

Proyecto COVID-19. Revisión Literatura

Alejandro Aranguren 201416350

Edgar Alonso Ramírez Hernández 201710834

Johan Sebastián Carvajal Salamanca 201921325

1. Introducción

Los tratamientos antivirales o la vacunación no estarán disponibles en el corto plazo para detener el avance del COVID- 19, por lo tanto, la pandemia causada por este virus seguirá avanzando en todo el mundo. En la medida en que estas soluciones farmacéuticas sigan sin estar al alcance de la humanidad, la única forma de controlar la propagación del COVID- 19 es por medio de la implementación de políticas y protocolos de salud pública que reduzcan el riesgo de muerte y contagio por este coronavirus: aislamiento social, fortalecimiento de la infraestructura del sector salud, pruebas de diagnóstico focalizadas, teletrabajo, entre otros.

El análisis sobre cuales medidas deberían tomar los gobiernos es todavía incierto debido a la gran incertidumbre que existe frente a este coronavirus. Sin embargo, esto no ha impedido el avance en algunas aproximaciones para modelar y medir el avance de la pandemia sobre la población. Gran parte de los gobiernos y las agencias de salud a nivel mundial tienen acceso a modelos de predicción para tomar decisiones sobre la implementación y evaluación de estas políticas de salud pública para mitigar los efectos del virus. En ese sentido, el desempeño de los modelos resulta fundamental para salvar un número de significativo de vidas humanas.

Los esfuerzos para modelar la propagación del virus han sido múltiples. Por lo tanto, resulta importante realizar una revisión de los principales modelos matemáticos usados y que aplicaciones se han realizado con métodos de aprendizaje de máquinas para estudiar la dinámica del COVID-19. Los modelos matemáticos acompañados de algoritmos de aprendizaje de máquinas que pueden aprender de los datos sin depender de reglas preestablecidas, parece una opción sensata para seguir mejorando el desempeño de estos modelos. Esta es la motivación de este proyecto.

2. Revisión de literatura

Ante la llegada de la pandemia del SARS-CoV-2 se han intensificado los estudios cuantitativos que buscan modelar el contagio y la propagación del virus. El estudio juicioso de las tasas efectivas de transmisión y de contagio son un insumo fundamental para la toma de decisiones en materia de política sanitaria, permitiendo medir la efectividad de medidas como el aislamiento social, cuarentenas, pruebas focalizadas etc. Los esfuerzos para modelar la propagación del virus han sido múltiples y, por lo tanto, resulta importante realizar una revisión de los principales modelos usados para el caso de la COVID-19.

Entre los modelos epidemiológicos más usados para el estudio de enfermedades como la Covid-19 son modelos determinísticos basados en ecuaciones diferenciales ordinarias. La familia más grande de modelos de este tipo la componen los llamados modelos SIR, por las siglas, que indican los diferentes estados por los que puede pasar la población: Susceptibles, Infectados y Recuperados (o en otras referencias a los recuperados también se les conoce como la población removida o fallecida) [Giordano et al., 2020, Mohler et al., 2020, Manrique Abril et al., 2020].

Estos modelos analizan las transiciones de la población por cada uno de los estados, estimando el número teórico de personas susceptibles a la enfermedad, la cantidad de infectados y el número de personas que no pueden transmitir la enfermedad pues han logrado la inmunidad total contra la reinfección (recuperados). Este último estado del modelo SIR, también pueden ser interpretados como la población que ha sido aislada socialmente (cuarentena), que ha sido inmunizada contra la infección (vacunación) o la cantidad de fallecidos a causa de la enfermedad [Castro et al., 2020].

Aunque los modelos SIR son la forma estándar de estudiar el comportamiento de enfermedades y su propagación, presenta serias limitaciones, pues no considera otros estados por los que podría pasar la población en el transcurso del virus. En ese sentido, el modelo a utilizar dependerá de los agentes infecciosos por los que se transmita la enfermedad a modelizar, pues pueden variar de una enfermedad a otra. [Manrique Abril et al., 2020]

Bajo la anterior premisa, surge una extensión de estos modelos conocida como modelos SIER, en el cual se adiciona un nuevo estado que referencia a la población que está expuesta a la enfermedad; es decir, que la enfermedad se encuentra en un período de incubación durante el cual no pueden infectar a otros [Bertozzi et al., 2020, Castro et al., 2020].

Sin embargo, los modelos SIR, SIER y otras extensiones de ésta clase de modelos ¹, tienen grandes limitaciones cuando se aplican a una enfermedad como el Covid-19 por las particularidades que tiene la enfermedad en cuanto a lo que tiene que ver con la etapa

¹Para una mejor descripción de modelos epidemiológicos y de tipo determinísticos, consultar [García Rovira et al., 2017]

de incubación del virus, la población asintomática o con síntomas leves, la población no diagnosticada, dificultades en la detección y confirmación oportuna de los infectados, entre otros problemas [Castro et al., 2020, Bertozzi et al., 2020, Parra et al., 2020]. También, se debe tener en cuenta las medidas de política pública que se han implementado para mitigar el impacto de la enfermedad, principalmente las relacionadas con el aislamiento social [Castro et al., 2020, Mohler et al., 2020, Mahmud et al., 2020].

Por otro lado, están los modelos estocásticos como otra alternativa para modelar la propagación del Covid- 19. La principal diferencia entre los modelos determinísticos como el modelo SIR y sus demás variaciones y, los modelos estocásticos, radica en que con los primeros es posible controlar todos los factores que intervienen en el estudio del fenómeno, y predecir sus resultados con exactitud; por eso bajo este modelo, un solo individuo puede causar una epidemia generalizada; mientras que, en los estocásticos no es posible controlar todos los factores concurrentes, de manera que los resultados no son únicos, sino que cada uno se genera con una probabilidad determinada y, por lo tanto, existe la posibilidad de que la epidemia se extinga [García Rovira et al., 2017].

Modelos flexibles como los estocásticos también han sido implementados para estudiar el fenómeno es cuestión. El modelo Hawkes expuesto por [Bertozzi et al., 2020], se basa en un proceso de ramificación para la estimación del número de reproducción y escala de tiempo. El punto importante de este modelo es poder representar el comportamiento de la probabilidad de contagio (distribución Poisson). Igualmente, este modelo se ajusta a la tasa de mortalidad observada en los datos y las tendencias acumulativo de casos. Dentro de sus particularidades es que se puede estimar de forma paramétrica y no paramétrica, al igual que se puede entender como la versión estocástica del modelo SIR [Rizoiu et al., 2018].

De forma similar, [Giordano et al., 2020] utilizan un modelo denominado SIDARTHE (por las siglas en ingles de susceptibles (S), infectados (I), diagnosticados (D), enfermos (A), reconocidos (R), amenazados (T), recuperados (H) y los fallecidos (E)). Al clasificar a la población en estos ocho diferentes estados de la infección, tiene la ventaja de realizar modelos a más largo plazo y mostrar el efectos de políticas públicas de mitigación como lo son las del distanciamiento social, pruebas de diagnóstico focalizadas, ocupación en Unidades de Cuidado Intenso - UCI-, entre otras, que los modelos determinísticos con dificultad pueden modelar.

Los modelos de aprendizaje de maquinas también son una alternativa para modelar el comportamiento del Covid- 19. Los modelos matemáticos determinísticos y estocásticos que se han revisado son los que se utilizan con mayor frecuencia en la implementación de modelos epidemiológicos; sin embargo, debido a la complejidad y al problema de gran escala en el desarrollo de estos modelos, el aprendizaje de máquinas ha llamado la atención, pues con estos enfoques se pueden desarrollar modelos con mayor capacidad de generali-

zación y mayor confiabilidad de predicción con horizontes de tiempo más largos [Ardabili et al., 2020].

Los modelos de aprendizaje de máquinas ya habían sido implementados para estudiar la propagación de otras enfermedades: para modelar el dengue se han utilizado las redes neuronales [Anno et al., 2019], redes neuronales bayesianas [Raja et al., 2019] y el algoritmo LogitBoost [Iqbal and Islam, 2019]. De igual forma, se había implementado Random Forest para modelar la gripa porcina [Liang et al., 2019] y el virus de la influenza [Tapak et al., 2019].

En cuanto al desarrollo de modelos de aprendizaje de máquinas para estudiar la propagación del Covid- 19, se identifica una brecha en relación con los desarrollos hechos con modelos matemáticos determinísticos y estocásticos, principalmente debido a que es muy reciente la aplicación del aprendizaje de maquinas al modelamiento de la propagación de enfermedades [Ardabili et al., 2020].

Hasta el momento, la literatura de modelos de aprendizaje de máquinas para modelar la propagación del Covid- 19, se concentra en realizar ejercicios comparativos con los modelos de la familia SIR [Mbaye Ndiaye et al., 2020, Baldé, 2020] y testear el desempeño de distintos algoritmos de aprendizaje de máquinas utilizando medidas de precisión y de ajuste como RSME, el valor Akaike y otras metricas de desempeño según los modelos testeados [Punn et al., 2020, Ardabili et al., 2020, Ribeiro et al., 2020].

Los modelos predictivos de esta clase tienen resultados variables, pues en algunos ejercicios muestran buen desempeño los algoritmos [Mbaye Ndiaye et al., 2020, Ardabili et al., 2020] y en otros trabajos el desempeño de estos modelos como el de Random Forest o la Ridge Regression no tienen buen desempeño en las métricas [Ribeiro et al., 2020]. Sin embargo, se resalta con mayor frecuencia el desempeño de los algoritmo SVR (Support Vector Regression) [Ribeiro et al., 2020] y PR (Polynomial Regression) [Punn et al., 2020] en los resultados predictivos de la propagación del Covid- 19.

Los trabajos que combinan modelos matemáticos de la familia SIR y algoritmos de aprendizaje de maquinas son aun menos frecuentes. Aunque la necesidad de utilizar estos algoritmos para reducir el sesgo que tienen los modelos determinísticos en relación con los cálculos que realizan de la población expuesta, infectada y recuperada provocados por forma de recolección de información de las agencias de salud [Ardabili et al., 2020, Zou et al., 2020] desarrollan un modelo SIER que es complementado con el algoritmo de aprendizaje de máquinas Gradiente de Optimización para estimar los valores óptimos de los parámetros de acuerdo con la minimización del MSE. Este modelo SuEIR, como lo han llamado los autores, evita el problema de los modelos determinísticos de no considerar el subreporte de la población infectada, haciendo de esta estimación aun más precisa [Zou et al., 2020].

3. Descripción de los datos

La principal fuente de información sobre la propagación del COVID- 19 en Colombia es el Instituto Nacional de Salud- INS (tabla 1). En la página web, a diario actualizan el número de contagiados por ciudad y algunas características poblaciones de los contagiados. Así mismo actualizan el estado de las personas enfermas, indicando si son si tienen la enfermedad activa, si se han recuperado o si han fallecido a causa del virus.

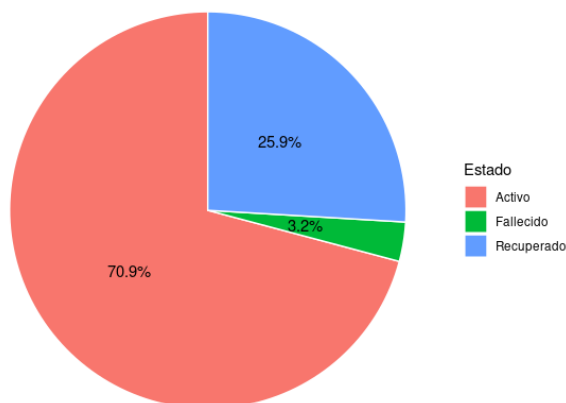
Tabla 1. Variables identificadas para modelar el COVID- 19 en Colombia

Variable	Periodicidad	Tipo de variable	Fuente
Tasa de contagio	Diaria	Numérica (conteo)	INS
Tasa de mortalidad	Diaria	Numérica (conteo)	INS
Tasa de recuperación	Diaria	Numérica (conteo)	INS

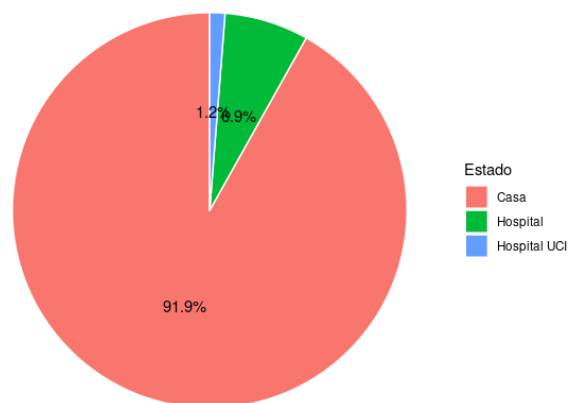
De acuerdo con el Instituto Nacional de Salud, autoridad científica y técnica encargada de registrar el avance del COVID- 19 en Colombia, había 26.688 casos confirmados, 6.913 recuperados y 853 fallecidos registrados al 28 de mayo de 2020. En la figura 1 se observa que el porcentaje de casos activos que requieren atención hospitalaria o cuidados en Unidad de Cuidados Intensivos (UCI) es mínimo si se compara con el total de activos.

Figura 1: Distribución de casos confirmados por estado

Distribución de confirmados por Estado



Distribución de activos por Estado

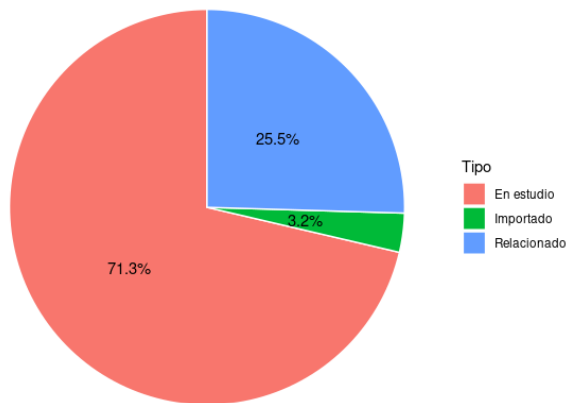


Fuente: Instituto Nacional de Salud

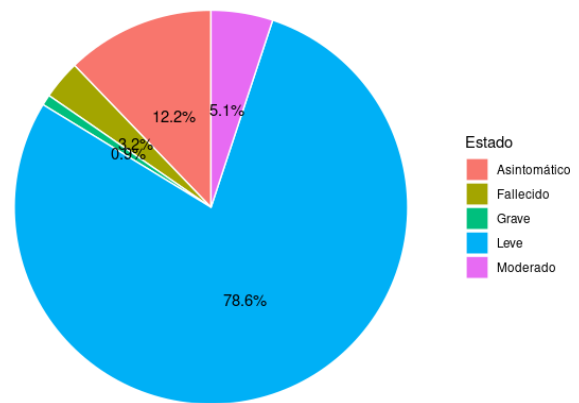
En la figura 2 se presenta la distribución de los casos confirmados por tipo y por nivel gravedad del caso. Como se puede observar a la fecha del presente trabajo, la mayoría de casos (78.6 %) presentan síntomas leves mientras que el 12.2 % son asintomáticos.

Figura 2: Distribución de casos por tipo

Distribución de confirmados por Tipo



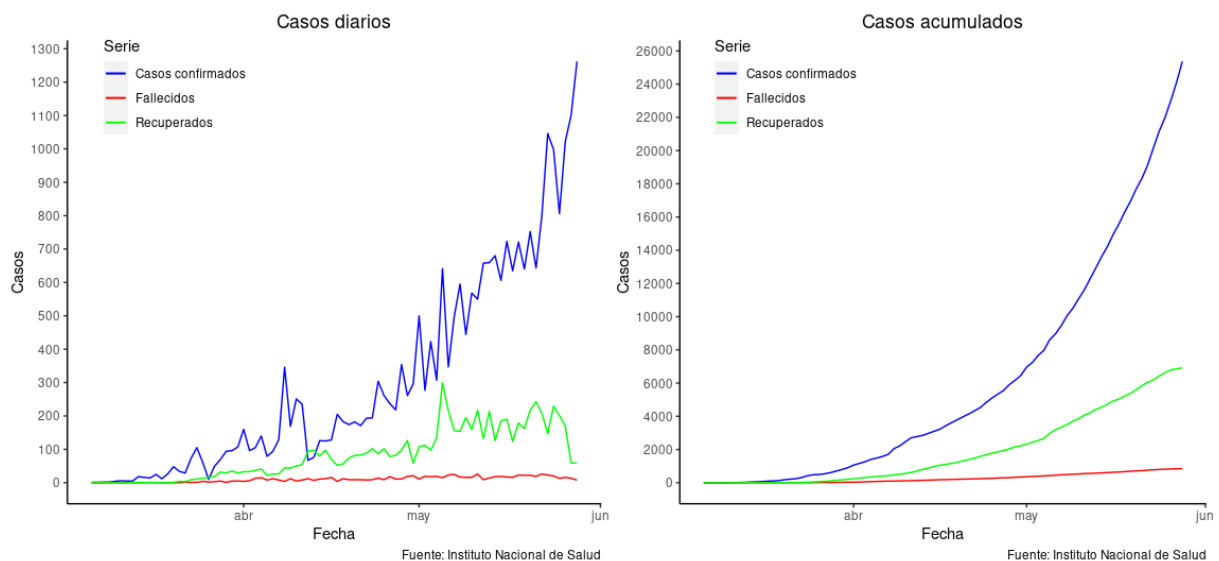
Distribución de confirmados por Estado de Gravedad



Fuente: Instituto Nacional de Salud

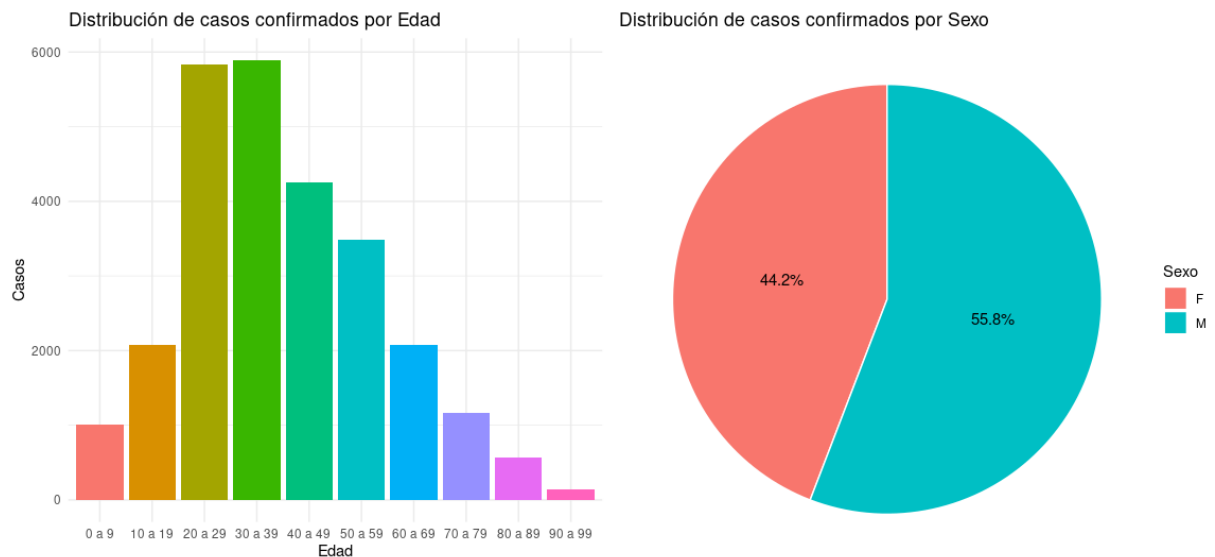
La figura 3 presenta la evolución de los casos, diarios y acumulados, en Colombia desde el 6 de marzo de 2020, día en el que se confirmó el primer diagnóstico de Covid-19 en el país. Pasados aproximadamente tres meses, desde el primer diagnóstico, el número de contagiados no para de crecer en Colombia. La figura 3 muestra también la evolución de los recuperados y de los fallecidos. Estas series también presentan una tendencia creciente a lo largo del tiempo.

Figura 3: Casos confirmados en Colombia



Existen variables demográficas que permiten identificar factores diferenciadores entre la población como: grupo de edad, sexo o ciudad de ubicación de los contagiados. A continuación se presentan gráficas descriptivas para los grupos de confirmados, fallecidos y recuperados. En la figura 4 se presenta la gráfica del consolidado de casos confirmados en Colombia.

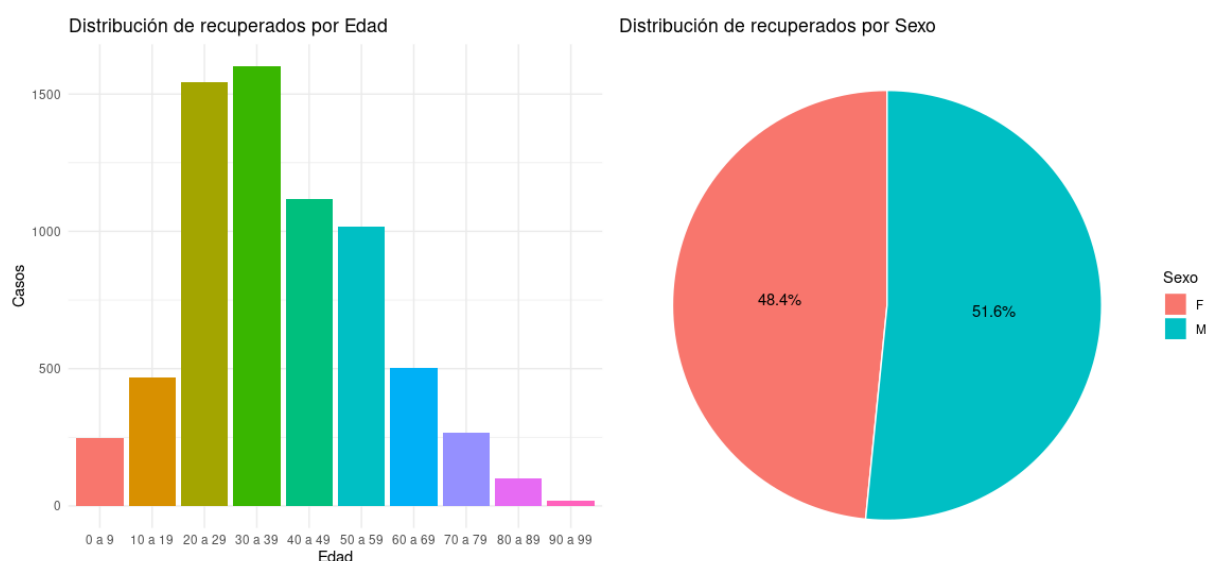
Figura 4: Casos confirmados por edad y sexo



Fuente: Instituto Nacional de Salud

En la gráfica 5 se observa que la distribución por edad y sexo es similar al del total de casos confirmados.

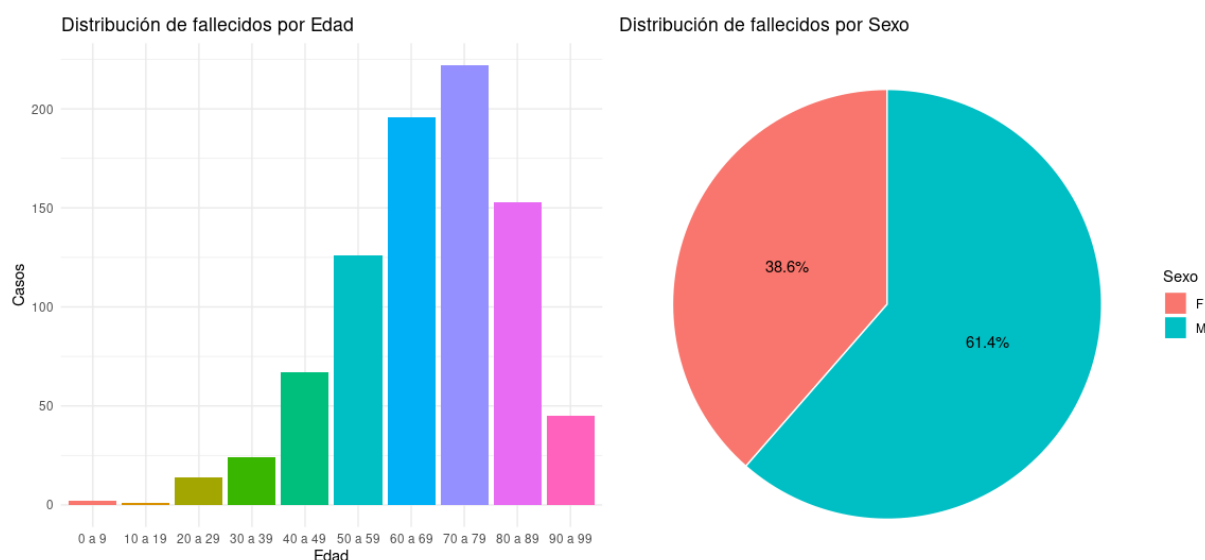
Figura 5: Recuperados por edad y sexo



Fuente: Instituto Nacional de Salud

En la gráfica 6 se observa que la distribución de edad de los fallecidos está claramente centrada en la población de mayor edad, siendo la mayoría de fallecidos con edades en los grupos de edades de personas mayores de 60 años. Así mismo, resalta que una gran proporción de fallecidos son hombres.

Figura 6: Recuperados por edad y sexo



Fuente: Instituto Nacional de Salud

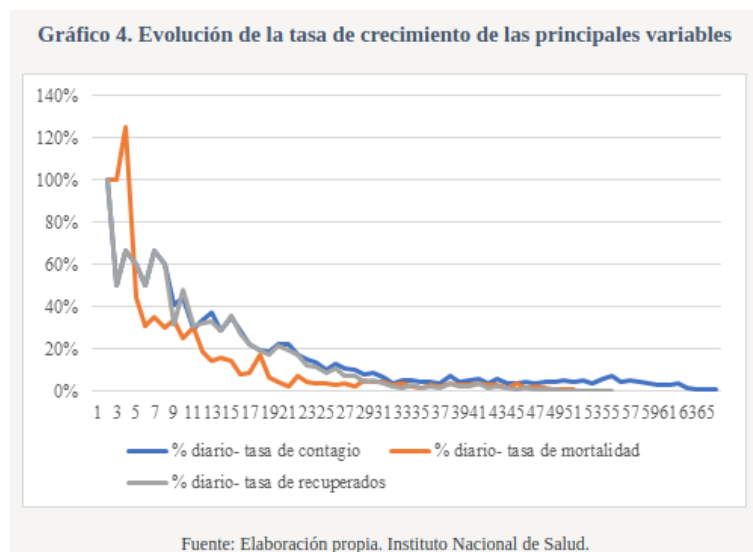
Aunque las características presentadas anteriormente son importantes para comprender el panorama del virus en el país, gran parte de la literatura desconsidera dichos atributos. Los modelos matemáticos adoptan supuestos sobre las características, o atributos de los parámetros que caracterizan a la población: todos los individuos de la población se comportan de la misma manera y toda la población es susceptible en la misma medida.

Tabla 2. Porcentaje de crecimiento promedio de las principales variables según el periodo de referencia

Variable	Crecimiento diario	Crecimiento cada tercer día	Crecimiento cada quinto día
Tasa de contagio	16,8%	38,5%	107,2%
Tasa de mortalidad	15,1%	36,2%	93,6%
Tasa de recuperación	17,5%	40,4%	115,5%

Finalmente, la tabla 2 y el gráfico 4 resumen las descriptivas de la dinámica del COVID- 19 en Colombia. Se observa como la tasa de recuperados empieza a superar en crecimiento a la tasa de contagio, indicando que Colombia se acerca al denominado “pico de la curva de contagio”, donde el crecimiento de contagiados empieza a reducirse sustancialmente (tabla 2). Así mismo, se identificó un aumento en la tasa de mortalidad en los primeros días, sin embargo, después se observa un descenso dramático (gráfico 4). De igual forma se identifica que la de tasa recuperados sigue muy de cerca la tasa de

contagiados.



La información presentada en esta sección tiene fecha de corte del 28 de mayo de 2020. En general se puede observar que las tasas de crecimiento promedio han descendido significativamente en relación con el comienzo de la pandemia en el país. Aunque esto parece indicar que la dinámica del virus en el país empieza llegar a su máximo, hasta que no se desarrolle una vacuna o un tratamiento antiviral, los modelos y las estadísticas seguirán siendo la única forma de evaluar la implementación de políticas de salud pública orientadas a la reducción del riesgo de contagio y de muerte por el coronavirus en el país.

Referencias

- Anno, S., Hara, T., Kai, H., Lee, M.-A., Chang, Y., Oyoshi, K., Mizukami, Y., and Tandon, T. (2019). Spatiotemporal dengue fever hotspots associated with climatic factors in taiwan including outbreak predictions based on machine-learning. *Geospatial health*, 14(2).
- Ardabili, S. F., Mosavi, A., Ghamisi, P., Ferdinand, F., Varkonyi-Koczy, A. R., Reuter, U., Rabczuk, T., and Atkinson, P. M. (2020). Covid-19 outbreak prediction with machine learning. *Available at SSRN 3580188*.
- Baldé, M. A. M. T. (2020). Fitting sir model to covid-19 pandemic data and comparative forecasting with machine learning. *medRxiv*.
- Bertozi, A. L., Franco, E., Mohler, G., Short, M. B., and Sledge, D. (2020). The challenges of modeling and forecasting the spread of COVID-19. page 19.
- Castro, P., De los Reyes, J. C., González, S., Merino, P., and Ponce, J. (2020). Modelización y Simulación de la propagación del virus SARS-COV-2 en Ecuador. *Escuela Politécnica Nacional de Ecuador*, page 12.
- García Rovira, L. et al. (2017). Modelos matemáticos compartimentales en epidemiología.

- Giordano, G., Blanchini, F., Bruno, R., Colaneri, P., Di Filippo, A., Di Matteo, A., and Colaneri, M. (2020). Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nature Medicine*.
- Iqbal, N. and Islam, M. (2019). Machine learning for dengue outbreak prediction: A performance evaluation of different prominent classifiers. *Informatica*, 43(3).
- Liang, R., Lu, Y., Qu, X., Su, Q., Li, C., Xia, S., Liu, Y., Zhang, Q., Cao, X., Chen, Q., et al. (2019). Prediction for global african swine fever outbreaks based on a combination of random forest algorithms and meteorological data. *Transboundary and emerging diseases*.
- Mahmud, M. S., Kamrujjaman, M., Jubyrea, J., Islam, M. S., and Islam, M. S. (2020). Quarantine vs social consciousness: A prediction to control covid-19 infection. *Journal of Applied Life Sciences International*, pages 20–27.
- Manrique Abril, F., González-Chordá, V. M., Gutiérrez Lesmes, O. A., Tellez Piñerez, C. F., Herrera-Amaya, G. M., et al. (2020). Modelo sir de la pandemia de covid-19 en colombia.
- Mbaye Ndiaye, B., Tendeng, L., and Seck, D. (2020). Analysis of the covid-19 pandemic by sir model and machine learning technics for forecasting. *arXiv*, pages arXiv–2004.
- Mohler, G., Schoenberg, F., Short, M. B., and Sledge, D. (2020). Analyzing the World-Wide Impact of Public Health Interventions on the Transmission Dynamics of COVID-19. page 6.
- Parra, N., Vargas, V., Florez, J. S., Ardila, L., and Viviescas, C. (2020). Delay-adjusted case fatality ratio and under-reporting estimates in Colombia. page 4.
- Punn, N. S., Sonbhadra, S. K., and Agarwal, S. (2020). Covid-19 epidemic analysis using machine learning and deep learning algorithms. *medRxiv*.
- Raja, D. B., Mallol, R., Ting, C. Y., Kamaludin, F., Ahmad, R., Ismail, S., Jayaraj, V. J., and Sundram, B. M. (2019). Artificial intelligence model as predictor for dengue outbreaks. *Malaysian Journal of Public Health Medicine*, 19(2):103–108.
- Ribeiro, M. H. D. M., da Silva, R. G., Mariani, V. C., and dos Santos Coelho, L. (2020). Short-term forecasting covid-19 cumulative confirmed cases: Perspectives for brazil. *Chaos, Solitons & Fractals*, page 109853.
- Rizoiu, M.-A., Mishra, S., Kong, Q., Carman, M., and Xie, L. (2018). Sir-hawkes: linking epidemic models and hawkes processes to model diffusions in finite populations. In *Proceedings of the 2018 World Wide Web Conference*, pages 419–428.
- Tapak, L., Hamidi, O., Fathian, M., and Karami, M. (2019). Comparative evaluation of time series models for predicting influenza outbreaks: application of influenza-like illness data from sentinel sites of healthcare centers in iran. *BMC research notes*, 12(1):353.
- Zou, D., Wang, L., Xu, P., Chen, J., Zhang, W., and Gu, Q. (2020). Epidemic model guided machine learning for covid-19 forecasts in the united states. *medRxiv*.