza asociados con esta predicción.
- Estimar el número de días que transcurren desde el inicio de síntomas de un paciente infectado por COVID-19 hasta que el paciente se recupera o fallece. Identificar si esta estimación varía con: sexo, edad, y/o comorbilidades. Concretamente, estimar:
  1. El tiempo en días que tarda un paciente en asistir a la clínica desde el inicio de los síntomas (ver la Figura 2).
  2. El tiempo en días que transcurre entre que el paciente ingresa a la clínica y la resolución del caso.

### **Avance (20 %):**

#### **Resultados parciales**

- Se realizó un programa que de forma automática obtiene la evolución del número de pacientes infectados por COVID-19 en la CDMX por unidad de salud, sector (IMSS, ISSSTE, PRIVADOS, etc) y delegación. Es indispensable tener bases de datos (remesas) en un formato pre-establecido y que no cambie con el tiempo (ver Cuadro 1).
- Se investigó la asociación estadística entre **edad y sexo** y un resultado positivo a COVID-19. El punto de partida fueron 1,414 individuos diagnosticados en la Ciudad de México, quienes manifestaron síntomas entre el 15 de enero y 02 de abril de 2020. Los resultados sugieren que dos variables demográficas: EDAD y SEXO tienen un impacto significativo en el diagnóstico final. Siendo el grupo de hombres **mayores de 40 años**, el que muestra una mayor probabilidad de dar positivo a COVID-19 (ver Figura 3).
- Se realizó un análisis exploratorio de las variables de interés. Entre ellas, la relación entre la fecha de inicio de síntomas, el ingreso y la defunción (ver Figura 4), así como la distribución de los casos positivos por ocupación (ver Figura 5) en la cual se detectó que los trabajadores de la salud, médicos y enfermeras representan al 18.19 % del total de infectados al 11/04/2020.

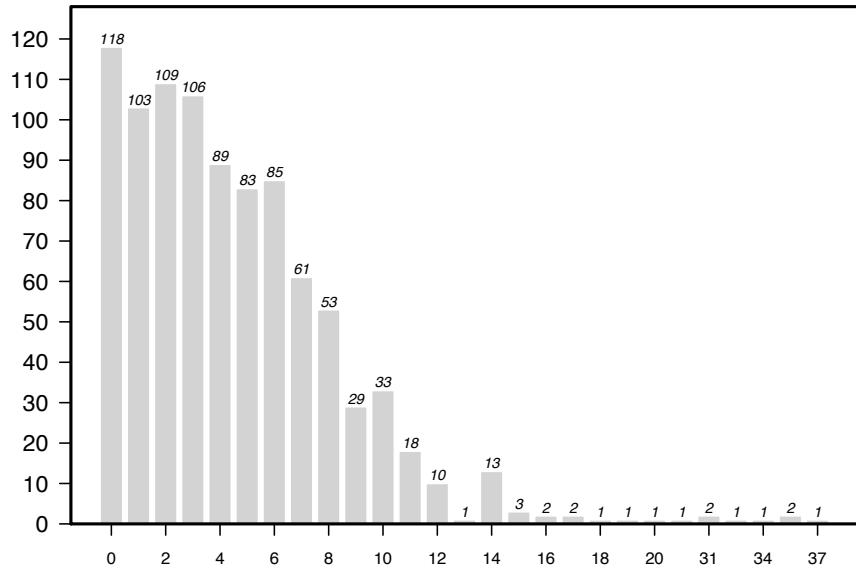


Figura 2: Distribución del número de días entre el inicio de los síntomas y que el paciente asiste a la unidad de salud.

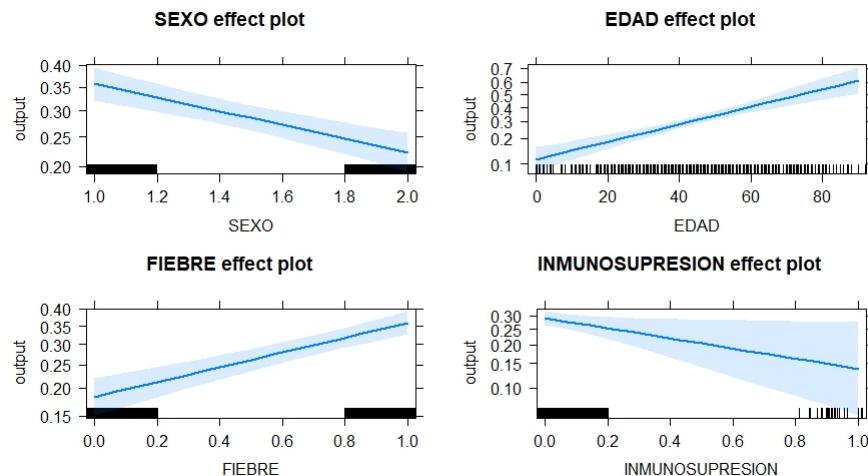


Figura 3: Efectos principales de los síntomas significativos.

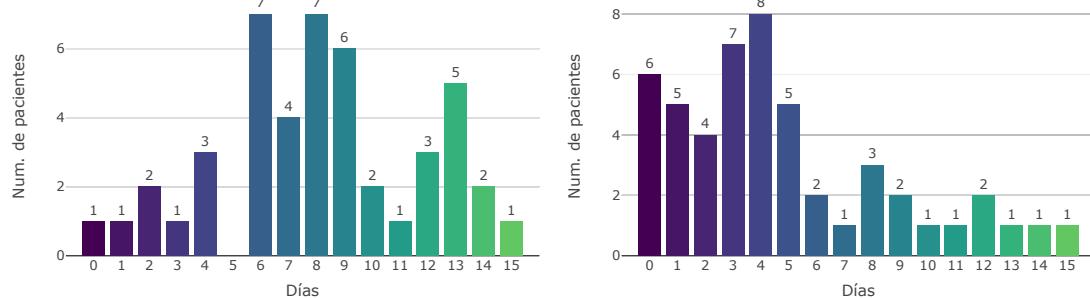


Figura 4: Distribución del número de días entre el inicio de síntomas y la defunción (izquierda) y el número de días entre en ingreso y la defunción al 11/04/2020.

Sector	Abril, 2	Abril, 3	Abril, 4	Abril, 5	Abril, 7	Abril, 8
SSA	167	212	242	277	300	376
IMSS	28	40	60	78	266	263
PRIVADA	117	156	158	158	205	220
ISSSTE	14	14	14	33	34	54
PEMEX	2	7	7	7	7	7
SEDENA	6	6	6	7	7	7
SEMAR	1	1	1	1	1	1
UNIVERSITARIO	0	0	0	0	1	1
CRUZ ROJA	0	0	0	0	0	0
ESTATAL	0	0	0	0	0	0
IMSS-OPORTUNIDADES	0	0	0	0	0	0

Cuadro 1: Evolución del número de infectados por COVID-19 por Sector

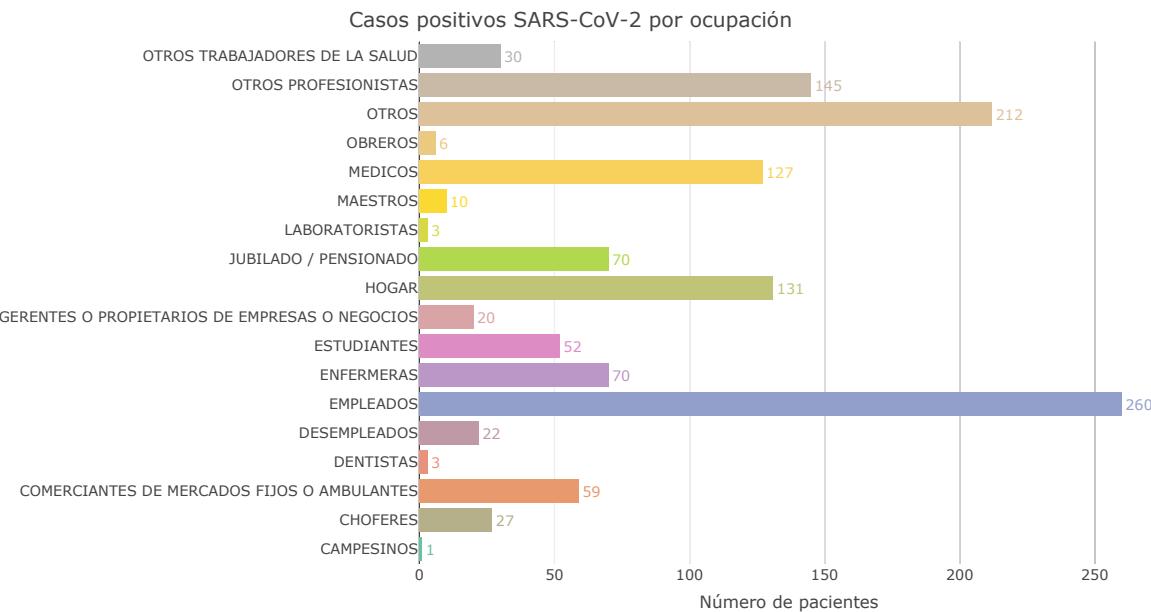


Figura 5: Distribución de pacientes infectados por ocupación al 11/04/2020.

### 3. Modelos epidemiológicos

#### 3.1. Incertidumbre y planeación

**Integrantes:** Isaac Pérez Castillo (IF), Ramsés Mena (IIMAS), Jorge X. Velasco (IM), Luis Benet (ICF), Denis Boyer (IF), Natalia Mantilla (FC), Lukas Nellen Filla (ICN), Carlos Nataren (IF), y Octavio Valenzuela (IA).

**Objetivos:** Parametrizar modelos epidemiológicos basados en el registro de casos y fallecimientos confirmados por SARS-CoV2. En una segunda etapa, incorporar factores demográficos, geográficos y económicos de la población. En forma central se usa estadística Bayesiana avanzada para medir la incertidumbre inherente a estos modelos y diseñar estrategias que contribuyan a evitar el rebase de las capacidades del sistema de salud, contemplando el peor escenario.

## Avance (40 %):

**Resultados:** El primer modelo se tiene listo y se están ultimando los detalles en un reporte técnico y la generación de un nota pública y presentación (**Avance 95 %**). Este es un modelo epidemiológico donde se introducen los siguientes compartimentos con el fin de retratar el proceso infeccioso y estimar necesidades del sistema de salud: Susceptible, Expuesto, Infectado, Hospitalizado, Crítico, Recuperado, Fallecido (Deceased en inglés) que siguen el siguiente conjunto de ecuaciones diferenciales:

$$\begin{aligned}
 \frac{dS(t)}{dt} &= -\beta(t) \frac{S(t)I(t)}{N}, \\
 \frac{dE(t)}{dt} &= \beta(t) \frac{S(t)I(t)}{N} - \frac{E(t)}{\tau_\ell}, \\
 \frac{dI(t)}{dt} &= \frac{E(t)}{\tau_\ell} - \frac{I(t)}{\tau_i}, \\
 \frac{dH(t)}{dt} &= (1-m) \frac{I(t)}{\tau_i} + (1-f) \frac{C(t)}{\tau_c} - \frac{H(t)}{\tau_h}, \\
 \frac{dC(t)}{dt} &= c \frac{H(t)}{\tau_h} - \frac{C(t)}{\tau_c}, \\
 \frac{dR(t)}{dt} &= m \frac{I(t)}{\tau_i} + (1-c) \frac{H(t)}{\tau_h}, \\
 \frac{dD(t)}{dt} &= f \frac{C(t)}{\tau_c},
 \end{aligned}$$

donde  $\beta$ ,  $\tau_i$ ,  $\tau_\ell$ ,  $\tau_h$ ,  $\tau_c$ ,  $m$ ,  $c$  y  $f$  son parámetros que tienen se ajustan con los datos. El resultado de la calibración y su correspondiente contraste con datos reales para México se puede observar en la Fig. 6. Una vez calibrado el modelo, se le utiliza como herramienta predictiva. Con

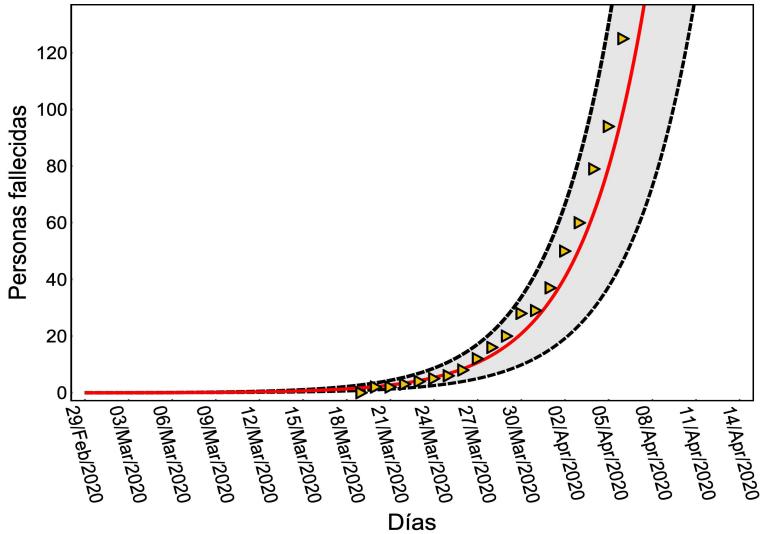


Figura 6: Modelo (línea roja), datos (símbolos), incertidumbre (líneas punteadas).

estadística Bayesiana se predice tanto la evolución típica de pacientes en estado crítico, como la incertidumbre en las predicciones que proviene de la falta de certeza en los parámetros del

modelo. Esta incertidumbre nos motiva a recomendar considerar el peor escenario posible de la epidemia para definir la fecha de implementación de medidas de contención. En la Fig. 7 se puede observar una predicción del modelo junto con el intervalo de confianza para cada momento.

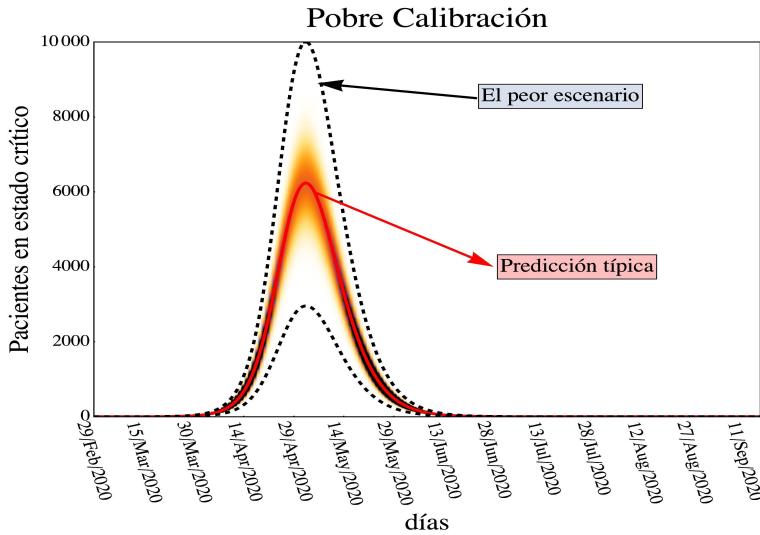


Figura 7: Predicción para pacientes críticos.

Finalmente, estudiando la evolución de pacientes críticos en el peor escenario se plantean planes de mitigación, para aplanar la curva, para que el número de pacientes críticos nunca exceda la capacidad de nuestros hospitales. El resultado se resume en la Fig. 10, teniendo en cuenta la pobreza de datos y/o la sensibilidad del modelo.

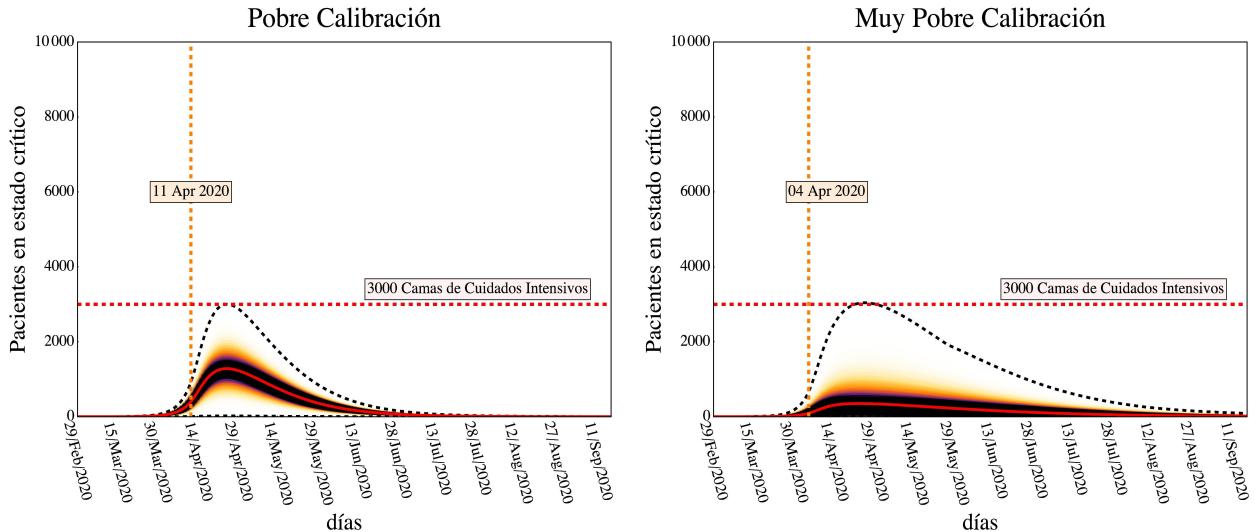


Figura 8: Día de activación para que planes de mitigación sean exitosos.

Los modelos más realistas ya están planteados y estamos en la fase de implementación para empezar a hacer simulaciones, primero con datos sintéticos (para controlar y entender el modelo) y luego con datos reales de la epidemia.

### 3.2. Evaluación de políticas de control

**Integrantes:** Natalia Mantilla (FC), Ramsés Mena (IIMAS), Carlos Rodríguez (IIMAS), Marco Herrera (FC) (confirmar, ampliar).

#### Objetivos:

1. Evaluar el impacto de las medidas de contención en México. Para ello, comparar el número de casos esperado de acuerdo con un modelo SEIR parametrizado para Covid-19 con la estimación de los casos existentes basada en el registro de muertes.
2. Investigar la relación entre estrategias de muestreo e incertidumbre en los datos para distintos países.

**Planteamiento:** Dada la existencia de casos subclínicos y asintomáticos de Covid-19, así como la escasez y alto costo de las pruebas de laboratorio que permiten confirmar el diagnóstico, el número de casos confirmados de Covid-19 representa apenas una fracción de los casos existentes. Por otro lado, las muertes que ocasiona esta enfermedad suelen registrarse y diagnosticarse con cierta fidelidad en estadísticas oficiales. Además, contamos con estimaciones de la tasa de letalidad del Covid-19 ( $\rho$ ), así como del tiempo que transcurre entre la presentación de síntomas y el deceso ( $\tau_d$ ) para distintas poblaciones. Esto nos permite estimar una serie de tiempo de los casos de Covid-19 en México y en otros países (con un intervalo de confianza) a partir del registro de las muertes que ha ocurrido en cada lugar. A su vez, da una estimación del momento de inicio del brote en cada población( $t_0^i$ ). Lo anterior sirve para:

1. Hacer un contraste de la evolución que muestran los datos con predicciones teóricas de modelos epidemiológicos básicos y evaluar el éxito de las medidas de contención.
2. Estimar la fracción de casos que está representada en el registro de casos confirmados para la población  $i$ ,  $\xi_i$ , dando un intervalo de confianza para este dato, y compararla por el número de pruebas de laboratorio por millón de habitantes  $P_i$ . Este contraste nos permite conocer qué grado de incertidumbre se asocia a cada  $P_i$ .

**Refinamiento:** Utilizar datos sociodemográficos para calcular la tasa de letalidad esperada en México (y de ser posible, en otros países). Hacer también estimaciones locales de parámetros como  $\tau_d$  y  $R_0$  para afinar las proyecciones teóricas.

## 4. Modelado de la pandemia por medio de un análisis estadístico de datos

**Integrantes:** Fabio De Colle (ICN), Carlos del Burgo (INAOE), Natalia Mantilla (FC), Ramsés H. Mena (IIMAS).

#### Avance (25 %)

**Objetivos:** Los modelos epidemiológicos dependen de parámetros que se pueden estimar utilizando los registros disponibles de casos, decesos y/o recuperados. Pero, siendo que el número de infectados es largamente sub-estimado, es muy difícil calibrar estos modelos. En este estudio,

nos proponemos analizar los datos que muestran la evolución de la pandemia como función del tiempo en varios países, y usarlos como referencia para predecir el comportamiento de las distintas fases de evolución de la pandemia para países que aún están en la fase de crecimiento exponencial pero que han implementado medidas de aislamiento (por ejemplo, México). Un ejemplo de los resultados se muestra en la Figura 9, que representa (en verde) la estimación del número de personas contagiadas como función del tiempo (siendo el día 0 el día de hoy), y en morado el comportamiento futuro asumiendo una evolución consistente con la de los países que han implementado medidas de aislamiento social.

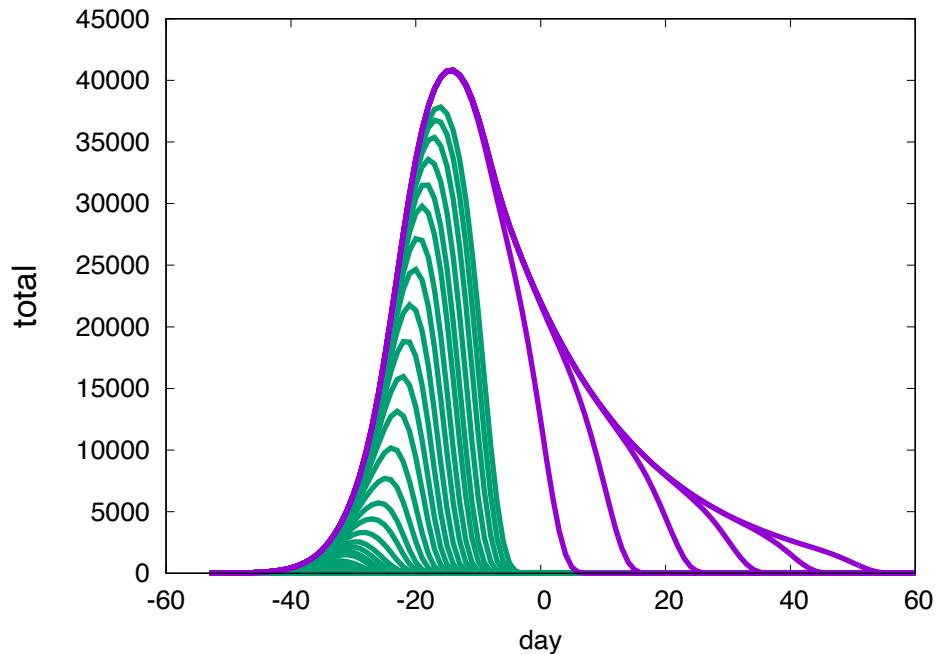


Figura 9: Estimación de contagiados como función del tiempo en España. El día de hoy esta representado como 0.

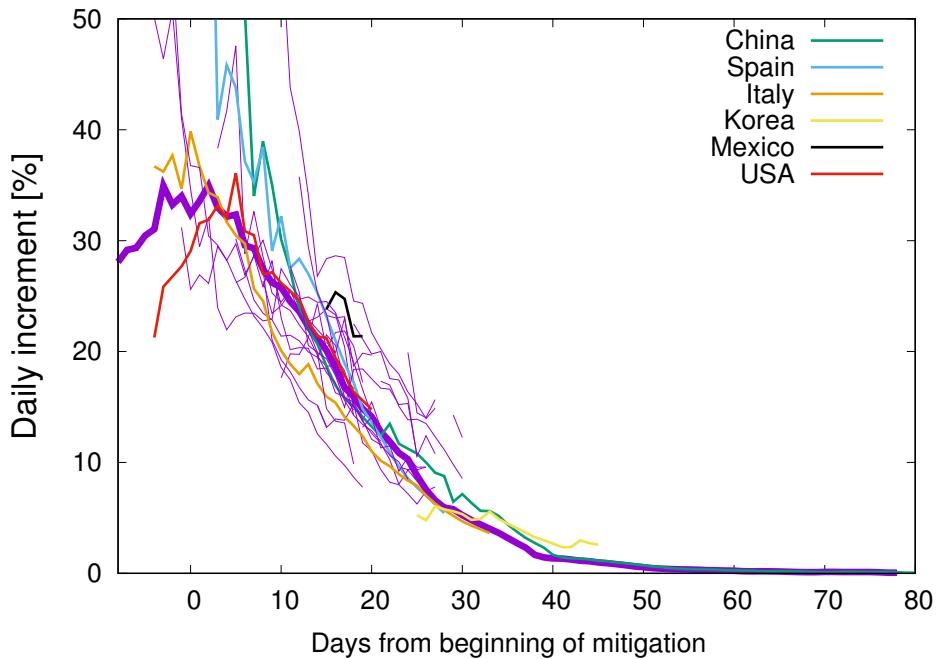


Figura 10: Tendencia de fallecimientos diarios desde el inicio de las medidas de aislamiento social para países con más de 100 fallecidos.

## 5. Estimaciones y modelaje estocástico de dinámica de propagación en una población para distintos grupos de riesgo

**Participantes.** Carlos I. Herrera-Nolasco (UNAM), Eugenia O'Reilly-Regueiro (IM), Sergio Iván López Ortega (FC), Marco Arieli Herrera-Valdez (FC). Este trabajo es parte de un reporte técnico que está siendo editado para su revisión por pares. Se pueden encontrar más detalles y referencias en <https://scab-unam.github.io/tsamCOVID-19/>.

**Objetivo.** Detectar similitudes macroscópicas en el comportamiento de las defunciones por COVID-19 con otros países.

Sabemos que factores como la fatalidad entre casos confirmados puede depender de una diversidad de factores como la calidad de los servicios de salud, y el acceso a dichos servicios, entre otros. Sin embargo, es posible observar algunas similitudes macroscópicas en el comportamiento de las defunciones por COVID-19 en lugares que podría pensarse que los fallecimientos presentarían comportamientos muy distintos, como China y Corea del Sur por un lado, e Italia, por otro. En Corea del Sur y China, el control es muy estricto y la cuantificación de casos ha sido masiva. En cambio en Italia, donde la estructura poblacional es distinta, ha habido más fallecimientos por COVID19, y ha sido rebasado el sistema de salud al grado de tener que ne-

gar el uso de respiradores a la gente. Sin embargo, en los tres países el cociente de fatalidad de casos (CFR por sus siglas en inglés) tiene muchas similitudes (Fig. 11, panel izquierdo), que se pueden explotar para obtener las contribuciones relativas de los fallecimientos por COVID-19 (Fig. 11, panel derecho) por cada grupo de edad al total de muertes observadas, y usar esas estimaciones.

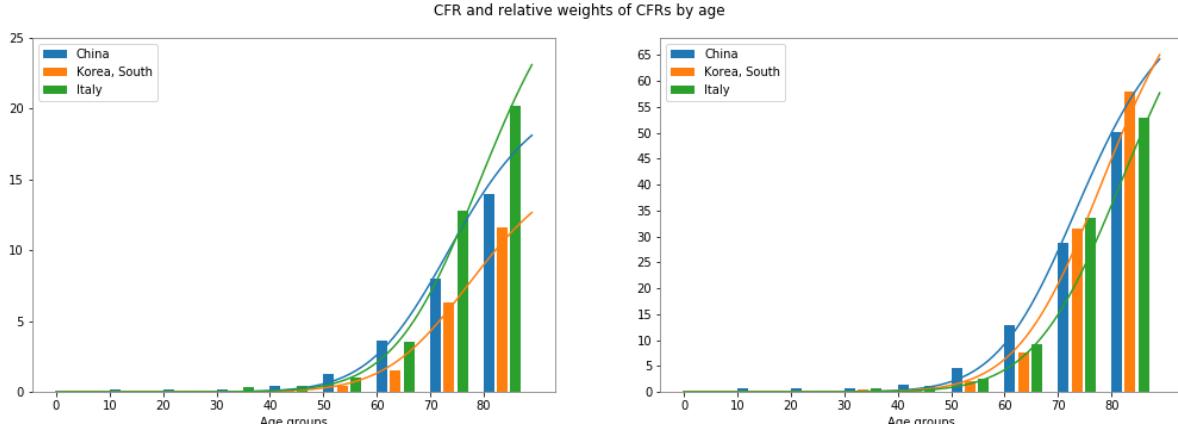


Figura 11: CFR para distintos grupos de edad usando datos sobre fatalidad de COVID-19 de China, Korea del Sur, e Italia. El panel izquierdo muestra los CFRs de los tres países. El panel derecho, muestra las contribuciones a las muertes inferidas a partir de los CFRs.

Por su similitud, es posible usar los pesos relativos de las defunciones de los distintos grupos de edad y hacer distintas predicciones para el caso de México.

Las proyecciones obtenidas usando los datos de los distintos países son similares. Hay que tomar en cuenta que estos datos no han sido ajustados con respecto a subreporte. Por ejemplo, sin ajustar los datos por subreporte, la estimación del 11 de abril de 2020 es de alrededor 150 muertes de adultos de más de 70 años en México. Usando un factor de ajuste por subreporte igual a 10, estaríamos estimando que el número de muertes por COVID-19 en México sería aproximadamente 1500.

Es importante mencionar que estas estimaciones no toman en cuenta la estructura poblacional en México, o la de los países tomados para el análisis, pero esa información está implícita en los datos de dichos países.

## Trabajo en progreso.

Estamos construyendo modelos estocásticos de dinámica de propagación basados en estimaciones cualitativas similares a las descritas arriba, con la intención de explicar mecanismos subyacentes a la propagación de la epidemia.

## 6. Análisis de marcadores de riesgo

**Integrantes:** Fabián García Nocetti (IIMAS), Gibran Fuentes Pineda (IIMAS), Blanca Hilda Vázquez Gómez (IIMAS)

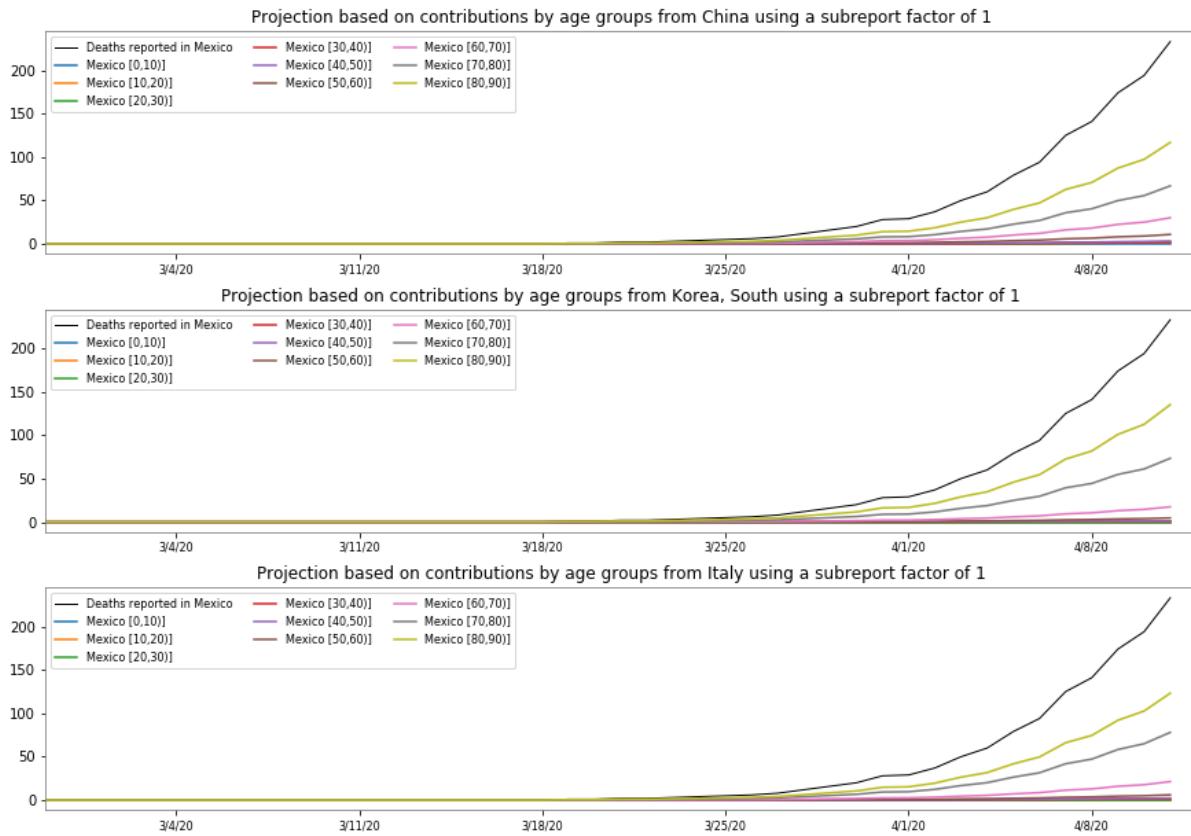


Figura 12: Estimaciones de fatalidad para México tomando en cuenta grupos de edad, con los datos de China, Korea del Sur, e Italia, calculado hasta el 11 de abril de 2020, y usando un factor de ajuste por subreporte igual a 1.

**Objetivos:** Entrenar y evaluar modelos basados en árboles de potenciación del gradiente para la predicción de mortalidad y de evolución clínica en pacientes contagiados con COVID-19 e identificar marcadores de riesgo dentro de las variables disponibles.

**Avance (20 %):**

**Resultados parciales:**

- Se identificaron marcadores preliminares de riesgo de mortalidad en pacientes confirmados con COVID-19 (ver figura 13). Los principales marcadores identificados son dificultad respiratoria, presencia de neumonía, edad avanzada, género masculino, ser ingresado a la UCI, obesidad, cianosis, fiebre e intubación. Lo que indican los marcadores es que la presencia de éstos incrementan la probabilidad de que el paciente fallezca. Llama la atención que uno de los marcadores de riesgo identificados es ser atendido en el IMSS. Sin embargo, cabe resaltar que la identificación de estos marcadores puede verse afectada por la baja cantidad de decesos en la base de datos (23 pacientes de un total de 2,663), de tal forma que los marcadores podrían no reflejar la realidad. Para la identificación de

estos marcadores se generó un modelo de predicción de mortalidad que obtuvo un área bajo de curva ROC de 0.89 (ver figura 14).

- Además de identificar marcadores de riesgo, se analizaron las interacciones entre ellos. En la figura 15 se observa que los pacientes mayores a 40 años tienen un mayor riesgo de muerte (valores SHAP mayores a 0). En particular, el grupo de mayor riesgo se caracteriza por presentar dificultad respiratoria, neumonía y padecer hipertensión y diabetes. Además de estos resultados preliminares se están analizando otros escenarios para detectar nuevos marcadores de riesgo.

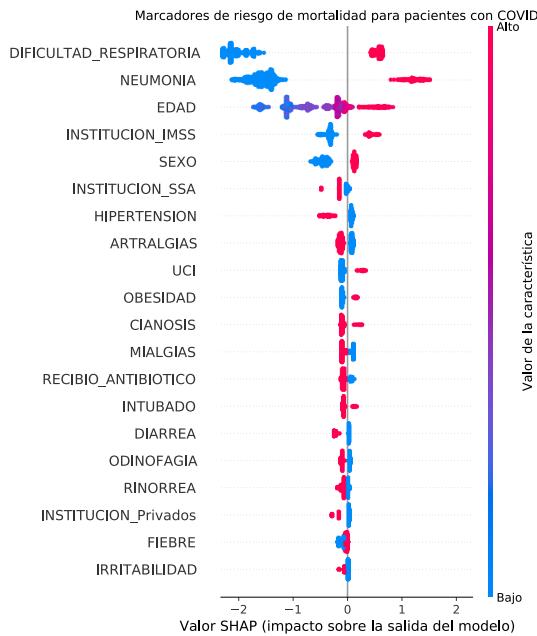


Figura 13: Marcadores de riesgo de mortalidad. La distribución de los marcadores se muestra en orden descendente, mientras que la ubicación horizontal muestra si el marcador influye en que la predicción sea de alto o bajo riesgo. Finalmente, el color indica si el valor de la variable es alta (color rojo) o baja (color azul) para esa observación.

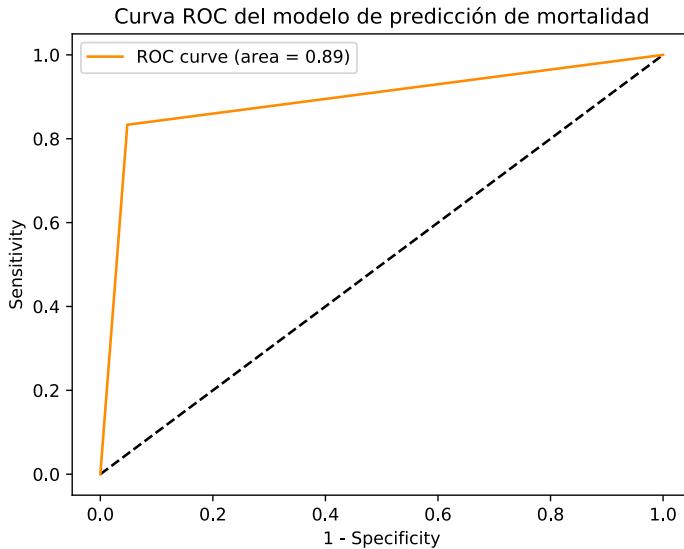


Figura 14: Curva ROC del modelo de predicción de mortalidad

## 8. Modelado matemático en pacientes con COVID-19

**Objetivos:** Desarrollar varios modelos matemáticos que se proponen para representar la dinámica del SARS-CoV-2 en pacientes infectados.

**Integrantes:** EA Hernandez-Vargas (IM-UJ), Jorge X. Velasco Hernández (IM-UJ)

**Avances:** Teniendo en cuenta los diferentes tiempos de inicio de la infección, se calculan conjuntos de parámetros que representan la infectividad del SARS-CoV-2 y se comparan con otras infecciones virales que también pueden causar pandemias. Uno de los modelos señala que el tiempo de infección del SARS-CoV-2 entre las células susceptibles (media de 30 días aproximadamente) es mucho más lento que el reportado para el Ébola (unas 3 veces más lento) y la gripe (60 veces más lento). El número de reproducciones dentro del huésped para el SARS-CoV-2 es coherente con los valores de infección por influenza (1,7-5,35). El mejor modelo para ajustar los datos fue incluir las respuestas inmunitarias, que sugieren una respuesta celular lenta que alcanza su punto máximo entre 5 y 10 días después de la aparición de los síntomas. Todos los modelos propuestos predicen que el SARS-CoV-2 puede replicarse muy lentamente en los primeros días después de la infección, y podría estar por debajo de los niveles de detección durante los primeros 4 días posteriores a la infección (ver Figura 19).

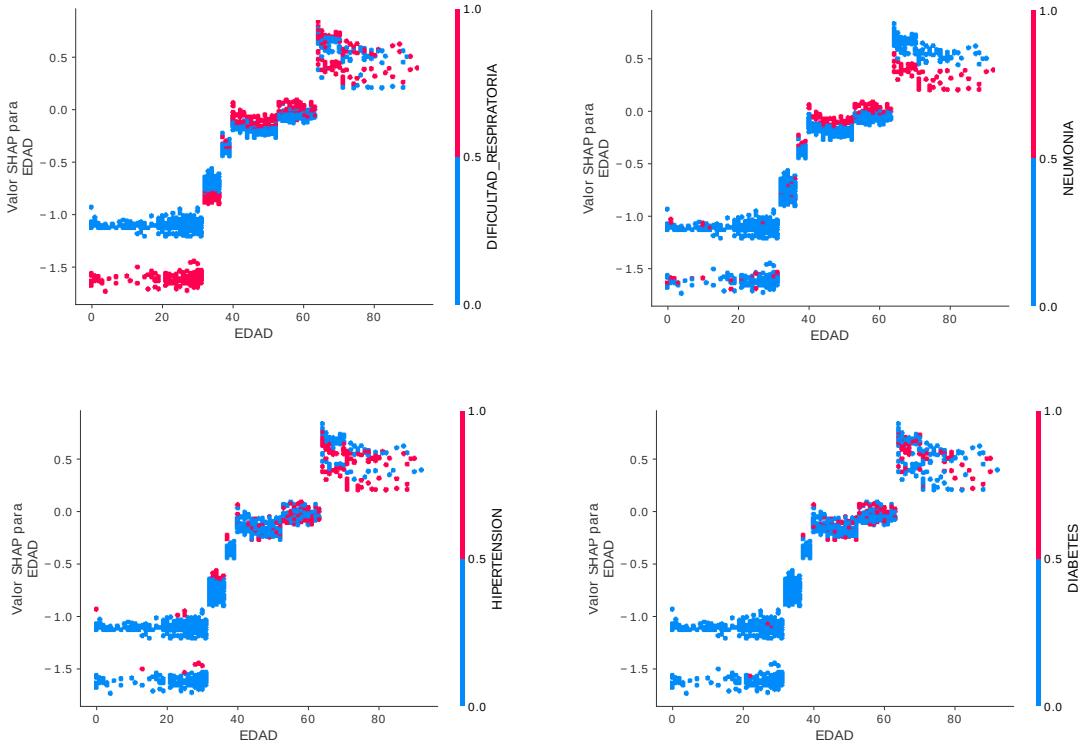


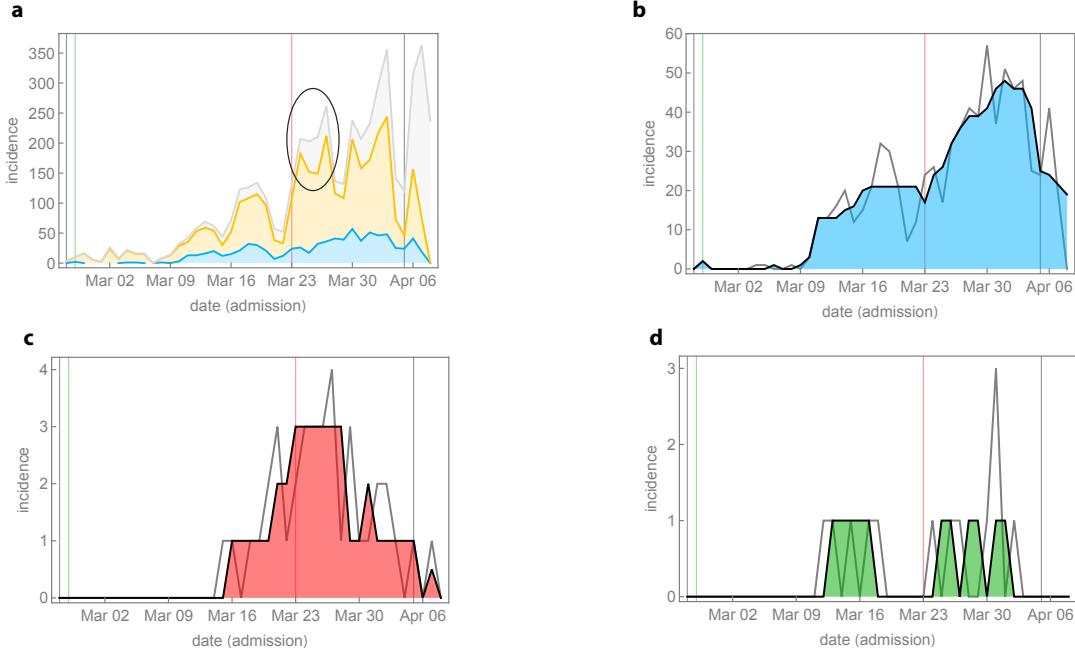
Figura 15: Interacciones entre marcadores que incrementan el riesgo de muerte en pacientes con COVID-19, donde cada punto es un paciente, el eje  $x$  describe una característica clínica y el eje  $y$  es el valor SHAP atribuido a esta característica. Un valor SHAP más alto representa un mayor riesgo de muerte debido a esa característica. Los diferentes colores representan un efecto de interacción entre las características.

## 9. Estudio de viabilidad para la estimación de escenarios en México para la mitigación de muerte por desbalance entre demanda y capacidad por la epidemia causada por el virus SARS-CoV-2

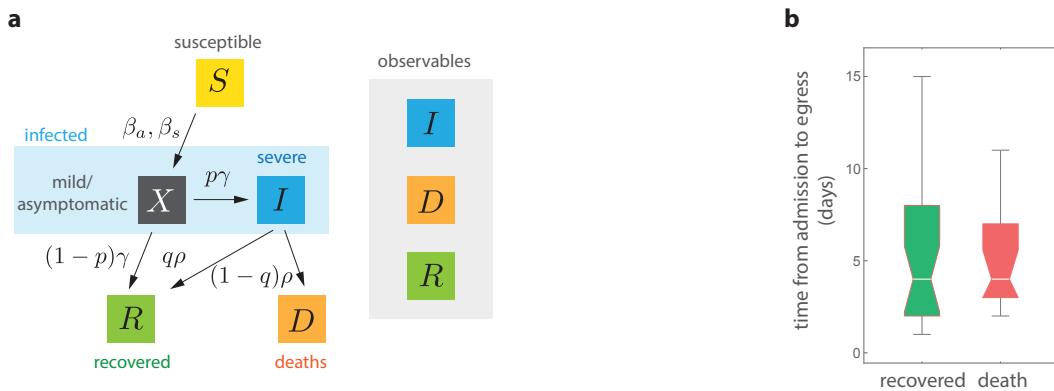
**Integrantes:** Gilberto Calvillo (IM-UC), Jesus Lopez (IM-UC), Jorge X. Velasco (IM), Erick Treviño (IM-UC).

**Objetivos:** Simulacion de escenarios en mitigación de casos de muerte por desbalance entre capacidad de unidades de atención especializada en hospitales y la demanda por casos de COVID-19.

**Avances:** Se presentan algunas consideraciones para la implementación de la metodología expuesta en [10] el cual pone a disposición de manera libre el código R en [<https://github.com/nhsbnssg-analytics/covid-simr-hospital-application>], misma que se utiliza aquí para obtener las Figuras 21 y 22



**Figura 16: Incidence of cases in Mexico City.** In the Figure, the green line on February 27 denotes the first COVID-19 positive case detected in Mexico City. The red line on March 23 marks the date in which the epidemic mitigation measures started. **a.** COVID-19 positive cases (blue), COVID-19 negative cases (yellow), and unknown cases (gray). Deep troughs are observed during weekends, and peaks the beginning and end of each week (see, e.g., marked ellipse). **b.** The original time series of COVID-19 positive cases (gray) and the filtered time series (black with blue filling). **c.** The original time series of COVID-19 positive deaths (gray) and the filtered time series (black with red filling). **d.** The original time series of COVID-19 positive recoveries (gray) and the filtered time series (black with green filling).



**Figura 17: A model to estimate the number of hidden infected individuals in Mexico City.** **a.** Block diagram for the population dynamics model. The observables available by the Sentinel surveillance system are marked in the gray box. **b.** Time difference from the date of admission to the date of recovery or death, for all COVID-19 positive cases that were hospitalized. For both groups, the median is 4 days.

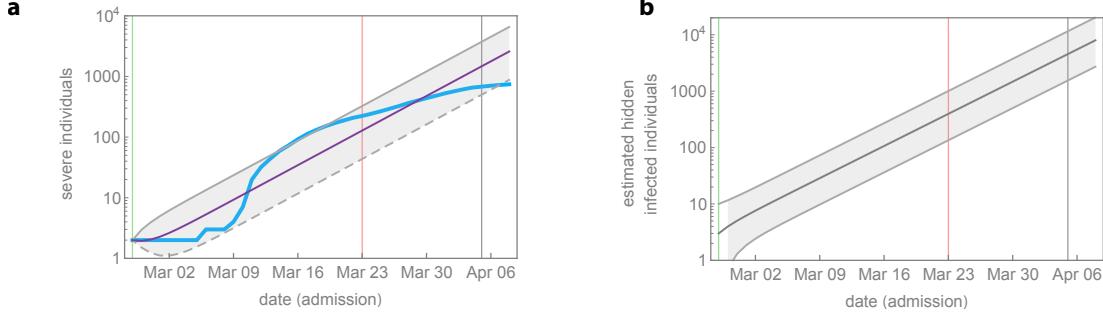


Figura 18: **Estimating the total number of COVID-19 infected in Mexico City.** **a.** The trajectory for the true number of infected individuals  $I(t)$  (blue). It is enveloped by two trajectories  $\hat{I}(t)$  produced by our model. **b.** Estimated number of infected individuals that are asymptomatic or have mild-symptoms using our model.

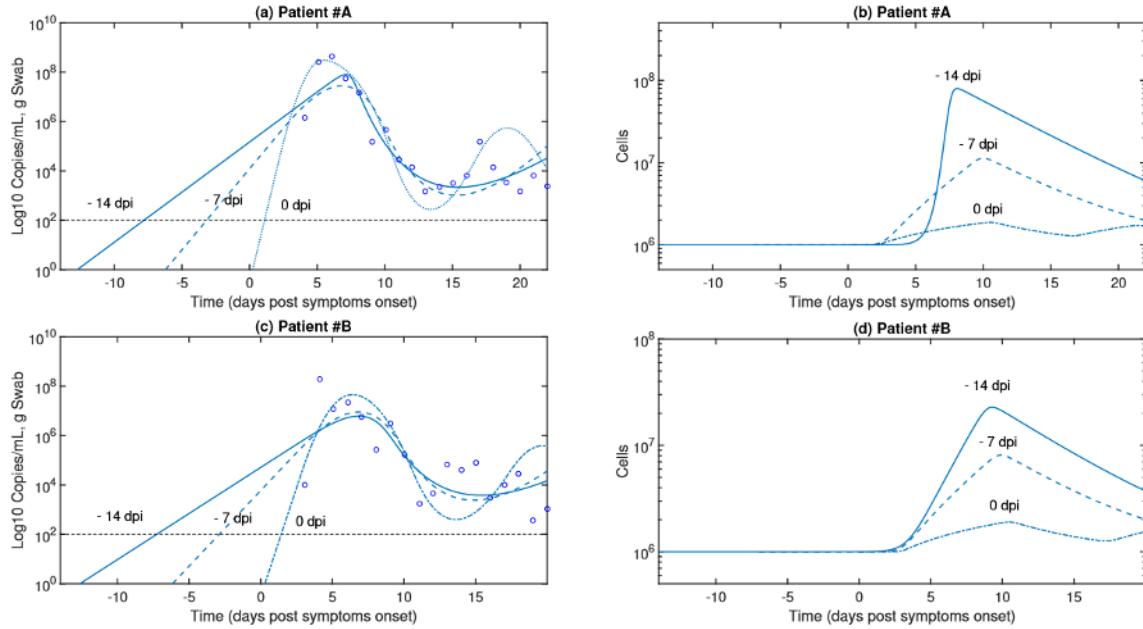


Figura 19: Modelo con respuestas inmunes para el SARS-CoV-2. Las líneas continuas son simulaciones. Los círculos azules representan los datos de pacientes con COVID-19. El tiempo de infección se consideró a los -14, -7 y 0 días después del inicio de los síntomas (dpi)

## El modelo de simulación

El modelo es una implementación de un sistema de colas operando con pérdidas (multi-channel queuing system operating with loss). Los pacientes se rechazan si no hay un “canal” (UCI Unidad de Cuidado Intensivo) disponible para ellos. Esto es en la notación de Kendall (1953):  $M(t)|G|C|C$  el cual se basa en un proceso de Poisson no-homogéneo. Para las conexiones perdidas, es decir admisiones hospitalarias rechazadas por falta de equipo, existe una probabilidad  $P_d$  de muerte. La implementación de [10] simula los siguientes eventos:

1. Llegada de pacientes requiriendo hospitalización (evento incondicional)

2. Admisión de paciente (evento condicional)
3. Paciente dado de alta (evento incondicional)
4. Admisión de paciente rechazada y paciente muere (evento condicional)
5. Admisión de paciente rechazada y paciente sobrevive (evento condicional)

## Los parámetros necesarios para la simulación

El sistema requiere la especificación de los siguientes parámetros:

- El modelo supone una distribución Gamma  $\Gamma(\alpha, \beta)$  para el número de días de hospitalización. Estimaciones realizadas en el artículo [2] determina  $\alpha = 8$  y  $\beta = 1$ .
- Capacidad en número de UCI's del hospital.
- Probabilidad  $P_d$  de **muerte por rechazo** dado que el paciente no es admitido. El artículo asume una probabilidad muy alta:  $P_d = 0,99$ . Lo hace en congruencia con lo documentado en la literatura en la que incluso para pacientes que son admitidos y atendidos la tasa es estimada entre 81 y 97 por ciento, c.f., [9].

## Requerimientos

1. Datos con dos columnas. Una columna de fechas y una segunda columna con el número de casos.
2. Diseñar escenarios.
3. Buena estimación de parámetros que va de la mano con el anterior punto.
4. Computo de alto rendimiento. En el ejemplo que se presenta más abajo, en un hospital pequeño (100 ICU's) con 100 repeticiones tarda 40 minutos y se recomiendan que sean al menos 1000 repeticiones y para hospitales grandes se deben considerar mas ICU's (en esta variable la complejidad aumenta mas allá de lo lineal).

El primer punto en relación a datos, es un requerimiento fuerte porque involucra “forecasting” de casos de infectados que se canalizan a un hospital. Por ello, la metodología requiere ser acompañada por el pronóstico de un modelo del estilo SIR. Sin embargo, este puede ser simple y en principio no requiere ninguna variante muy sofisticada en que se agregan nuevos estados, como por ejemplo en el modelo SEIR.

Además se debe considerar que se requiere de un procedimiento confiable para pasar de estimaciones que normalmente se hacen a escala nacional o estatal a un nivel mas desagregado. En general, tener la capacidad de agregar vs desagregar estimaciones.

No se puede dejar de mencionar que será necesario tener un mecanismo para pasar de la estimación del modelo SIR, particularmente la curva  $I$ , a una estimación de cuantos de estos casos se dirigen a un hospital. Por ejemplo, el Subsecretario Dr. Lopez Gatell mencionó que la mayoría de la gente no presentará síntomas. Por ejemplo, el porcentaje estimado de infectados que efectivamente llegan al hospital es aproximadamente 12 % (dado por el modelo centinela). De estos que llegan al hospital entre el 20 % y 30 % son graves y requieren una UCI. Esos son

los que deben ingresarse y los que provocarán el cuello de botella. Es decir aproximadamente entre el 2.4% y el 3.6% de los infectados requieren una UCI. Desde luego esto es dinámico y hay que ver si un modelo con llegadas de Poisson es adecuado. No obstante, la metodología aquí propuesta permite el explorar con escenarios. Por ejemplo al varia la tasa de infección.

## Ejemplo

Se calcula una curva de infectados del modelo SIR con las siguientes especificaciones. El sistema SIR está definido por el sistema de ecuaciones ordinarias

$$\begin{aligned}\frac{dS}{dt} &= -\beta S(t)I(t) \\ \frac{dI}{dt} &= \beta S(t)I(t) - \gamma I(t) \\ \frac{dR}{dt} &= \gamma I(t).\end{aligned}$$

Se generaron curvas para el sistema mediante solución numérica con parámetros

$$\beta = 0,1, \quad \gamma = 0,05, \quad S_0 = ,99, I_0 = ,1, R_0 = 0.$$

La Figura 20 presenta las curvas del modelo SIR obtenidas numéricamente.

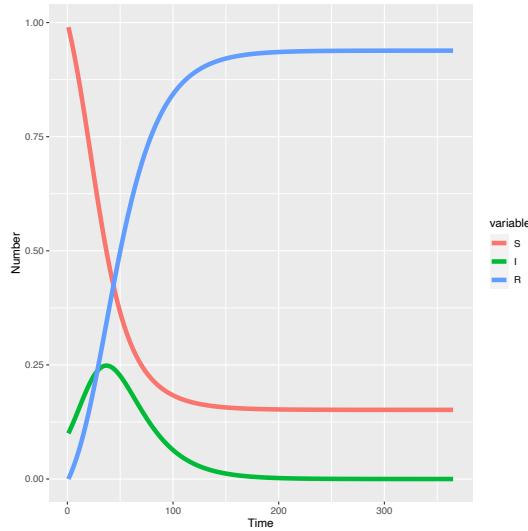


Figura 20: Curva SIR. Parámetros ficticios con la especificación del Apéndice ???. En rojo la curva  $S$  de Susceptibles, en verde la de Infectados  $I$ , en azul la de Recuperados.

Teniendo la curva  $I$  de infectados por el sistema SIR, **suponemos** que  $I$  será igual al número de casos que se recibirán en hospitales. Esta curva se escala con un factor (ahora mismo arbitrario) de  $1e3$  y se usa como input al modelo de simulación. Se fija la capacidad a  $1e2$  UCI's. El resultado de simulación se despliega en la Figura 21.

A este nivel de “Toy model” ya podemos dejar entrever el tipo de análisis que puede realizarse. Por ejemplo, la relación entre capacidad y muertes no es lineal y es sensible a la condición inicial. Esto lo vemos con una segunda simulación que en comparación con la anterior lo único

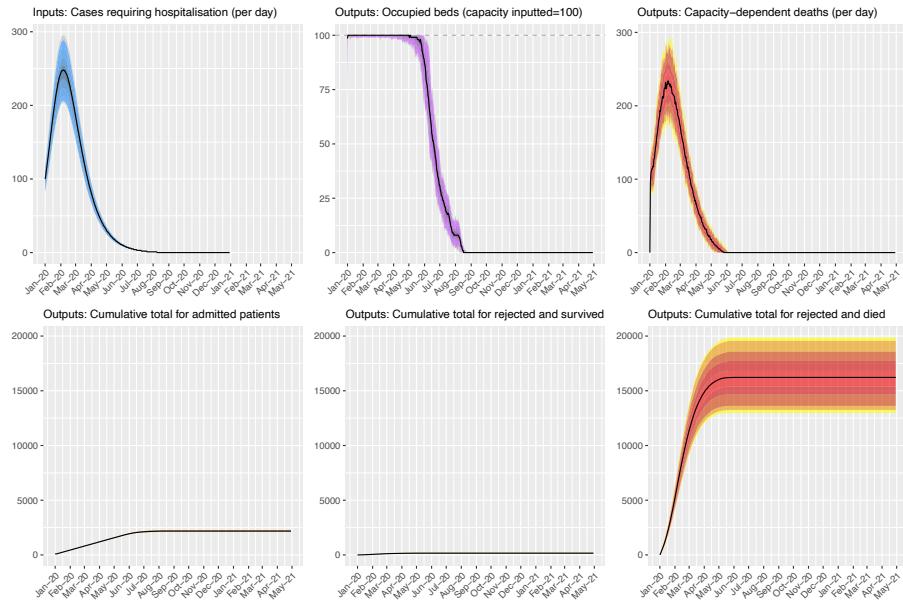


Figura 21: Curvas por simulación de casos recibidos y su atención. En la primera fila leyendo de izquierda a derecha vemos en el primer cuadro la curva de casos requiriendo hospitalización, a continuación la ocupación de camas, en la última las muertes dado que no hubo capacidad de recepción. En la segunda fila vemos casos acumulados, en el primer cuadro los pacientes admitidos, en el segundo los casos rechazados y que sobrevivieron, finalmente en el último cuadro se aprecian casos de defunción que no fueron admitidos por falta de capacidad.

que cambia es la capacidad que se ha asignado e igual a 200. En este caso la curva del número acumulado de casos de fallecimientos se contrae como se puede apreciar en la Figura 22. Sin embargo, en el caso concreto del ejemplo es notorio que de inicio estaba rebasado por mucho la capacidad contra la demanda y a pesar que se duplicó la capacidad no hubo una reducción significativa. Es interesante porque significa para el caso hipotético de este ejemplo, que se deben buscar otras estrategias de mitigación complementarias a simplemente duplicar capacidades. Hay que experimentar más.

## Necesidades para trabajo futuro

- Se requiere un inventario de hospitales así como ICU's disponibles.
- Se requiere comenzar a hacer pruebas de computo eficiente.
- Se requieren datos para estimar curvas SIR para contar con la curva de Infectados  $I$ . Idealmente se requiere contar con una base de datos de 4 columnas (fecha, casos, muertes y recuperados). Pues, con más información, mejores son las estimaciones de los parámetros  $\beta$  y  $\gamma$  en el modelo SIR.
- Sin embargo, se debe discutir este punto porque la curva  $I$  no necesariamente es la curva de gente que se dirige a un hospital y entonces será necesario calibrar. También será necesario ajustar para hacer concordar los niveles de (des)agregación entre las estimaciones de las curvas SIR y los inventarios de hospitales y UCI's.

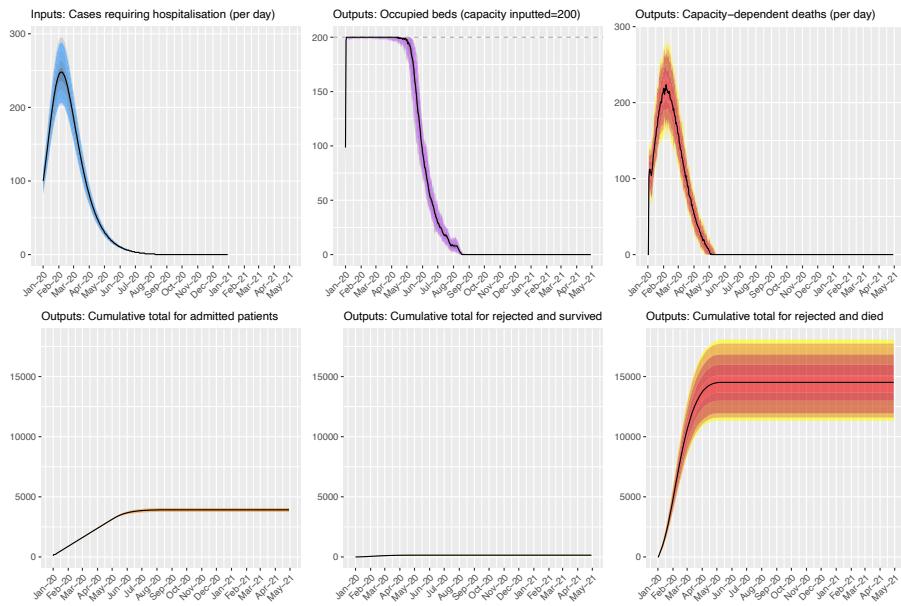


Figura 22: Curvas por simulación de casos recibidos y su atención. En la primera fila leyendo de izquierda a derecha vemos en el primer cuadro la curva de casos requiriendo hospitalización, a continuación la ocupación de camas, en la última las muertes dado que no hubo capacidad de recepción. En la segunda fila vemos casos acumulados, en el primer cuadro los pacientes admitidos, en el segundo los casos rechazados y que sobrevivieron, finalmente en el último cuadro se aprecian casos de defunción que no fueron admitidos por falta de capacidad.

- Ajustes de modelos SIR seguramente ya se estan desarrollando en otros grupos de trabajo y entonces habrá que coordinarse.
- También se puede adaptar con estimaciones del modelo de series de tiempo SIR (TSIR) [3] y la implementación en paquete tsIR de R.
- Se podría trabajar con los parámetros  $\alpha = 8$ ,  $\beta = 1$  para la distribución Gamma. Aun así sería interesante realizar estimación para el caso de México y para ello se requieren días de atención desde el ingreso de un paciente con Covid-19.

## 10. Detectando el impacto del transporte público sobre la transmisión del COVID-19 en la Ciudad de México.

**Integrantes:** Luis Álvarez Icaza (IIIng) , Jesús del Río (IIIng), María Lárraga (IIIng), Emmanuel Torres (IIIng), Jorge X. Velasco (IM).

**Objetivo:** Desarrollo de un modelo que reproduzca la dinámica espacio temporal de la propagación de COVID en la Ciudad de México y a la vez permita evaluar la efectividad de diferentes estrategias en México orientadas a retrasar la dispersión de la enfermedad derivadas del uso masivo del transporte público. Particularmente, se pone énfasis en los efectos derivados del uso de transporte (disminución) sobre la dispersión de la enfermedad que permitan establecer

estrategias de mitigación y su efectividad. Además, con el modelo se pretende también generar una herramienta para la detección de alto riesgo en la ciudad de México y determinar un índice que relacione estas zonas con factores de marginación relacionados con pobreza y servicios que permita detectar zonas de alta vulnerabilidad. De tal manera que con el trabajo realizado no se pretenden evaluar diferentes estrategias (o políticas) de control tanto desde el sector salud como desde el uso del transporte público.

#### **Avances:**

- Proponemos un modelo de propagación de la epidemia de COVID-19 basado en el uso del transporte público usando datos de la encuesta origen-destino de la Ciudad de México de INEGI 2017 y datos del uso de transporte público publicados MOOVIT y su cambio desde el 15 de enero a la fecha., para variar la cantidad de viajes entre 86 distritos de 16 delegaciones a lo largo del tiempo.
- Para este propósito se desarrollará un modelo híbrido basado en redes y en modelos SEIR (Susceptible, Expuesto, modificado y Recuperado) modificados para la consideración de escenarios de mitigación. De tal manera que la, la dinámica de infección toma en cuenta una visión de doble perspectiva del proceso de propagación de la epidemia de COVID-19; la cual involucra el contagio tanto en lugares específicos (como trabajos y hogares), como en el transporte público.
- Además, los viajes en el tiempo entre los diferentes distritos varían con base en los cambios de uso de transporte reportados para la Ciudad de México y no son estáticos. El modelo es discreto en tiempo y espacio y se basa en metapoblaciones para mantener la simplicidad del modelo.
- Particularmente, investigaremos la tasa de contribución del viaje mediante transporte público, en el proceso de propagación de la epidemia del individuo y evaluamos el impacto de la tasa de contribución del viaje de transporte público mediante la evaluación del volumen de personas infecciosas. El modelo se calibrará con datos de SEDESA y la DGE.
- Se usarán datos de la SEDESA para detectar zonas de riesgo, haciendo uso de herramientas de Google Maps, herramientas computacionales, y estadística. Con base en esta estimación se pretende determinar un índice de vulnerabilidad por distrito que tome en cuenta datos de marginación e infraestructura para la identificación de zonas vulnerables (quizás relacionar la efectividad de las estrategias con base en el número de camas de cuidado intensivo que serían necesarias de acuerdo con cada estrategia).

#### **Trabajo en progreso:**

- Actualmente, hemos descargado los datos de la SEDESA (casos confirmados) a través de Google Maps por delegación y C.P. Hemos logrado tener un mapa de riesgo actual para la Ciudad de México. Así como un mapa de calor para ver la propagación con base en el número de casos día a día.
- También hemos analizado datos de la encuesta origen destino para identificar los distritos en las delegaciones y los viajes que se realizan entre los mismos a o lo largo del día. Así, como la densidad poblacional de cada zona. A través de los cuales se han construido ya tablas de datos necesarias para la modelación.

- Tenemos ya un prototipo de modelo desarrollado, estamos en la fase de verificación del mismo. En el modelo cada nodo representa un distrito y cada arista entre dos distritos un número de viajes por día. Este número de viajes se cambia en el tiempo con base en datos y las estrategias de control bajo estudio. Se conjugan tres procesos infecciosos: los que suceden por viajes de ida y los que suceden durante el viaje de regreso, y los que suceden al interior de cada distrito considerando una población mezclada (varía en el día). Así se considera una interacción local, donde la población que se mueve por un medio de transporte puede contagiarse y una interacción global donde para cada distrito se considera una persona puede contagiarse al interior y moverse, cuya probabilidad de contagio es menor que para la interacción local.
- El modelo considera segmentación de la población por edades y sexo con finalidad de simular de mejor manera la dinámica de la enfermedad y la movilidad tanto al interior, como al exterior de los distritos.
- Lo que sigue es finalizar de detallar el modelo, la implementación y análisis.

## 11. COVID-19 forecasting

**Objective:** To present a description of the methodology used in [7] to reproduced the forecasting scenarios of number of infection cases in four countries: United States, Slovenia, Iran and Germany. We also use this method to predict possible scenarios of number of cases in Mexico and in particular in CDMX.

**Participants:** Nancy Leticia Gonzalez Morales (IM-UJ), Ruth Corona Moreno (IM-UJ), Jorge X. Velasco Hernández (IM-UJ)

The algorithm in [7] requires only the daily values of confirmed cases. Assuming the length of the data is  $n$  the average growth rate for the last  $m$  values in database according to [7] is determined by the equation (1)

$$G_\Delta = \frac{1}{m} \sum_{i=n-m}^{n-1} \left( \frac{x_i}{x_{i-1}} - 1 \right) \quad (1)$$

This rate  $G_\Delta$  is considered as a constant daily increase during the last  $m$  days. The forecast is made with equation (2) taking into account that after  $h$  days the majority of infected will recover, and that after  $d$  days a fraction  $p$  will die.

$$x_{i+1}^* = x_{i+1} - px_{i-d} - (1-p)x_{i-h} \quad (2)$$

These results are computed for thirty different average daily growth rates, starting from a  $G_\Delta$  increased and decreasing in equal intervals toward growth rate zero, so that  $x_{i+1}$  is defined as in equation (3) where  $i \in \{n-1, n, \dots, n+(m-1)\}$ ,  $q \in \{0, \dots, 29\}$  and  $M$  is the maximum growth rate.

$$x_{i+1} = x_i \left( 1 + (1 + G_\Delta)M - q \left( \frac{M}{30} \right) \right) \quad (3)$$

Parameter	Description	Value
$h$	Infectious period	$h \in [7, 14]$
$p$	Case fatality rate	0,04
$d$	Period of time until dead	21
$m$	Number of last days to estimate average growth rate	14

Cuadro 2: Parameter description

Although estimations of the values  $h$ ,  $d$ , and  $p$  for COVID-19 vary in literature, the authors in [7] pointed out that the goal is not to determine accurately but use them to evaluate possible forecasting scenarios. They also argued that sensible variations in  $h$ ,  $d$ , and  $p$  do not affect the forecast that significantly. Whereas, the key factor is the average growth rate  $G_\Delta$ , determined by equation (2).

### Important remarks

- Daily values of confirmed cases should be referred as cumulative confirmed cases. And prevalence can be also forecast using the algorithm in [7] since take into account two types of infected individuals, a proportion  $(1 - p)$  of those who will be recovered after  $h$  days and a proportion  $(1 - p)$  of those who will be recovered after  $d$  days. Thus, it is important to describe propoerly the type of predictions performed.
- The proportion  $p$  is referred as death rate, however, this value indicates the proportion of infected individuals who will die, hence this should be referred as the case fatality rate (CFR). Besides, the value for  $p$  in the paper is assumed to be 0.04, however CFRs vary widely between countries, from 0.2 % in Germany to 7.7 % in Italy.
- Databases used in [7] do not correspond to the total daily cumulative cases from the indicated date.
- Predictions can be adjust using values of average grow rate and cfr computed by 1 and recent cfr from data up to date. We present the computed estimations for these values using the equations.
- Some predictions in article [7] show decreasing behaviour. However, it is not possible for cumulative cases because the predictions are based on the variations of the maximum growth rate (a positive value) obtained from the algorithm and hence the number of cumulative cases should be in the best scenario to remain constant as it is shown in the plots we obtained.

## Forecasting Simulations

The algorithm in [7] can be used to forecast possible scenarios for both, the total cumulative cases and the prevalence of positive cases for an horizon in time which depends on the period of recovery. Therefore and due to the important remarks we present simulation for both predictions.

On the other hand, in predictions of the prevalence of positive cases, it is possible to observe decreases because parameters as death rate, days to heal and days to death must be considered

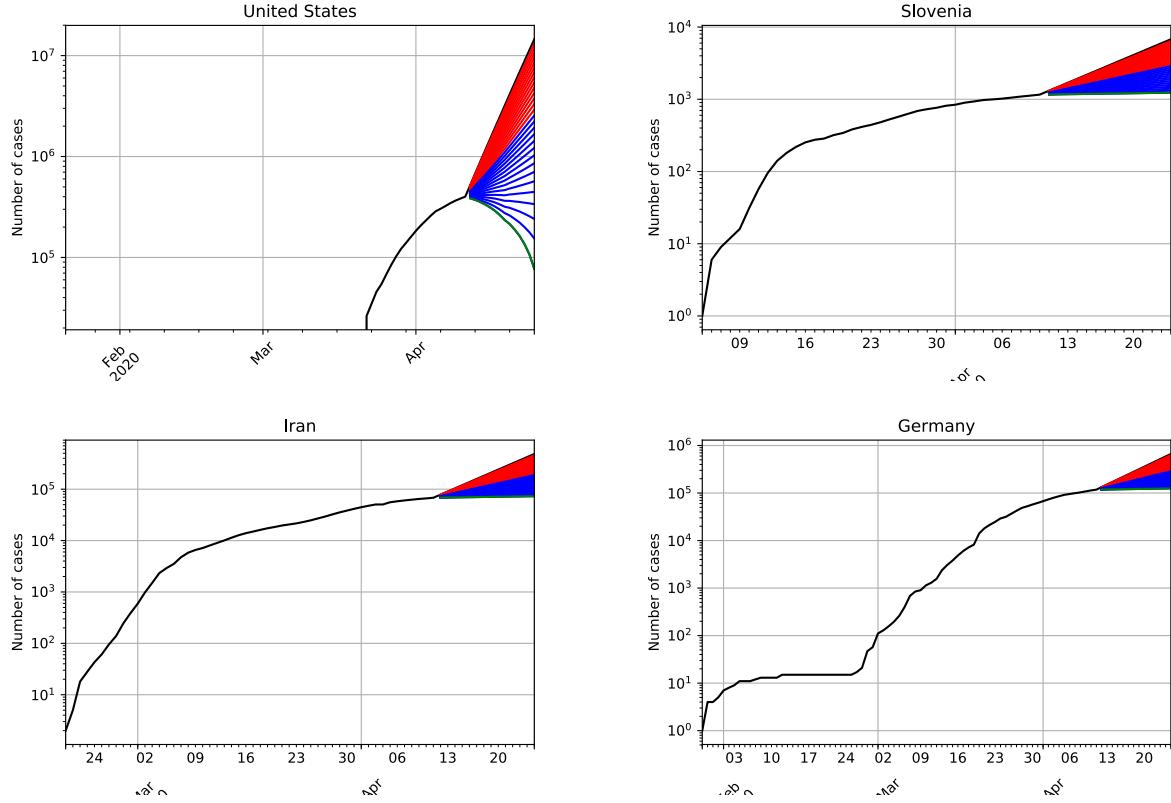


Figura 23: Forecasts of cumulative COVID-19 cases in United States, Slovenia, Iran and Germany. Black line denotes the actual data according to database of [? ]. The line that divides red lines and blue ones denotes the continuation of the trend if nothing would change. The uppermost solid red line denotes the prediction obtained if the maximal daily growth rate recorded during the past 14 days would increase by 20 %, while the lowermost line denotes the prediction if the daily growth rate would drop to zero, i.e. the number of cumulative cases up to the last recorded day remain constant.

to subtract from the total number of predicted cases as shown in following graphs where the values of these parameters are 0.04, 14 and 21 respectively as in [7].

It is important to notice that the prevalence forecasting takes the last  $d$  days since this is the time in which individuals are taken to die up to April 11th, 2020. Before, this shift prevalence are taken as zero. However, this assumption does not affect the predicted values since these are calculated with the last  $d$  days associated to the required approximate time for an infected individual dies.

## Forecasting simulations for Mexico and CDMX

In this section, we present forecasting simulations for both cumulative and prevalence cases in Mexico and CDMX. These predictions are performed using parameters as indicated in 2.

In the prevalence forecasts the reduction scenarios depends on the number of recovered and deceased people. Interpretation for this predictions requires to take into account the variations of the case fatality rates in Mexico since these values change over time and these are remarkable increasing due to preexisting health conditions in infected individuals.

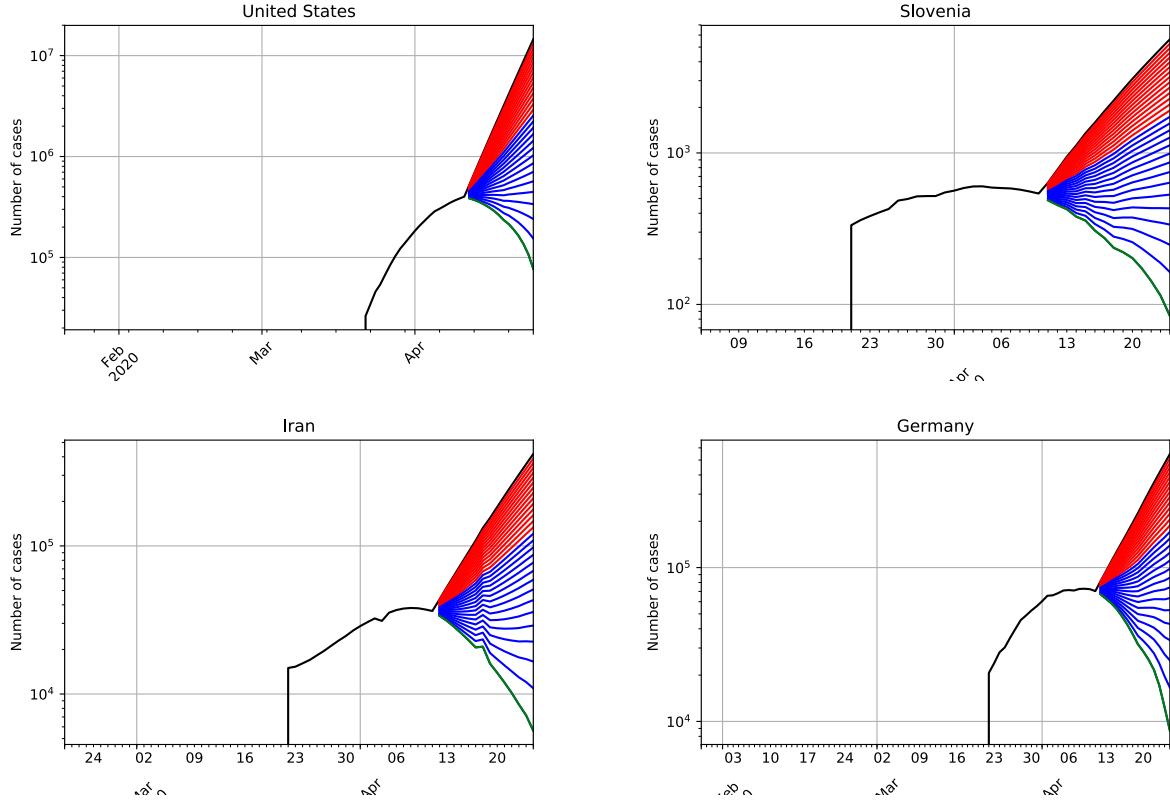


Figura 24: Forecasts of prevalence of positive COVID-19 cases after April 11th, 2020 up to 14 days later in United States, Slovenia, Iran and Germany. Black line denotes the actual data according to database of ECDC European centre for disease prevention and control. The line that divides red lines and blue ones denotes the continuation of the trend if nothing would change. The uppermost solid red line denotes the prediction obtained if the maximal daily growth rate recorded during the past 14 days would increase by 20 %, while the lowermost line denotes the prediction if the daily growth rate would drop to zero, i.e. the number of cases will tend to zero.

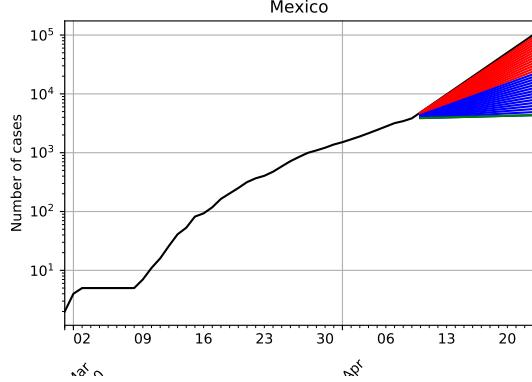
Simulations for CDMX show a following trend of an order of  $10^3$  by April 11th, however the reported cases was 1117 by that time, therefore we are in the horizon to see if we remains or change to another predicted outcome.

## Further results

We also calculate the minimum, maximum and average growth rate for each data set (Table 3). These calculation were made using the algorithm in [7]. The python code implementation is found in Authors' GitHub links.

*Case fatality rate* is defined as the ratio between confirmed deaths and cases, and up to April 11th, the followings values are the corresponding ones for each country.

Cumulative cases forecast



Prevalence forecast

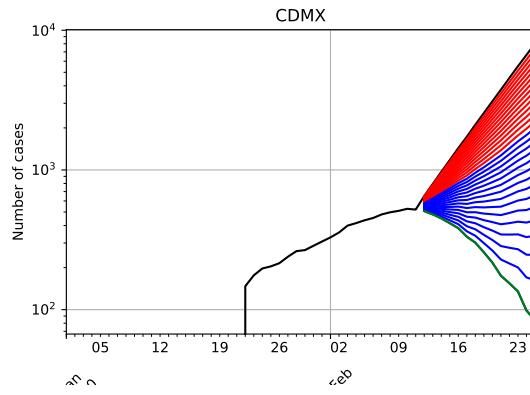
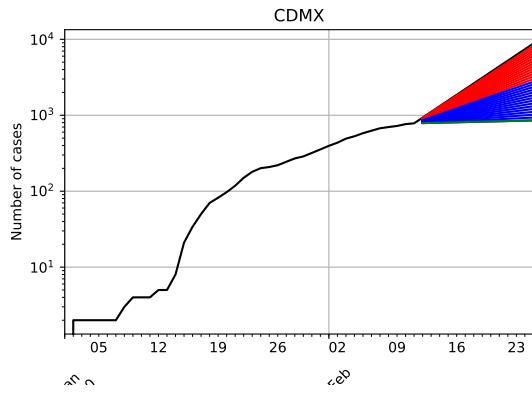
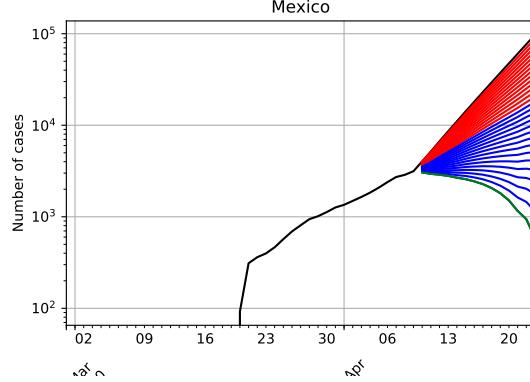


Figura 25: Cumulative confirmed cases in Mexico can reach and order of  $10^4$  cases about April 23rd if the trend remains. Whereas in the worst scenario the number could reach up to an order  $10^5$ . Meanwhile the prevalence predictions exhibit a wider range of possible scenarios. Cumulative confirmed cases in CDMX exceeds an order of  $10^3$  cases about April 15th if the trend remains. Whereas in the worst scenario the number could reach up to an order  $10^4$  by April 23rd. Meanwhile the prevalence predictions also exhibit a wider range of possible scenarios.

Country	min growthrate	max growthrate	average growthrate
USA	0.07623279896488011	0.22901629635290294	0.1189618517997764
IRAN	0.0	0.12542601252278676	0.05515815372759896
SLOVENIA	0.020470829068577334	0.11202531645569626	0.04450916072122805
GERMANY	0.036406077956397276	0.10965769082326408	0.06542385215888666
MEXICO	0.0817353033637203	0.21924686192468615	0.12774276309241947
CDMX	0.024869109947643908	0.15724137931034488	0.07949495911236551

Cuadro 3: Grow rates of cumulative prediction for each dataset

## Discussion

- We can perform simulations more accurately using case fatality rates for each country instead of the assumptions made in the paper [7].
- Using this method, we can also predict scenarios for sub classes of confirmed positive

individuals such as groups by age and preexisting health conditions.

- It is also possible to test this method applying it to data in previous days to evaluate which scenario each country is following with respect to growth rates increasing.

## 12. Report on the consequences of the mitigation measures announced on March 23 and March 30.

**Participants:** Manuel Acuna (U. Sonora), Mario Santana (C. CONACyT), Jorge X. Velasco (IM).

### Introduction: general assumptions

The mitigation measures announced on March 23 and March 30, attempt to reduce the contact rate between individuals. CoVID-19 is transmitted however by both symptomatic and asymptomatic individuals. Asymptomatic transmission exists, it is not negligible, but its true magnitude is still under study. The consequence of this fact is that the contact rate of symptomatic and asymptomatic individuals are different, thus in this model we incorporate the difference in the expression for the force of infection. The diagrammatic representation of the basic model is presented in Figure 26.

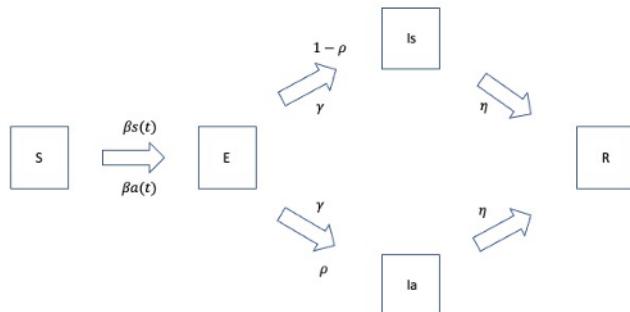


Figura 26: Diagramatic model. S, susceptible, E, exposed, Is, Ia, symptomatic and asymptomatic, R, recovered individuals respectively.  $\beta_k$  with  $k = s, k = a$  contact rates for symptomatic and asymptomatic individuals,  $1/\gamma$  incubation period,  $1/\eta$  infectious period,  $\rho$  proportion of the exposed cases that become asymptomatic carriers.

The contact rate has the form illustrated in Figure 27. The subindex  $k$  refers to symptomatic or asymptomatic individuals. The time from the reporting of the first cases and the enforcement of social distancing mitigation measures is given by  $T_\theta =$  March 23. After March 23 the original contact rate  $\beta_k$  starts to decline towards a contact rate target  $q_1 \beta_k$  where  $q_1$  is the desired proportion in which the contact rate is expected to be reduced after  $\theta$  days. This time is the ?learning? time, the time it takes the population to reach the desired target reduction. In México, on March 30 the suspension of non-essential activities was announced. In this model, this action is seen as a reinforcement of the already ongoing reduction in contact rate.

We postulate a time  $1/\omega$  that is related to the average time an individual is committed to all the behavioral requirements of effective social distancing (not leaving home, washing hands, frequently cleaning common surfaces, using masks, etc). This means that we assume that the

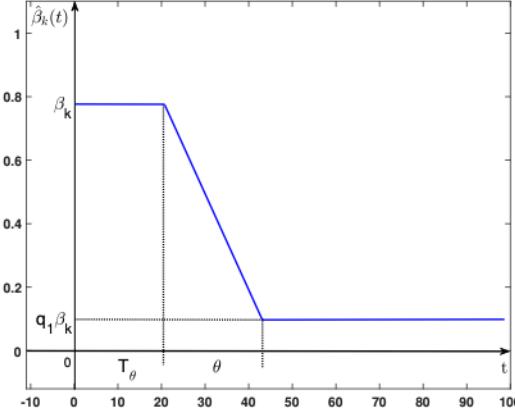


Figura 27: Illustration of the general form and parameters for the variable contact rate. See text for description.

mitigation measures are leaky, not everybody follows them.

We assume that a proportion  $q$  of the population is effectively applying the mitigation (stay-at-home) measures indicated by the health authorities. The population is thus subdivided into two separate parts beginning March 23: those that follow the mitigation measures, and those that do not. See Figure 28

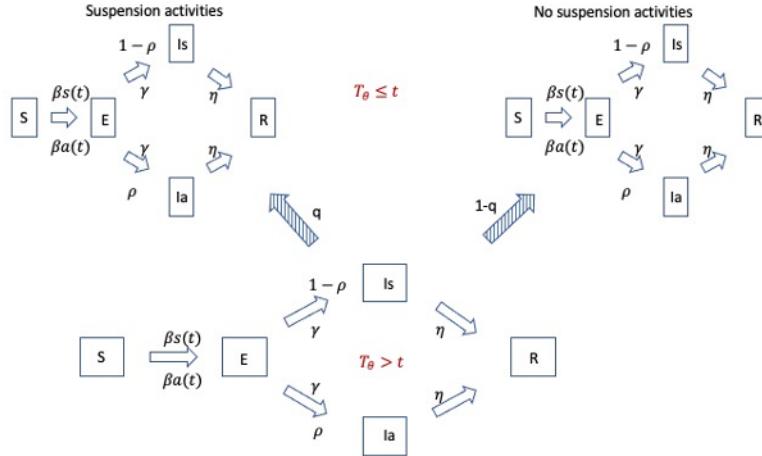


Figura 28: On  $T_\theta$  = March 23, the population is effectively split into two subpopulations: one that subject to the mitigation measures, and other that does not follow them. See text for further description

We assume that both subpopulations do get a reduction in contact rates but with different target. We therefore set  $q_1\beta_k$  to be the target contact rate for the population abiding to the stay-at-home instructions and  $q_2\beta_k$  be the target contact rate for those that do not follow the indication; of course  $q_1 < q_2$ , the reduction of those at home is better than those that do not stay at home.

We assume that a person follows or does not follow the mitigation indicatives. This amounts

to assume that if in a household one person does not follow the instruction he or she elevates the risk of the whole household and thus the whole household belongs to the subpopulation that does not follow the mitigation indicatives. A more realistic model can postulate mixing between these two populations. This will be presented in a forthcoming update of the present model.

To end this section the parameters for the incubation, infection periods and proportion of asymptomatics throughout the simulations are  $1/\gamma = 5,1$ ,  $\rho = 0,45$  [5, 6]. The parameters  $\eta$ ,  $\beta_a$  and  $\beta_s$  were estimated from data.

## Results

Results are presented according to scenarios. These are defined in terms of  $\theta_1$ ,  $\theta_2$  the learning times for stay-at-home and not stay-at-home, respectively;  $q_1$ ,  $q_2$  the target contact rate for stay-at-home and not stay-at-home, respectively and  $1/\omega$  the average time to abandon the stay-at-home indicatives.

We use  $\theta_1 = 150$ ,  $\theta_2 = 150$ ; these values were chosen so that the contact rate reduction of the stay-at-home and outside home subpopulations where on average, of 23 % and 18 % respectively, by April 30th, a very conservative estimation. Then we assume that the contact rate decreases linearly until reaching the target reductions of  $q_1 = 0,1$  and  $q_2 = 0,3$  for the stay-at-home and outside home subpopulations (reductions of 90 % and 70 %) respectively.

### Illustrative example: varying $q$ , the proportion of the population obeying the mitigation indicatives.

In this example, we set  $1/\omega = 200$  days, average time an individual is committed to the behavioral requirements of effective social distancing.

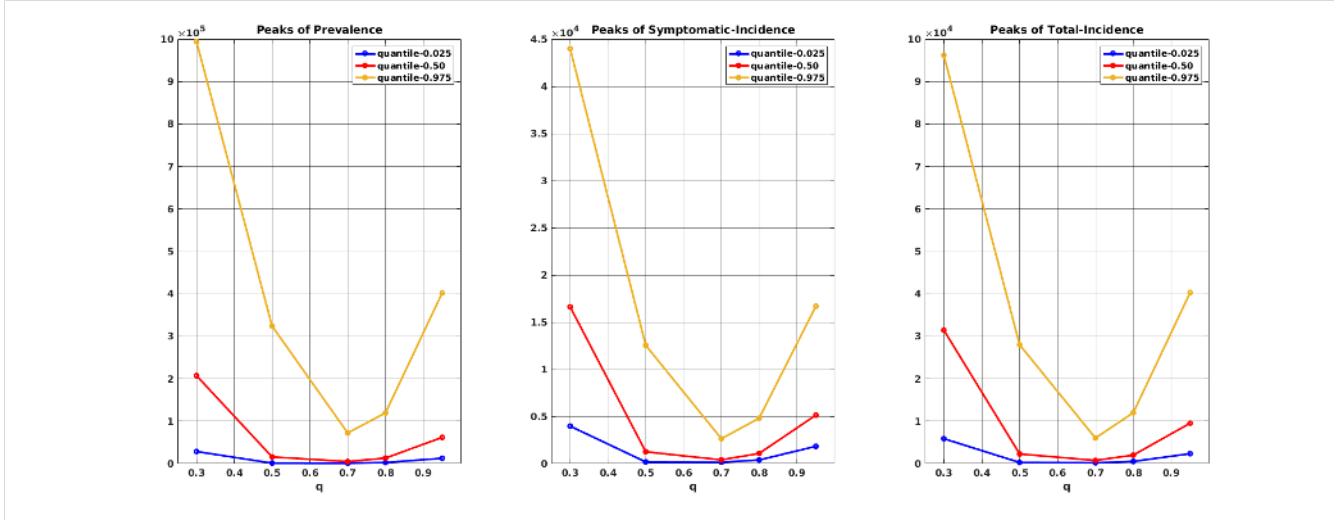
1. Dates to reach maximum prevalence for different values of  $q$ , the proportion of individuals that obey the stay-at-home indicative

		Quantile		
		0.025	0.5	0.975
$q$	0.3	July 04	July 22	August 08
	0.5	July 03	July 28	August 21
	0.7	May 16	June 04	August 10
	0.8	May 24	June 08	July 08
	0.95	June 03	June 14	July 03

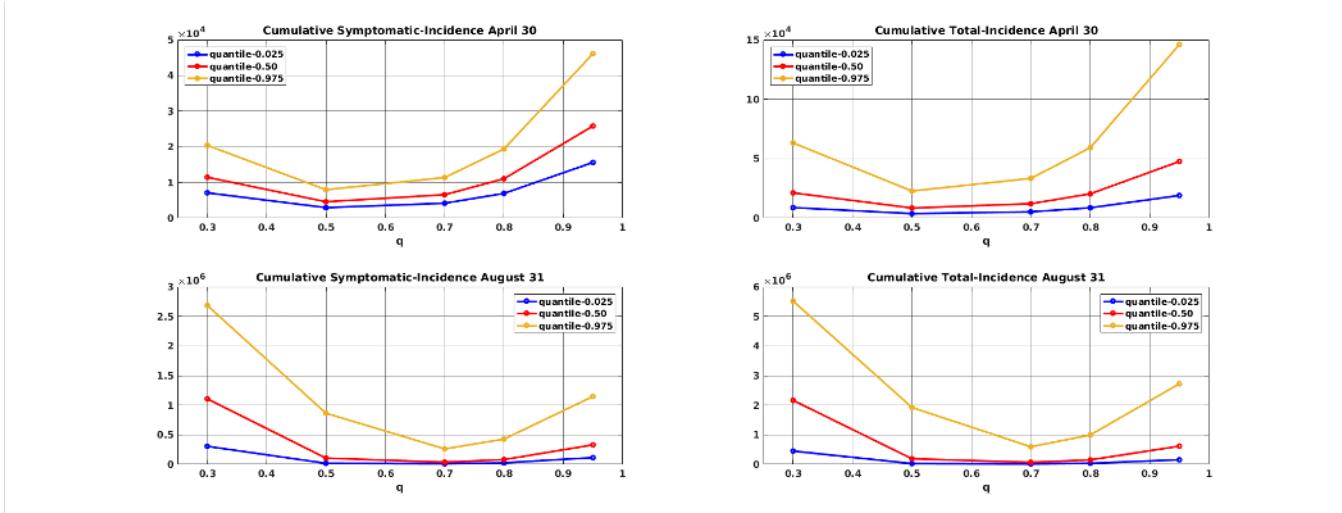
2. Dates to reach maximum incidence (symptomatic + asymptomatic cases) for different values of  $q$ , the proportion of individuals that obey the stay-at-home indicative
3. Maximum prevalence, symptomatic incidence and total incidence (symptomatic + asymptomatic) as a function of  $q$ , the proportion of individuals that obey the stay-at-home indicative. Observe the existence of a minimum. This reflects the balance that the model implicitly assumes between  $q$  and the “leak” rate  $\omega$ . When  $q$  is low, few people is committed to the mitigation indicatives and the leaking rate has no effect. As  $q$  increases and more

		Quantile		
		0.025	0.5	0.975
$q$	0.3	June 23	July 11	July 25
	0.5	June 24	July 17	August 05
	0.7	May 07	May 23	July 27
	0.8	May 15	May 28	June 19
	0.95	May 25	June 03	June 17

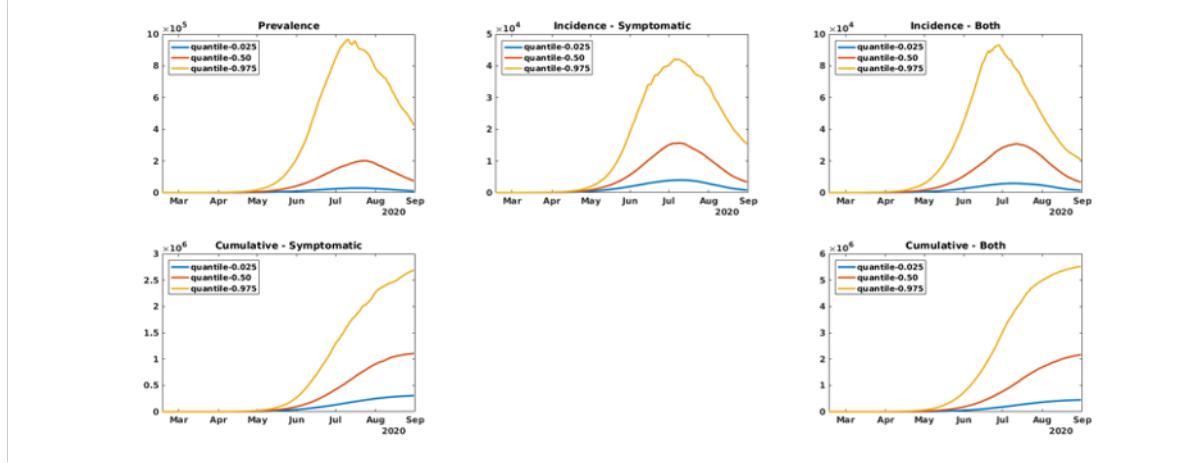
people is following the mitigation indicatives, the net flux of people abandoning social distancing grows larger until it counterbalances the effect of  $q$  (for  $q$  near 1).



4. Same effect as in 3. can be seen here for the cumulative incidence. We plot three quantiles for April 30th, the date where supposedly social distancing indicatives will end, and for August 30th past the most probable dates of maximum incidence. Note that the trade-off between  $q$  and  $\omega$  still holds.



5. Prevalence, symptomatic incidence, total incidence and envelop (quantiles) cumulative incidence curves for  $q = 0,3$  (30 % of the populations follows the mitigation indicatives). In the figures “both” means symptomatic and asymptomatic cases grouped together.



## Conclusions

The forecasts produced by this model are limited by the nature of the mathematical framework and by the quality and completeness of the data we have available to make inferences. The mathematical limitations are common to most of the models used to describe COVID-19. This include a high sensitivity to initial conditions and parameter estimates which result in a wide variability for the intervals associated to the range of model solutions. In practical termos, this model produces a wide range of scenarios depending upon the baseline parameters. We have presented here an illustrative example. We have available upon request a simulator where the interested user may explore graphically the behavior of the model as the parameters of interest are varied.

We have incorporated a variable contact rate in the form of a step function to mimic the learning process that this instruction of behavioral change (namely, the Susana Distancia program) involves. We are now implementing the next step: the liberation from the mobility restrictions planned, in principle, for April 30th. This would imply, among other things, the increase of the contact rate achieved during the stay-at-home program. Our initial simulations indicate the unavoidable appearance of a second epidemic wave resulting from the liberation of susceptible individuals into a population far from the percentage of herd immunity required to prevent a new outbreak. We are exploring the plausible scenarios for mitigating this new wave as functions of the way the contact rate is increased upon the lifting of restrictions.

The true percentage of the population following in a significant manner the stay-at-home program is unknown but it is likely to be low. In Table 1 we have listed several possible outcomes for the dates of maximum prevalence or incidence. The scenario reported by the Secretaria de Salud of a maximum incidence reached on late May or early June is only compatible with a relatively high proportion of compliance (above  $q = ,7$ ). For lower values of  $q$ , the peak is predicted to occur further into June and most likely until July.

In 3 of the Results, we pointed out the existence of a minimum for the incidence as  $q$  varies between 0 and 1. As stated then, this reflects the balance that the model implicitly assumes between  $q$  and the “leak” rate  $\omega$ . When  $q$  is low, few people is committed to the mitigation indicatives and the leaking rate has no effect. As  $q$  increases and more people is following the mitigation indicatives, the net flux of people abandoning social distancing grows larger until it counterbalances the effect of  $q$  (for  $q$  near 1).

## Referencias

- [1] Acuna-Zegarra, A., M. Santana-Cibrian, and J. X. Velasco-Hernández (2020, 12 April 2020). Report on the consequences of the mitigation measures announced on march 23 and march 30. *Technical report: Nodo Multidisciplinario de Matemáticas Aplicadas (NoMMA)*.
- [2] Deasy, J., E. Rocheteau, K. Kohler, D. J. Stubbs, P. Barbiero, P. Liò, and A. Ercole (2020). Forecasting ultra-early intensive care strain from covid-19 in england. *medRxiv*.
- [3] Finkenstaedt, B. F. and B. T. Grenfell (2000). Time series modelling of childhood diseases: a dynamical systems approach. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* @ (2), 187–205.
- [4] Flaxman, S., S. Mishra, A. Gandy, H. J. T. Unwin, H. Coupland, T. A. Mellan, T. Berah, J. W. Eaton, P. N. P. Guzman, N. Schmit, L. Cilloni, K. E. C. Ainslie, I. Blake, A. Boonyasiri, O. Boyd, L. Cattarino, C. Ciavarella, L. Cooper, Z. Cucunubá, G. Cuomo-dannenburg, A. Dighe, B. Djaafara, I. Dorigatti, and S. V. Elsland (2020). Estimating the number of infections and the impact of non- pharmaceutical interventions on COVID-19 in 11 European countries. (March), 1–35.
- [5] Nishiura, H., T. Kobayashi, A. Suzuki, S.-M. Jung, K. Hayashi, R. Kinoshita, and T. Miyama (2020). Estimation of the asymptomatic ratio of novel coronavirus infections (covid-19). *Journal of Infectious Diseases*.
- [6] Nishiura, H., N. M. Linton, and A. R. Akhmetzhanov (2020). Serial interval of novel coronavirus (covid-19) infections. *Journal of Infectious Diseases*.
- [7] Perc, M., N. Gorisek Miksic, M. Slavinec, and A. Stozer (2020). Forecasting covid-19. *Frontiers in Physics* 8, 127.
- [8] Statista (2020). Percentage of covid-19 cases in the united states from february 12 to march 16, 2020 that resulted in hospitalization.
- [9] Weiss, P. and D. R. Murdoch (2020). Clinical course and mortality risk of severe covid-19. *The Lancet*.
- [10] Wood, R. M., C. J. McWilliams, M. J. Thomas, C. P. Bourdeaux, and C. Vasilakis (2020). Covid-19 scenario modelling for the mitigation of capacity-dependent deaths in intensive care: computer simulation study. *medRxiv*.

Reporte de avances  
Análisis de datos COVID-19  
Dr. Jorge Luis Pérez González  
jorge.perez@iimas.unam.mx

**Objetivos:**

- 1) Clasificar automáticamente pacientes con COVID-19 y no COVID-19, usando el conjunto de datos de la Secretaría de Salud de CDMX.
- 2) Clasificar automáticamente entre pacientes diagnosticados con COVID-19 de bajo riesgo y alto riesgo.

**Análisis 1- Clasificación automática de pacientes COVID-19**

Para este análisis se tomaron los resultados de diagnóstico dados por el INDRE, donde se analizaron un total de 800 casos balanceados entre las dos clases (extraído del concentrado de casos al 5 de abril). El algoritmo empleado fue un bosque aleatorio, con hiperparámetros optimizados. Los datos de entrada fue información sobre viajes, demográfica, enfermedades crónicas y sintomatología. Los resultados encontrados con la validación cruzada de 10 vías son:

- Sensibilidad:  $67 \pm 11\%$
- Especificidad:  $62 \pm 9\%$
- Accuracy:  $63 \pm 8\%$
- AUC-ROC:  $67 \pm 8\%$

Dado que el porcentaje ideal es 100% y el peor 50% en todas las métricas, se considera que no es factible hacer diagnóstico certero mediante estos datos. En la figura 1, se muestra el detalle del la curva ROC.

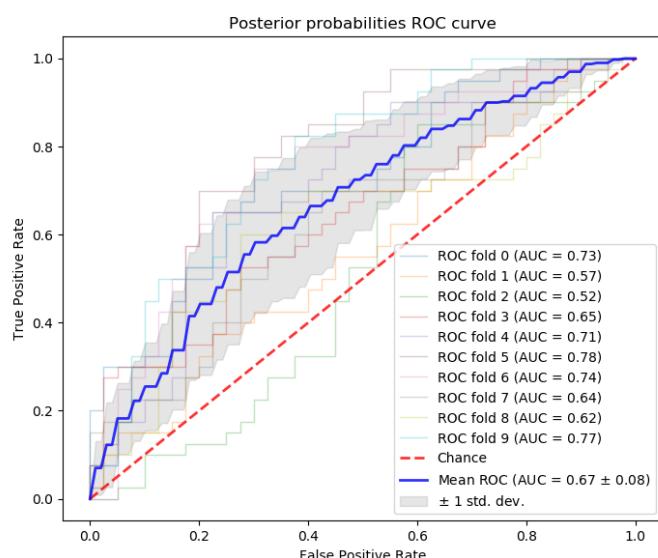


Fig. 1 Área bajo la curva ROC, de la clasificación entre pacientes COVID-19 y no COVID-19

Adicionalmente, se efectuó un análisis de la importancia de las características usadas en la clasificación (figura 2). Dicho análisis se efectuó mediante el algoritmo de árboles aleatorios, con un total de 1000 iteraciones. Las características están ordenadas de acuerdo a su nivel de importancia. Se puede observar, que la edad es por mucho la característica mas influyente, seguido a la delegación de procedencia del paciente. Esto se puede deber a que, existen cúmulos de brotes a diversos niveles y en distintas delegaciones de la CDMX. Las demás características son concordantes con lo reportado ya hasta el momento, donde la rinitis, odinofagia, fiebre y dolor torácico resaltan en el diagnóstico de pacientes con COVID-19. Por otro lado, la insuficiencia hepática o renal, los meses de embarazo y convulsiones parecen aportar poca información en el diagnóstico.

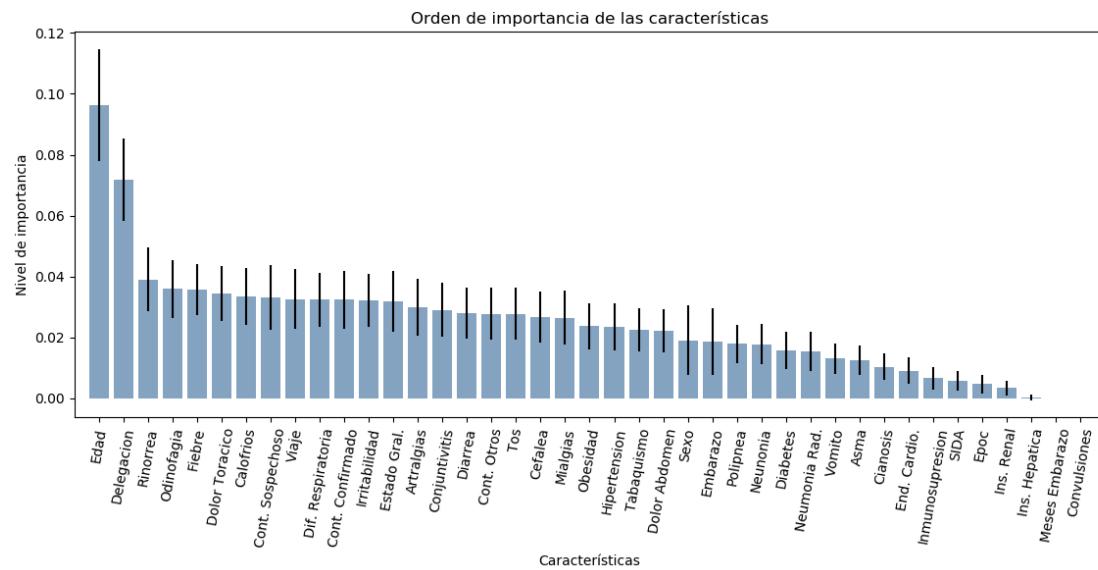


Fig. 2 Relevancia de las características en la clasificación de pacientes con COVID-19 y no COVID-19

## Análisis 2- Clasificación automática de pacientes con COVID-19 de alto riesgo y bajo riesgo

Para este análisis se tomaron los resultados de diagnóstico dados por el INDRE, donde se analizaron un total de 280 casos diagnosticados con COVID-19. Los 280 pacientes se dividieron en dos grupos:

- Grupo 1: Caso no grave o dado de alta
- Grupo 2: Caso grave, intubado o defunción.

Los dos grupos están balanceados (todos los datos fueron extraídos del concentrado de casos al 5 de abril). El algoritmo empleado fue un bosque aleatorio, con hiperparámetros optimizados. Los datos de entrada fue información sobre viajes, demográfica, enfermedades crónicas y sintomatología. Los resultados encontrados con la validación cruzada de 10 vías son:

- Sensibilidad:  $64 \pm 25\%$
- Especificidad:  $93 \pm 4\%$
- Accuracy:  $86 \pm 5\%$
- AUC-ROC:  $90 \pm 4\%$

Dado que el porcentaje ideal es 100% y el peor 50% en todas las métricas, se observa que se tiene un buen resultado para clasificar entre pacientes de bajo y alto riesgo. Sin embargo, hay una baja

sensibilidad ( $64 \pm 25\%$ ), eso indica que varios pacientes de alto riesgo son clasificados como de bajo riesgo, lo cual representa un problema. Se cree, que es necesario contar con mas datos en este etapa para poder comprobar los resultados obtenidos y generar una calculadora de riesgo. En la figura 3 se muestra el detalle del desempeño con respecto al área bajo la curva ROC.

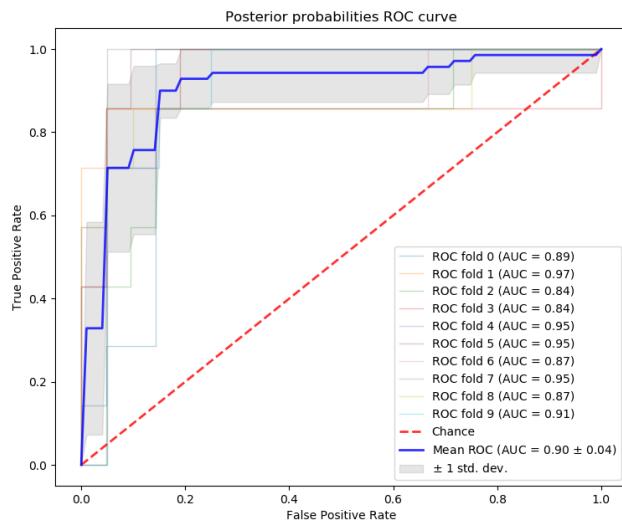


Fig. 3 Área bajo la curva ROC, de la clasificación entre pacientes con COVID-19 de alto y bajo riesgo.

Adicionalmente, se efectuó un análisis de la importancia de las características usadas en la clasificación (figura 4). Dicho análisis se efectuó mediante el algoritmo de árboles aleatorios, con un total de 1000 iteraciones. Las características están ordenadas de acuerdo a su nivel de importancia. Se puede observar que el diagnóstico de neumonía clínica y radiológica, la dificultad respiratoria, la edad y la polipnea son indicadores relevantes para identificar pacientes de alto riesgo. Por otro lado, las convulsiones, meses de embarazo, insuficiencia hepática, la inmunodepresión o el SIDA, no son indicadores de riesgo significativo para un paciente diagnosticado con COVID-19.

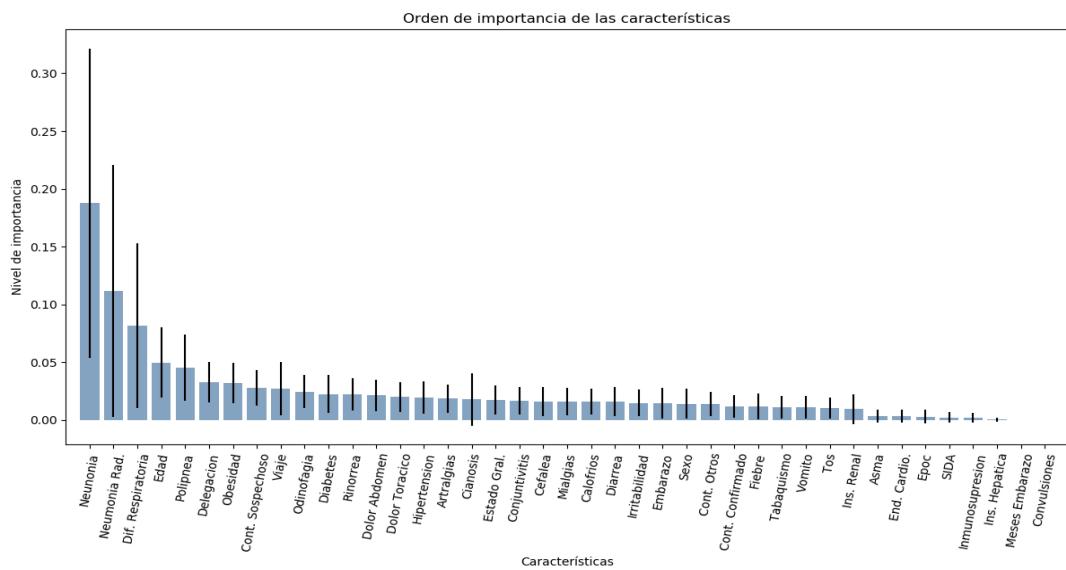


Fig. 4 Relevancia de las características en la clasificación de pacientes con COVID-19 de alto y bajo riesgo.

## **Conclusiones**

Todos los resultados son preliminares, es necesario contar con mayor número de casos para poder dar un análisis final. En general se puede observar que es difícil clasificar correctamente pacientes con y sin COVID-19. Sin embargo, se considera que el segundo análisis presentado podría ayudar a evaluar el riesgo de un paciente infectado con COVID-19, dada su tasa de asertividad. Por otro lado, se confirman varios de los indicadores ya conocidos para el diagnóstico y el nivel de riesgo usando métodos de reconocimiento de patrones.

## **Preguntas**

¿Se puede contar con los datos actualizados a la fecha?

¿Cuentan con estudios de Rayos X o Tomografía para cada paciente?

Me gustaría tener mas detalle de los pacientes etiquetados como “tratamiento y seguimiento”, ¿Son de alto riesgo?, ¿Están dados de alta? o ¿Cuál es la situación de ellos?

## Informe: Logística Hospitales

### Objetivo inicial:

- Generar un modelo para distribuir los pacientes entre los hospitales disponibles, no solo infectados por coronavirus también enfermos de otras tantas enfermedades como distribuirlos de manera eficiente. Es un problema de optimización y logística.

### Integrantes:

- Helena Gómez (IIMAS)
- Pilar Angeles (IIMAS)
- Victor Lomas (IIMAS)
- Luis Pineda (IIMAS)
- Monica Vazquez (IIMAS)
- Nora Perez (IIMAS)
- Mario Peña (IIMAS)
- Tamara Alcántara Concepción (DGTIC)
- Katya Rodriguez (IIMAS)
- Sergio Padilla (IIMAS)
- Gibran Fuentes (IIMAS)
- Caleb Rascon (IIMAS)
- Oscar Esquivel (IIMAS)
- Francisco Viveros
- Adrian Soto

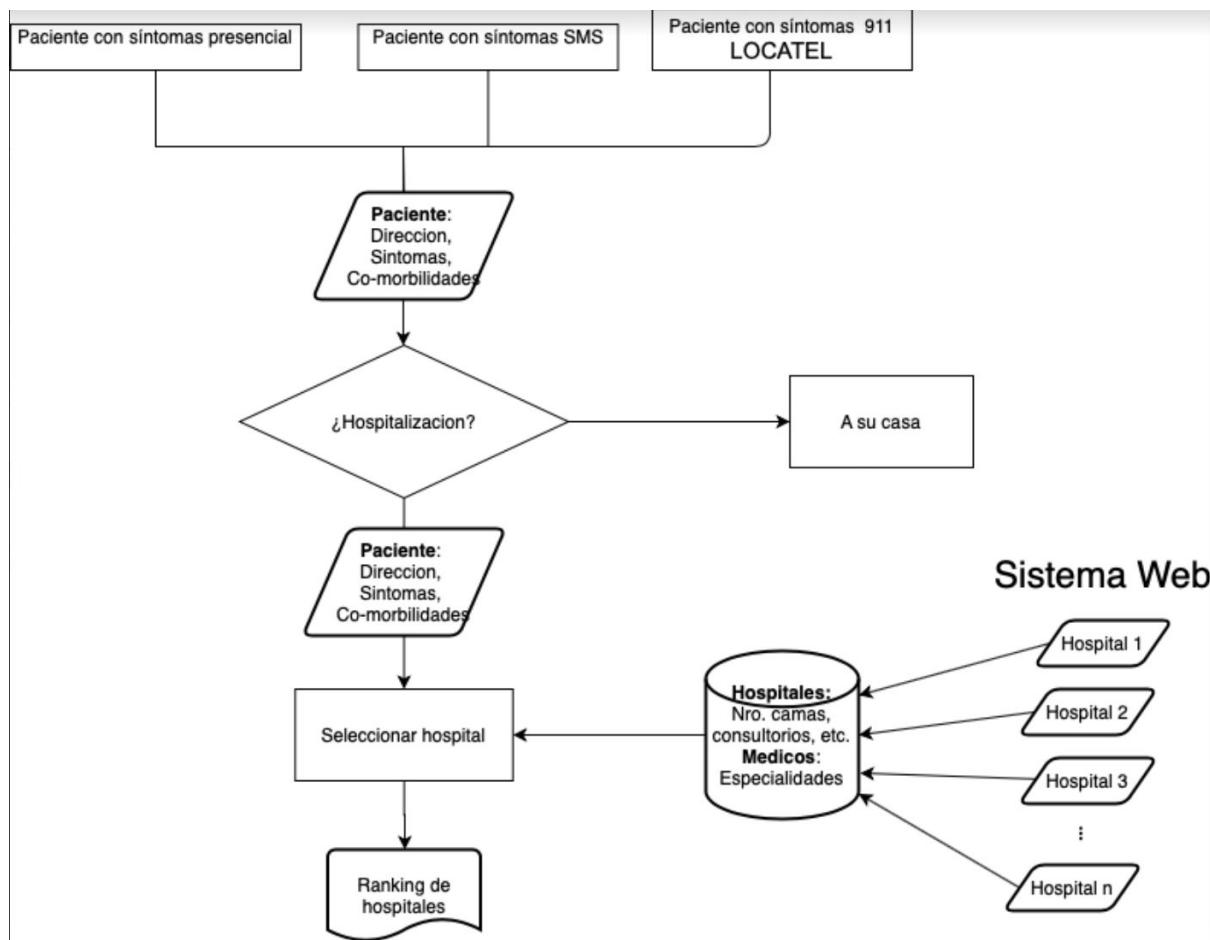
### Trabajos realizados:

Hemos trabajo en el diseño y dimensionamiento del sistema. Ya se tiene una propuesta de Diagrama Entidad-Relación y ya se escogieron las herramientas tecnológicas. La arquitectura estará basada en un sistema web con conexión asíncrona con una base de datos SQL. El Backend y Frontend se espera sean completamente independientes para el potencial crecimiento del sistema así como para garantizar la seguridad de los datos.

La metodología para asignación de hospital a un paciente se abordará desde el punto de vista de optimización. El objetivo es optimizar los recursos de los hospitales y reducir el tiempo de estadía de los pacientes en el mismo. Para mejorar el flujo de pacientes se modelará como un problema de colas.

Se están analizando formas de evaluar el sistema mediante simulaciones de diferentes escenarios, que permitan optimizar los recursos del sistema de salud en términos de ocupación de camas y personal sanitario. Los resultados ayudarán a seleccionar y priorizar las acciones necesarias para reducir el tiempo de estadía de los pacientes y, en consecuencia, reducir los tiempos de espera y el congestionamiento en los hospitales.

Proponemos el siguiente diagrama de flujo de datos en el que nosotros estamos trabajando en el proceso de selección de hospital y el sistema web de carga de datos de los hospitales:



**SUBGRUPO GEO**  
Colaboraciones COVID-19

El objetivo de este subgrupo, dentro del Grupo de Trabajo COVID19 Ambiente/COVID datos modelos, es reunir a diversos colaboradores, de diferentes entidades académicas de la UNAM (y otras instituciones), con capacidades en el manejo de sistemas de información geográfica (SIG) y en el desarrollo de plataformas de modelamiento de fenómenos y procesos naturales o antropogénicos (biológicos, geocientífico-ambientales o epidemiológicos), a encontrar soluciones de visualización de los resultados de varios grupos de trabajo en forma georreferenciada y hacerlos compatibles con otras plataformas usadas en el ámbito oficial (federal, estatal o municipal).

En este subgrupo se encuentran al momento participando catorce académicos de las siguientes instituciones: Instituto de Geografía (IGG), Centro de Investigación en Geografía Ambiental (CIGA), Instituto de Ciencias Nucleares (ICN), Instituto de Investigación en Matemáticas Avanzadas y Sistemas, Instituto de Física, (IF), Instituto de Investigación en Ecosistemas y Sustentabilidad (IIES), Instituto de Geofísica (IGF), Centro de Ciencias de la Atmósfera (CCA), Centro de Estudios Mexicanos – Costa Rica (CEM-CR), Facultad de Ingeniería (FI) y de la Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO).

Se han realizado varias reuniones virtuales utilizando la plataforma zoom con el fin de integrar el grupo y definir sus objetivos. En particular, se ha presentado la plataforma desarrollada en el ICN y en el IGG, así como la del CIGA.

Se ha discutido acerca de los contenidos del Atlas Nacional de Riesgos desarrollado en el Centro Nacional de Prevención de Desastres (CENAPRED) de la Secretaría de Seguridad y Protección Ciudadana (SSPC) del Gobierno Federal. La idea primordial es vislumbrar la mejor forma de integrar bases de datos del sector académico y del sector público, además de asegurar la accesibilidad.

Para lo anterior, se ha establecido contacto con el CENAPRED, quienes invitaron a una sesión virtual con el Dr. Dwight Dyer y el Ing. Enrique Guevara para presentar la disposición de la UNAM a colaborar con las autoridades federales en la presente coyuntura de la contingencia COVID-19. En esta reunión participó el Coordinador de la Investigación Científica, Dr. William Lee Alardín, el Dr. Héctor Benítez del IIMAS, el Dr. Carlos Valdés González del CEM-CR y el Dr. Hugo Delgado Granados del Instituto de Geofísica. Estuvieron presentes, la Directora del CONACyT, Dra. María Elena Álvarez Buylla y varios investigadores del CONACyT y centros CONACyT. Después de la presentación se discutió acerca de la disponibilidad de datos y de enviar un resumen de las capacidades y acciones con las que la UNAM puede contribuir.

En comunicación posterior con el Ing. Guevara se acordó el acceso a una aplicación en línea donde se encuentra la información con la que cuenta a la fecha el CENAPRED, con el fin de comenzar a vislumbrar la manera de hacer fluir la información en ambos sentidos.

**Participantes del Sub Grupo GEO:**

Adrián Ghilardi  
Agustín Fernández Equiart  
Alejandro Pérez Riascos  
Boris Escalante  
Carlos Valdés González  
Christopher Stephens  
Gabriel Ernesto García peña  
Héctor Rescendiz  
Hugo delgado Granados  
Leticia Flores  
Manuel Suárez  
Nelly Ramírez  
Roberto Bonifaz  
Víctor Velasco



## Iniciativa UNAM COVID19

# Reporte Grupo de Trabajo de EPI-SPECIES

**12 de abril 2020**

**Preparado por Christopher R. Stephens**

### **Resumen**

El grupo de EPI-SPECIES consiste en investigadores, técnicos y estudiantes de varias dependencias de la UNAM, de varios posgrados de la UNAM, y del CONABIO, CENAPRECE, IMSS y el Hospital General. Actualmente hay más que 12 personas trabajando en el proyecto.

El enfoque principal del grupo en el contexto la emergencia sanitaria de SARS-Cov-2 es producir modelos para predecir riesgos multifactoriales para los principales estados clínicos y epidemiológicos de la enfermedad (la “demanda”) y que ayudan en la planeación del sistema de salud (la “oferta”).

Se distingue el trabajo del grupo por su énfasis en la creación y uso de Plataformas de análisis de datos que:

- i) Se crea modelos usando métodos de aprendizaje de máquina de tipo Bayesiano
- ii) Que pueden usar como insumos cualquier tipo de dato – categórico, espacio-temporal de distintas resoluciones (rasters, puntos geo-referenciados etc.)
- iii) Que crean modelos predictivos nuevos con cientos o hasta miles de variables predictores (factores de riesgo) dentro de segundos o minutos

- iv) Que pueden ser usadas por tomadores de decisión sin entrenamiento previo en la modelaje o aprendizaje de máquinas
- v) Que pueden ser y han sido usadas para el análisis de otras enfermedades o fenómenos donde el análisis de riesgos dinámicos y adaptativos es importante – zoonosis, Enfermedades no-transmisibles crónico-degenerativas etc.

El trabajo actual consiste en convertir la Plataforma SPECIES ([species.conabio.unam.mx](http://species.conabio.unam.mx), [species.c3.unam.mx](http://species.c3.unam.mx)) a un contexto epidemiológico donde se puede alimentar el sistema con datos epidemiológicos, sociales y clínicos versus datos epidemiológicos. Se ha logrado cambiar la rejilla que se usa el sistema para poder analizar riesgos y crear modelos predictivos al nivel de estado, municipio o AGEB. Se ha creado modelos iniciales de distribución de confirmados al nivel de estado usando como factores de riesgo factores socio-demográficos y socio-económicos del Censo 2010 del INEGI, y sospechosos al nivel de municipios. También se ha preparado datos de la distribución y capacidades de todos los hospitales y clínicas en México hasta el nivel de AGEB que actualmente están siendo cargados en la Plataforma.

Actualmente, usando los datos de SEDESA se ha creado 3 modelos iniciales: i) un modelo para distinguir entre los pacientes probados para SARS-CoV-2 quienes resultaron positivos versus los que no eran positivos; ii) un modelo para predecir quien está más en riesgo de morir de SARS-CoV-2 dado eran positivos para su presencia; iii) un modelo para distinguir entre positivos para SARS-CoV-2 y ellos quienes eran positivos para la presencia de otro virus. En los apéndices están varios pre-análisis de algunos de los factores de riesgo, incluyendo presencia de síntomas, co-morbilidades, factores demográficos y socio-económicos. Estamos actualmente analizando los demás factores.

Las Plataformas permiten la creación de tales modelos en una forma muy eficiente. Se necesita interactuar con expertos de otras disciplinas – médicos, epidemiólogos etc. Para aprovechar de sus capacidades.

## Modelos Predictivos

Se están creando dos tipos de modelo predictivos para distintas clases de interés clínico, epidemiológico y de salud pública. Uno es asociado con datos de la SEDESA de los pacientes ingresados en hospitales en la CDMX. Estos datos son “intrínsecas” en que son datos del paciente mismo. El segundo tipo de modelo es asociado con el “nicho” de las personas – es decir el entorno socio-demográfico, socio-económico, climatológico y ambiental etc. Se planea integrar los dos tipos de modelo.

Ambos tipos de modelo están siendo creados en Plataformas de Inteligencia Epidemiológica (Proyecto 42 - <http://132.247.183.47/#/admin/naive> y EPI-SPECIES - [http://species.c3.unam.mx/covid19/geoportal\\_v0.1.html](http://species.c3.unam.mx/covid19/geoportal_v0.1.html)) que pueden ser alimentados con datos de la SEDESA y de otras fuentes relevantes. Las Plataformas permiten la creación de un modelo predictivo Bayesiano del tipo  $P(C | X(t))$ , donde  $C$  es el grupo clínico/epidemiológico de interés y  $X(t) = (X_1, X_2, \dots, X_N)$  es un vector de factores de riesgo tanto de factores intrínsecas como extrínsecas. Estas Plataformas también serán disponibles a distintos grupos de interés con distintos niveles de seguridad dependiendo de la naturaleza de los datos. Modelos predictivos para una clase de interés  $C$  con cientos o miles de predictores pueden ser creados muy rápidamente.

De los datos de SEDESA:

1) Modelos actualmente en mano:

- i) Modelo predictivo para los pacientes quienes han sido probados positivos para la presencia de SARS-CoV-2 versus pacientes quienes no han sido probados positivos para la presencia de SARS-CoV-2 (comprende de negativos, rechazados, positivos para otros virus etc.)
- ii) Modelo predictivo para los difuntos – ellos quienes eran positivos para la presencia de SARS-CoV-2 quienes murieron. Nota que es sencillo producir modelos para cualquier estado de la variable EVOLUCION (CASO GRAVE, CASO NO GRAVE, EN TRATAMIENTO etc.) que significa que se puede construir modelos para pasar de cualquier estado clínico a cualquier otro.
- iii) Modelo predictivo para los pacientes comprobados como positivos para un virus otro que SARS-CoV-2 – otros coronovirus, H1N1, B, enterovirus, rinovirus. Este modelo permite ver si el cuadro clínico se difiere entre los distintos tipos de virus.

Nota que cada modelo analiza el conjunto completo de variables que están en los archivos proporcionados por la SEDESA permitiendo distintas perspectivas respecto a los distintos estados clínicos y epidemiológicos de relevancia. Algunas perspectivas importantes son: perfiles socio-demográficos y socio-económicos de los pacientes, antecedentes clínicos, su sintomatología etc. Igual, se puede determinar el grado de predictibilidad de cada estado clínico – positivo/negativo, difunto, hospitalizado, recuperado etc.

2) Modelos planeados:

- i) Cuadros de síntomas discriminatorios para distinguir entre positivos/negativos (para no desperdiciar recursos)
- ii) Cuadros de síntomas discriminatorios para distinguir entre positivos de SARS-Cov-2 versus otras enfermedades respiratorias.
- iii) Modelos geográficos al nivel de municipio (o al nivel de AGEB si se logra tener mejor georreferenciación) para incluir factores ambientales, socio-demográficos y socio-económicos. Esto puede ser particularmente útil dado que estos factores pueden ser proxies para conducta y es la conducta que determina el ciclo de transmisión.

## Anexo 1:

Modelo 1)i)

Factores de riesgo

Cuadro Clínico

Síntomas Principales

Síntoma	Presencia en paciente probado	Probabilidad ser confirmado con síntoma ( $P(C/X)$ )	Probabilidad ser confirmado	Verosimilitud ( $P(X/C)$ )	Epsilon	RISK RATIO	ODDS RATIO	Número de confirmados con la síntoma	Número con la síntoma	Número confirmados	NTotal
FIEBRE	SI	0.28	0.23	0.80	5.79	2.01	2.40	741	2647	929	3997
FIEBRE	NO	0.14	0.23	0.20	-8.09	0.50	0.42	188	1349	929	3997
FIEBRE	SE IGNORA	0.00	0.23	0.00	-0.55	-	-	0	1	929	3997
TOS	SI	0.24	0.23	0.85	0.38	1.08	1.10	790	3359	929	3997
TOS	NO	0.22	0.23	0.15	-0.87	0.93	0.91	139	638	929	3997
DOLOR TORACICO	SI	0.24	0.23	0.36	1.08	1.08	1.11	335	1369	929	3997
DOLOR TORACICO	NO	0.23	0.23	0.63	-0.46	0.96	0.94	587	2568	929	3997
DOLOR TORACICO	SE IGNORA	0.12	0.23	0.01	-2.12	-	-	7	60	929	3997
DISNEA (DIFICULTAD RESPIRATORIA)	SI	0.31	0.23	0.40	6.60	1.58	1.85	369	1176	929	3997
DISNEA (DIFICULTAD RESPIRATORIA)	NO	0.20	0.23	0.60	-4.30	0.63	0.54	559	2820	929	3997
DISNEA (DIFICULTAD RESPIRATORIA)	SE IGNORA	1.00	0.23	0.00	1.82	-	-	1	1	929	3997

Figura 1: Cuadro clínico de positivos de SARS-CoV-2 de los síntomas principales  
Síntomas Secundarios

En Figura 1 se ve el cuadro clínico diagnostico que distingue entre los positivos de la prueba de SARS-COV-2 y los demás. Se nota que fiebre si es como esperado un factor muy notable con un riesgo de 140% más. También de la verosimilitud que cubre muchos de los pacientes probados – 80%. Sin embargo, si una condición para recibir una prueba en primer lugar es mostrar síntoma de fiebre entonces es tautológico. Interesantemente, 85% muestran el síntoma de tos, pero no distingue entre positivos y los demás. Tampoco distingue dolor torácico. Sin embargo, Disnea si es un factor que distingue entre positivos y los demás

Síntoma	Presencia en paciente probado	Probabilidad ser confirmado con síntoma	Probabilidad ser confirmado	Verosimilitud (P(X C))	Epsilon	RISK RATIO	ODDS RATIO	Número de confirmados con la síntoma	Número con la síntoma	Número confirmados	NTotal
ATRALGIAS (DOLOR EN LAS ARTICULACIONES)	SI	0.27	0.23	0.59	3.68	1.36	1.49	552	2071	929	3997
ATRALGIAS (DOLOR EN LAS ARTICULACIONES)	NO	0.20	0.23	0.40	-3.45	0.76	0.70	373	1876	929	3997
ATRALGIAS (DOLOR EN LAS ARTICULACIONES)	SE IGNORA	0.08	0.23	0.00	-2.55	0.34	0.28	4	50	929	3997
ATAQUE AL ESTADO GENERAL	SI	0.26	0.23	0.66	2.68	1.29	1.40	616	2411	929	3997
ATAQUE AL ESTADO GENERAL	NO	0.20	0.23	0.33	-2.99	0.79	0.74	309	1543	929	3997
ATAQUE AL ESTADO GENERAL	SE IGNORA	0.09	0.23	0.00	-2.16	0.40	0.34	4	43	929	3997
CALOFRIOS	SI	0.28	0.23	0.55	4.42	1.41	1.57	511	1853	929	3997
CALOFRIOS	NO	0.20	0.23	0.45	-3.82	0.73	0.66	414	2099	929	3997
CALOFRIOS	SE IGNORA	0.09	0.23	0.00	-2.28	0.38	0.32	4	45	929	3997
CEFALEA (DOLOR DE CABEZA)	SI	0.23	0.23	0.79	0.03	1.00	1.01	731	3142	929	3997
CEFALEA (DOLOR DE CABEZA)	NO	0.23	0.23	0.21	-0.06	1.00	0.99	198	855	929	3997
CIANOSIS	SI	0.30	0.23	0.05	1.89	1.30	1.44	42	140	929	3997
CIANOSIS	NO	0.23	0.23	0.95	-0.14	0.92	0.90	879	3798	929	3997
CIANOSIS	SE IGNORA	0.14	0.23	0.01	-1.76	0.58	0.51	8	59	929	3997
CONJUNTIVITIS	SI	0.22	0.23	0.19	-0.71	0.94	0.93	172	776	929	3997
CONJUNTIVITIS	NO	0.24	0.23	0.81	0.57	1.09	1.12	750	3169	929	3997
CONJUNTIVITIS	SE IGNORA	0.13	0.23	0.01	-1.67	0.58	0.51	7	52	929	3997
DIARREA	SI	0.27	0.23	0.25	2.78	1.23	1.32	231	847	929	3997
DIARREA	NO	0.22	0.23	0.75	-1.43	0.81	0.76	698	3149	929	3997
DIARREA	SE IGNORA	0.00	0.23	0.00	-0.55	0.00	0.00	0	1	929	3997
DOLOR ABDOMINAL	SI	0.24	0.23	0.17	0.19	1.02	1.02	155	658	929	3997
DOLOR ABDOMINAL	NO	0.23	0.23	0.83	0.14	1.03	1.03	767	3285	929	3997
DOLOR ABDOMINAL	SE IGNORA	0.13	0.23	0.01	-1.79	0.55	0.49	7	54	929	3997
IRRITABILIDAD	SI	0.24	0.23	0.24	0.79	1.06	1.08	219	899	929	3997
IRRITABILIDAD	NO	0.23	0.23	0.76	-0.43	0.94	0.92	710	3098	929	3997
MIALGIAS (DOLOR MUSCULAR)	SI	0.26	0.23	0.67	3.58	1.40	1.55	618	2344	929	3997
MIALGIAS (DOLOR MUSCULAR)	NO	0.19	0.23	0.33	-3.82	0.74	0.68	308	1603	929	3997
MIALGIAS (DOLOR MUSCULAR)	SE IGNORA	0.06	0.23	0.00	-2.89	0.26	0.21	3	50	929	3997
ODINOFAGIA (DOLOR DE GARGANTA AL TRAGAR ALIMENTOS)	SI	0.20	0.23	0.46	-3.44	0.75	0.68	431	2144	929	3997
ODINOFAGIA (DOLOR DE GARGANTA AL TRAGAR ALIMENTOS)	NO	0.27	0.23	0.53	4.05	1.37	1.51	492	1804	929	3997
ODINOFAGIA (DOLOR DE GARGANTA AL TRAGAR ALIMENTOS)	SE IGNORA	0.12	0.23	0.01	-1.82	0.52	0.46	6	49	929	3997
POLIPNEA (RESPIRACION NORMAL)	SI	0.32	0.23	0.18	4.89	1.47	1.70	171	531	929	3997
POLIPNEA (RESPIRACION NORMAL)	NO	0.22	0.23	0.81	-1.68	0.73	0.65	751	3409	929	3997
POLIPNEA (RESPIRACION NORMAL)	SE IGNORA	0.12	0.23	0.01	-1.96	0.52	0.46	7	57	929	3997
RINORREA (NARIZ TAPADA)	SI	0.20	0.23	0.43	-3.73	0.74	0.67	400	2026	929	3997
RINORREA (NARIZ TAPADA)	NO	0.27	0.23	0.56	4.04	1.38	1.52	523	1928	929	3997
RINORREA (NARIZ TAPADA)	SE IGNORA	0.14	0.23	0.01	-1.44	0.60	0.53	6	43	929	3997
VOMITO	SI	0.24	0.23	0.08	0.41	1.05	1.06	77	318	929	3997
VOMITO	NO	0.23	0.23	0.91	0.12	1.04	1.05	845	3622	929	3997
VOMITO	SE IGNORA	0.12	0.23	0.01	-1.96	0.52	0.46	7	57	929	3997

Figura 2: Cuadro clínico de positivos de SARS-CoV-2 de los síntomas secundarios

En los síntomas “secundarios” asociados con los positivos se nota varios que son más discriminatorios que los “principales” y que pueden ser usados para propósitos de triage. Entre ellos se nota particularmente polipnea, mialgias y calofríos. Además se nota que la ausencia de rinorrea y odinofagia son predictivos de ser positivo, es decir si estos síntomas aparecen es menos probable que es coronavirus.

## Factores Demográficos

### Genero

GENERO	Probabilidad ser confirmado (P(C/X))	Verosimilitud (P(X/C))	Epsilon	RISK RATIO	ODDS RATIO	Número de confirmados de ese genero	Número de ese genero	Número confirmados	NTotal
FEMENINO	0.18	0.39	-5.99	0.61	0.52	365	2065	929	3997
MASCULINO	0.29	0.61	6.19	1.65	1.92	564	1932	929	3997

Figura 3: Riesgo relativo de hombres/mujeres ser positivos de SARS-CoV-2

Los hombres son 2 veces más que las mujeres ser diagnosticado con SARS-CoV-2 una vez probado

## Edad

Se nota en Figura 4 claramente el aumento en riesgo de ser positivo de SARS-CoV-2 como función de edad con un aumento en riesgo de 88% para ellos en el rango de edad 63-95.

Edad	Probabilidad ser confirmado con esa edad ( $P(C/X)$ )	Probabilidad ser confirmado	Verosimilitud ( $P(X/C)$ )	Epsilon	RISK RATIO	ODDS RATIO	Número de confirmados con esa edad	Número con esa edad	Número confirmados	NTotal
0-23	0.10	0.23	0.04	-6.33	0.40	0.33	39	397	929	3997
23-28	0.14	0.23	0.06	-4.35	0.59	0.52	59	415	929	3997
28-31	0.18	0.23	0.09	-2.63	0.75	0.70	79	440	929	3997
31-35	0.18	0.23	0.08	-2.76	0.74	0.68	77	436	929	3997
35-39	0.23	0.23	0.10	0.03	1.00	1.00	93	399	929	3997
39-44	0.23	0.23	0.10	0.10	1.01	1.01	91	388	929	3997
44-49	0.29	0.23	0.12	2.70	1.29	1.40	110	378	929	3997
49-55	0.33	0.23	0.14	4.47	1.48	1.72	126	383	929	3997
55-63	0.32	0.23	0.13	4.10	1.44	1.66	119	369	929	3997
63-95	0.35	0.23	0.15	5.37	1.58	1.88	136	392	929	3997

Figura 4: Riesgo relativo de ser positivos de SARS-CoV-2 como función de edad

## Factores Socio-económicos

En Figura 5 se nota el riesgo como función de ocupación. La categoría de Jubilado/Pensionado es claramente vinculado con edad. Se agrada ver que hasta el momento el grado de riesgo de distintos trabajadores de salud no ha sido significativamente menor que el promedio. Vale notar que Choferes y Maestros tienen una alta probabilidad de ser positivos sin embargo los números de estas categorías no son suficientemente altos para ver estadísticamente significantes. Sin embargo, será importante monitorear la situación dado que ambas ocupaciones típicamente requieren contacto frecuente con otros.

Ocupacion	Probabilidad ser confirmado con síntoma ( $P(C/X)$ )	Probabilidad ser confirmado	Verosimilitud ( $P(X/C)$ )	Epsilon	RISK RATIO	ODDS RATIO	Número de confirmados con esa ocupacion	Número con esa ocupacion	Número confirmados	NTotal
CAMPESINOS	0.00	0.23	0.00	-0.95	0.00	0.00	0	3	929	3997
CHOFERES	0.31	0.23	0.02	1.29	1.33	1.48	16	52	929	3997
COMERCIANTES DE MERCADOS FIJOS O AMBULANTES	0.29	0.23	0.05	1.82	1.27	1.38	49	168	929	3997
DENTISTAS	0.21	0.23	0.00	-0.16	0.92	0.90	3	14	929	3997
DESEMPELADOS	0.11	0.23	0.02	-3.23	0.48	0.41	15	132	929	3997
EMPLEADOS	0.22	0.23	0.19	-1.11	0.91	0.89	174	806	929	3997
ENFERMERAS	0.18	0.23	0.05	-2.14	0.74	0.69	44	251	929	3997
ESTUDIANTES	0.16	0.23	0.04	-2.94	0.65	0.59	40	258	929	3997
GERENTES O PROPIETARIOS DE EMPRESAS O NEGOCIOS	0.39	0.23	0.02	2.57	1.68	2.11	19	49	929	3997
HOGAR	0.28	0.23	0.10	2.25	1.25	1.35	96	338	929	3997
JUBILADO / PENSIONADO	0.39	0.23	0.05	4.08	1.70	2.14	48	124	929	3997
LABORATORISTAS	0.18	0.23	0.00	-0.55	0.76	0.71	3	17	929	3997
MAESTROS	0.32	0.23	0.01	1.12	1.39	1.57	9	28	929	3997
MEDICOS	0.15	0.23	0.10	-4.95	0.60	0.53	90	609	929	3997
OBREROS	0.18	0.23	0.00	-0.40	0.78	0.73	2	11	929	3997
OTROS	0.28	0.23	0.19	2.56	1.23	1.32	172	624	929	3997
OTROS PROFESIONISTAS	0.33	0.23	0.14	4.56	1.49	1.73	129	391	929	3997
OTROS TRABAJADORES DE LA SALUD	0.16	0.23	0.02	-1.79	0.70	0.64	20	122	929	3997

Figura 5: Riesgo relativo de ser positivos de SARS-CoV-2 como función de ocupación

## Antecedentes de salud

### Comorbilidades

En Figura 6 se ve el vínculo entre la presencia de las comorbilidades en los pacientes y la probabilidad de ser probado positivo para SARS-CoV-2. Interesantemente, ser asmático es un factor “protector” en que personas con asma son mucho menos probables ser positivos. Sospecho que esto puede ser porque pueden tener síntomas similares – como dificultad en respirar – y esto los lleva a querer reportar síntomas análogos a SARS-CoV-2 aunque son debidos a su asma. Hipertensión y diabetes son factores de riesgo, probablemente por tener sistemas inmunes menos eficientes.

Comorbilidad	Presencia de la comorbilidad	Probabilidad ser confirmado con comorbilidad (P(C/X))	Probabilidad ser confirmado	Verosimilitud (P(X/C))	Epsilon	RISK RATIO	ODDS RATIO	Número de confirmados con la comorbilidad	Número con la comorbilidad	Número confirmados	NTotal
ASMA	SI	0.14	0.23	0.03	-3.33	0.58	0.51	31	224	929	3997
ASMA	NO	0.24	0.23	0.97	0.81	1.72	1.94	898	3773	929	3997
DIABETES	SI	0.28	0.23	0.12	2.19	1.23	1.31	112	402	929	3997
DIABETES	NO	0.23	0.23	0.88	-0.73	0.82	0.76	817	3595	929	3997
ENFCARDI	SI	0.16	0.23	0.02	-1.88	0.67	0.61	18	114	929	3997
ENFCARDI	NO	0.23	0.23	0.98	0.32	1.49	1.63	911	3883	929	3997
EPOC	SI	0.24	0.23	0.02	0.26	1.05	1.07	21	86	929	3997
EPOC	NO	0.23	0.23	0.98	-0.04	0.95	0.94	908	3911	929	3997
HIPERTEN	SI	0.31	0.23	0.19	4.44	1.42	1.61	174	558	929	3997
HIPERTEN	NO	0.22	0.23	0.81	-1.78	0.71	0.62	755	3438	929	3997
HIPERTEN	SE IGNORA	0.00	0.23	0.00	-0.55	0.00	0.00	0	1	929	3997
INMUSUPR	SI	0.11	0.23	0.02	-3.40	0.45	0.39	14	131	929	3997
INMUSUPR	NO	0.24	0.23	0.98	0.63	2.21	2.59	915	3866	929	3997
INSRENCR	SI	0.19	0.23	0.01	-0.81	0.82	0.78	13	68	929	3997
INSRENCR	NO	0.23	0.23	0.99	0.11	1.22	1.29	916	3929	929	3997
OBESIDAD	SI	0.25	0.23	0.15	0.99	1.09	1.12	141	564	929	3997
OBESIDAD	NO	0.23	0.23	0.85	-0.40	0.92	0.89	788	3433	929	3997
OTRACON	SI	0.20	0.23	0.06	-1.28	0.85	0.82	58	289	929	3997
OTRACON	NO	0.23	0.23	0.94	0.37	1.17	1.23	871	3707	929	3997
OTRACON	SE IGNORA	0.00	0.23	0.00	-0.55	0.00	0.00	0	1	929	3997
TABAQUIS	SI	0.23	0.23	0.12	-0.37	0.97	0.96	110	488	929	3997
TABAQUIS	NO	0.23	0.23	0.88	0.14	1.04	1.05	819	3509	929	3997
VIH/SIDA	SI	0.26	0.23	0.02	0.55	1.13	1.18	15	57	929	3997
VIH/SIDA	NO	0.23	0.23	0.98	-0.07	0.88	0.85	914	3940	929	3997

Figura 6: Riesgo relativo de ser positivo de SARS-CoV-2 como función de las comorbilidades presentes en el paciente

### Modelo 1)ii) (Difuntos)

#### Factores de riesgo

#### Factores Demográficos

#### Edad

En Figura 7 se ve el riesgo relativo de morir de SARS-CoV-2 como función de edad. Como uno espera el riesgo relativo es muy elevado para gente mayor de 55. Interesantemente la mortalidad es mayor para el grupo de 55-63 comparado con el grupo de 63-95. Sin embargo todavía las estadísticas no son muy amplias.

Edad	Probabilidad ser difunto con esa edad ( $P(C/X)$ )	Probabilidad ser difunto	Verosimilitud ( $P(X/C)$ )	Epsilon	RISK RATIO	ODDS RATIO	Número de difuntos con esa edad	Número con esa edad	Número difuntos	NTotal
0-23	0.00	0.04	0.00	-1.32	0.00	0.00	0	39	40	929
23-28	0.00	0.04	0.00	-1.63	0.00	0.00	0	59	40	929
28-31	0.01	0.04	0.03	-1.33	0.28	0.27	1	79	40	929
31-35	0.03	0.04	0.05	-0.74	0.58	0.57	2	77	40	929
35-39	0.01	0.04	0.03	-1.53	0.23	0.22	1	93	40	929
39-44	0.07	0.04	0.15	1.08	1.63	1.67	6	91	40	929
44-49	0.04	0.04	0.10	-0.35	0.83	0.82	4	110	40	929
49-55	0.02	0.04	0.08	-1.06	0.52	0.50	3	126	40	929
55-63	0.11	0.04	0.33	3.56	3.28	3.56	13	119	40	929
63-95	0.07	0.04	0.25	1.75	1.94	2.02	10	136	40	929

Figura 7: Riesgo relativo de ser difunto de SARS-CoV-2 como función de edad

## Antecedentes de salud

### Comorbilidades

En Figura 8 se ve el impacto verdadero de la presencia de comorbilidades en el riesgo de mortalidad. Diabetes por ejemplo se asocia con un riesgo excesivo de 390%, mientras para EPOC es 679% y para Insuficiencia renal es 1473%. Hasta con obesidad misma tiene un riesgo elevado de 223%. Estos son números muy grandes y significa que estas personas necesitan ser protegidas, especialmente en los grupo socio-económicamente vulnerables.

Comorbilidad	Presencia de la comorbilidad	Probabilidad ser difunto con síntoma ( $P(C/X)$ )	Probabilidad ser difunto	Verosimilitud ( $P(X/C)$ )	Epsilon	RISK RATIO	ODDS RATIO	Número de difuntos con la comorbilidad	Número con la comorbilidad	Número difuntos	NTotal
ASMA	SI	0.00	0.04	0.00	-1.18	0.00	0.00	0	31	40	929
ASMA	NO	0.04	0.04	1.00	0.22	-	-	40	898	40	929
DIABETES	SI	0.13	0.04	0.38	4.74	4.38	4.90	15	112	40	929
DIABETES	NO	0.03	0.04	0.63	-1.75	0.23	0.20	25	817	40	929
ENFCARDI	SI	0.06	0.04	0.03	0.26	1.30	1.32	1	18	40	929
ENFCARDI	NO	0.04	0.04	0.98	-0.04	0.77	0.76	39	911	40	929
EPOC	SI	0.24	0.04	0.13	4.40	6.18	7.79	5	21	40	929
EPOC	NO	0.04	0.04	0.88	-0.67	0.16	0.13	35	908	40	929
HIPERTEN	SI	0.09	0.04	0.38	2.80	2.60	2.75	15	174	40	929
HIPERTEN	NO	0.03	0.04	0.63	-1.35	0.38	0.36	25	755	40	929
INMUSUPR	SI	0.14	0.04	0.05	1.84	3.44	3.85	2	14	40	929
INMUSUPR	NO	0.04	0.04	0.95	-0.23	0.29	0.26	38	915	40	929
INSRENCR	SI	0.38	0.04	0.13	6.07	10.07	15.73	5	13	40	929
INSRENCR	NO	0.04	0.04	0.88	-0.72	0.10	0.06	35	916	40	929
OBESIDAD	SI	0.10	0.04	0.35	3.29	3.01	3.23	14	141	40	929
OBESIDAD	NO	0.03	0.04	0.65	-1.39	0.33	0.31	26	788	40	929
OTRACON	SI	0.05	0.04	0.08	0.33	1.22	1.23	3	58	40	929
OTRACON	NO	0.04	0.04	0.93	-0.08	0.82	0.81	37	871	40	929
TABAQUIS	SI	0.03	0.04	0.08	-0.82	0.60	0.59	3	110	40	929
TABAQUIS	NO	0.05	0.04	0.93	0.30	1.66	1.69	37	819	40	929
VIH/SIDA	SI	0.00	0.04	0.00	-0.82	0.00	0.00	0	15	40	929
VIH/SIDA	NO	0.04	0.04	1.00	0.11	-	-	40	914	40	929

Figura 8: Riesgo relativo de ser difunto de SARS-CoV-2 como función de las comorbilidades presentes en el paciente

## Desempeño de los Modelos Predictivos

Un aspecto importante de los modelos predictivos es como tan precisos son en sus predicciones y como pueden ser usados para ayudar a la SEDESA. Abajo en Figura 9 se ve el desempeño del modelo predictivo para predecir quien morirá entre ellos quienes han sido probados positivos para SAR-Cov-2. La proporción de difuntos es de 4% entonces se puede notar que el modelo puede predecir quien morirá hasta 600% mejor que el modelo basal. En la gráfica el eje horizontal se mide el score que es un resultado del modelo Bayesiano. En este ejemplo, 20 representa el score máximo y 1 el score mínimo. Se ve claramente que el % de aciertos se incrementa con el score, es decir que el modelo predice exitosamente quien morirá.



Figura 9: Desempeño del modelo que predice mortalidad entre ellos positivos para SAR-Cov-2.

En Figura 10 se ve el desempeño del modelo que predice quien será positivo entre ellos quienes recibieron la prueba. La proporción de positivos entre los probados es 25.9% y el modelo llega a una probabilidad predicha de 80% para los pacientes más probables ser positivos. En este caso el score máximo es normalizado a 50 y el score mínimo es 1.

Ambos modelos juntos demuestran el potencial de los modelos predictivos predecir los distintos estados clínicos y epidemiológicos y su evolución de uno al otro.

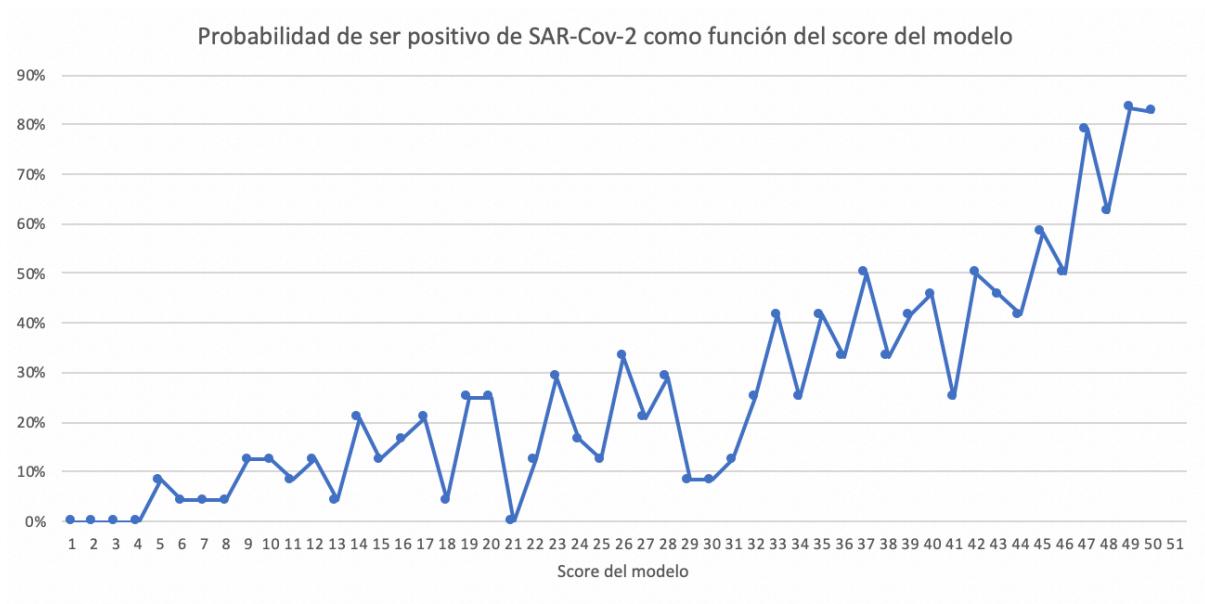


Figura 10: Desempeño del modelo que predice quien será positivo para la prueba de SAR-Cov-2.

## **Grupo Vulnerabilidad socioeconómica**

**Fecha: 10 de abril de 2020, Mérida, Yucatán**

### **Nota de avances**

#### **Integrantes:**

Dr. Jorge Luis Pérez - IIMAS-Mérida, UNAM

Dra. Maricarmen Jorge - IIMAS-Mérida, UNAM

Dra. María Josefa Santos - IIS, UNAM

Lic. Laura Calvo - IIS, UNAM

Dra. Rebeca de Gortari - IIS, UNAM

Dra. Yvon Angulo - IIS, UNAM

Dr. Carlos Valdés - UNAM, Centro de Estudios de México, Costa Rica

Dr. Jesús M. Siqueiros - IIMAS-Mérida, UNAM (Coordinador de grupo)

### **Proyecto general:**

Estudio para la generación de un mapa de vulnerabilidad socioeconómica frente a la crisis sanitaria del COVID-19

### **Pregunta central:**

¿Cuál es la distribución espacial de la vulnerabilidad socio-económica de la población del estado de Yucatán y/o nacional frente a la crisis del COVID-19?

### **Conceptos básicos (Definiciones nuestras):**

**Vulnerabilidad:** Es la (in)capacidad de una persona/familia/área etc., para responder ante la exposición a un riesgo (evento catastrófico). Mis referencias a la noción de vulnerabilidad vienen de adaptación al cambio climático (ver referencias):

**Vulnerabilidad socioeconómica:** La capacidad con la que cuenta una persona/familia/área etc., para responder a las presiones económicas provocadas por una contingencia --sanitaria en este caso.

### **Objetivos**

1. Identificar indicadores a partir de la encuesta que informen sobre las capacidades de respuesta y la vulnerabilidad socioeconómica de la población a las medidas de contención frente al COVID-19.
2. Generar con un mapa de la distribución de la vulnerabilidad socio-económica de la población del estado de Yucatán y posteriormente a escala nacional frente a la crisis del COVID-19.

### **Metas**

1. La meta de este proyecto es contar con un mapa georeferenciado de vulnerabilidad socioeconómica que sirva para informar a tomadores de decisiones qué zonas del estado o el país, se encuentran en una situación de menor capacidad para resistir durante los períodos de actividad económica baja debido a las medidas de contención sanitaria.

2. También se tiene como meta un producto que sea la base para el desarrollo de modelos de generación de escenarios que sirvan como heurística y reflexión sobre el impacto de distintas estrategias de mitigación sobre la componente socioeconómica.

**Organización del trabajo:**

El grupo de trabajo se ha dividido en dos sub-grupos para trabajar dos aspectos del proyecto en paralelo. Un primer grupo está conectado de la formulación de una encuesta para hacer un mapeo a escala de los hogares e identificar sus capacidades socioeconómicas de adaptación frente a las medidas de cuarentena y post-cuarentena.

**Grupo 1:**

**Proyecto:**

Encuesta para generar un mapa de vulnerabilidad socio-económica frente a la crisis sanitaria del COVID-19

El grupo está conformado por Maricarmen Jorge, María Josefa Santos, Yvon Angúlo, Jorge Luis Pérez, Rebeca de Gortari y Jesús M. Siqueiros

Este grupo está centrado en el diseño, elaboración e implementación de la encuesta. La encuesta consiste en 49 reactivos divididos en 6 categorías: 1) Perfil socio-demográfico; 2) riesgo de exposición al virus; 3) capacidades económicas; 4) capacidades sociales; 5) perfil de salud; 6) capacidad de respuesta. Link a cuestionario: <https://forms.gle/cMMyaAQmSjdAahif8>

El cuestionario será aplicado primero en el estado de Yucatán, posteriormente en la Ciudad de México y finalmente, se espera aplicarlo al resto del país.

**Avances:**

Diseño del cuestionario: 100%

Implementación en plataforma digital: 100%

Última revisión de formato a la encuesta: 70%

Redacción de protocolo para evaluación de comité de ética: 90%

Otros documentos para evaluación de comité de ética: 90%

Prueba piloto: 0%

Aplicación del cuestionario: 0%

El plan de trabajo para la semana entrante consiste:

- 1) Hacer la prueba piloto de la encuesta.
- 2) Simultáneamente, el protocolo pasará por el Comité de Ética del Centro Hideyo Noguchi, UADY —así está planeado-.
- 3) Diseño de estrategia para de difusión de la encuesta en el estado de Yucatán, CDMX y resto del país.

Grupo 2.

Identificación de indicadores a escala Municipal para mapa de vulnerabilidad.

El grupo está conformado por Laura Calvo, Carlos Valdés, Helena Gómez (asesoría) y Jesús M. Siqueiros

En foco de este grupo está en identificar en fuentes secundarias datos a nivel municipal para generar un mapa que de contexto socio-económico a los resultados de la encuesta y que en conjunto se pueda proponer una ponderación de vulnerabilidad para su posterior visualización en un mapa.

En este momento se tienen identificadas de manera preliminar una serie indicadores de infraestructura y otros de tipos social y económico (identificados inicialmente por Laura Calvo y María Josefa Santos). Estamos al tanto de que existen otros grupos con trabajo similar, y que es importante entrar en contacto con ellos.

**Resumen:**

El objetivo de este grupo es generar a través de una encuesta y datos de fuentes secundarias un índice de vulnerabilidad socio-económica con una representación gráfica en un mapa. Estos organizados en dos grupos. El grupo 1, que trabaja sobre la encuesta es el que va más avanzado y espera a inicios de la siguiente semana hacer la prueba piloto del mismo. El segundo grupo es más reciente y está concentrado en la identificación de variables vinculadas a nuestra pregunta de trabajo que describan la vulnerabilidad a escala municipal. En el orden de las cosas, iniciaremos con el estado de Yucatán, posteriormente con la CDMX y finalmente con el resto del país.

## **Informe de actividades - Análisis de Twitter**

### **Objetivos:**

- Crear un sistema automático de vigilancia Covid-19 mediante Twitter.
- Analizar mensajes de la red social Twitter, para evaluar el comportamiento de las personas, estados de ánimos, desinformación, popularidad de las medidas dadas por el gobierno, y además monitorear usuarios con posibles síntomas de coronavirus.

### **Grupo de Trabajo:**

- Gemma Bel Enguix - Instituto de Ingeniería
- Helena Gómez Adorno - Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas
- Gerardo Sierra Martínez - Instituto de Ingeniería
- Gabriel Castillo - Instituto de Ingeniería

### **Trabajos realizados**

Se ha comenzado a recolectar mensajes de la red social Twitter a partir del 1 de abril. Debido a la restricción de cantidad de mensajes que permite la API de Twitter, los mensajes se almacenan en 2 bases de datos tanto en el Instituto de Ingeniería como en el IIMAS.

En las últimas dos semanas se han recogido cerca de 1,300,000 tuits en cada servidor, lo que supone, después de la eliminación de los registros repetidos, unos 2,500,000 mensajes. Para el almacenamiento de estos datos se puso en funcionamiento un servidor de bases de datos no relacionales y un programa que con base en el API de Twitter recolecta los mensajes. Se está solicitando a la DGTIC la puesta en funcionamiento de un servidor público que permita poner en funcionamiento muchos más programas y que de manera concurrente se recolecten los mensajes a fin de tener una cantidad mayor por día. Todos los tuits han sido filtrados para que exclusivamente se consideren aquellos emitidos dentro de la república mexicana.

Por otra parte, se ha comenzado con el estudio de una muestra de registros que permitan orientar el diseño de aplicaciones.

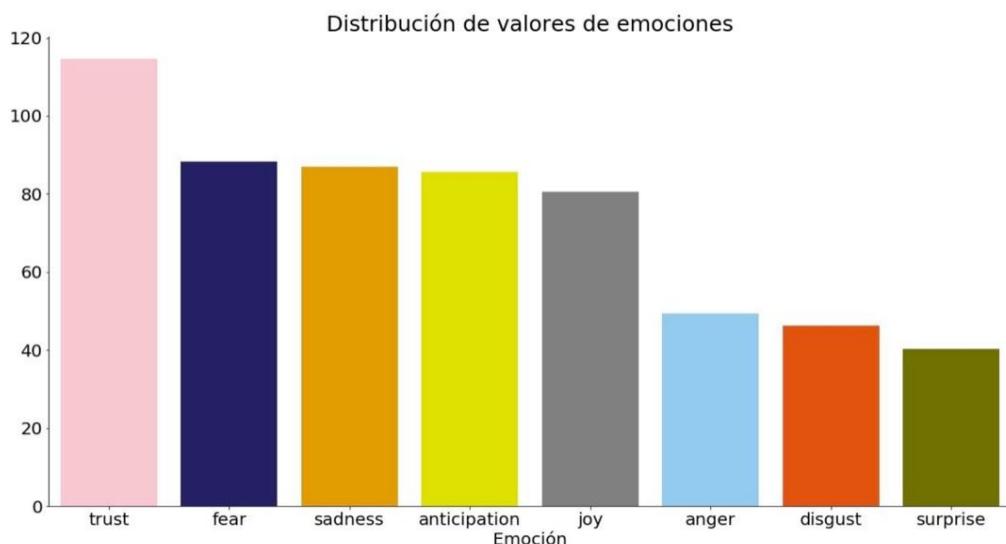
Para el sistema automático de vigilancia, la primera fase ha sido la de análisis. En este estadio se han analizado de forma manual 12.000 tuits con el fin de elaborar los siguientes repositorios:

- A. Diccionario de variantes de término relacionados con COVID-19, tanto científicas como coloquiales.

- B. Diccionario de hashtags relacionados con la enfermedad.
- C. Diccionario de sintomatología.
- D. Patrones que permitan reconocer estructuras que hacen referencia a síntomas o características de la enfermedad.

Con todo este trabajo previo realizado, durante esta semana se van a implementar programas que permitan detectar todos estos elementos automáticamente.

Para el análisis del comportamiento de usuarios en Twitter, se están evaluando métodos de clasificación automática de sentimientos basados en aprendizaje supervisado y basado en lexicones. Hemos evaluado los métodos con un conjunto preliminar de 7000 tweets aleatorios y los resultados indican una alta prevalencia de sentimiento de confianza por parte de los usuarios:



En paralelo estamos trabajando en el diseño de una página web donde se reflejen, en tiempo real todos los análisis realizados. La estructura de la página tendrá un mapa interactivo de la República mexicana, que permite la segmentación de información en los 32 estados que componen el país.

Además incluirá componentes denominados tarjetas los cuales facilitan la presentación de interés general.

Como herramienta de control de versiones utilizaremos Github que servirá además de repositorio para la página web y resultados de los análisis.

Se generará un cron en Linux que permita realizar la recolección de resultados en un formato fácil de interpretar (JSON por ejemplos) y subirlo al repositorio de datos.

Con esos archivos JSON, solo leerlos desde la página web. Lo anterior para no tener problemas con servicios entre servidores.