



UNIVERSIDAD AUTÓNOMA DEL ESTADO DE HIDALGO

INSTITUTO DE CIENCIAS DE LA SALUD

Maestría en Ciencias Biomédicas y de la Salud

TESIS

**DESARROLLO Y VALIDACIÓN EPIDEMIOLÓGICA DE UN
MODELO DE PREDICCIÓN PARA PACIENTES COVID-19
EN ÁLVARO OBREGÓN, DE LA CIUDAD DE MÉXICO**

**Para obtener el título de
Maestra en Ciencias Biomédicas y de la Salud**

PRESENTA

Ing. Mariel Victorino Aguilar

**Director (a)
Dr. Abel Lerma Talamantes**

**Codirectora
Dra. Claudia Lerma González**

**Comité tutorial
Dra. Rebeca María Elena Guzmán Saldaña
Dra. Lilián Elizabeth Bosques Brugada
Dr. Luis Israel Ledesma Amaya**

Pachuca de Soto, Hgo., México., septiembre 2023



07/09/2023

Asunto: Autorización de Impresión

Mtra. Ojuky del Rocío Islas Maldonado
Directora de Administración Escolar
Presente.

El Comité Tutorial de la TESIS del programa educativo de posgrado titulada "Desarrollo y validación epidemiológica de un modelo de predicción para pacientes COVID-19 en Álvaro Obregón, de la ciudad de México", realizado por la sustentante Mariel Victorino Aguilar con número de cuenta 244818 perteneciente al programa de Maestría en Ciencias Biomédicas y de la Salud, una vez que ha revisado, analizado y evaluado el documento recepcional de acuerdo a lo estipulado en el Artículo 110 del Reglamento de Estudios de Posgrado, tiene a bien extender la presente:

AUTORIZACIÓN DE IMPRESIÓN


Por lo que la sustentante deberá cumplir los requisitos del Reglamento de Estudios de Posgrado y con lo establecido en el proceso de grado vigente.


Atentamente

"Amor, Orden y Progreso"


Pachuca, Hidalgo a 07 de septiembre de 2023

El Comité Tutorial



 Abel Lerma Talamantes
 Director

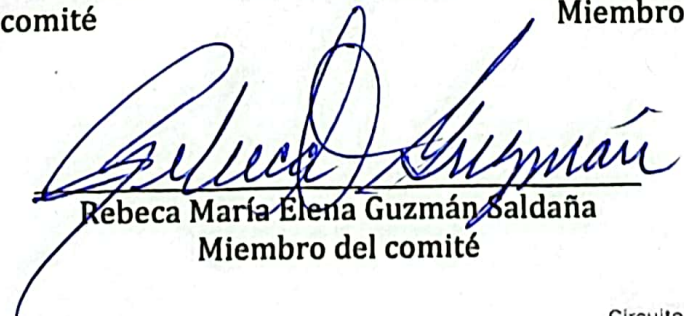

 Claudia Lerma González
 Codirectora




 Luis Israel Ledesma Amaya
 Miembro del comité

ÁREA ACADÉMICA DE MEDICINA


 Lilian Elizabeth Bosques Brugada
 Miembro del comité


 Rebeca María Elena Guzmán Saldaña
 Miembro del comité

Círculo ex-Hacienda La Concepción s/n Carretera
 Pachuca Actopan, San Agustín Tlaxiaca, Hidalgo,
 México. C.P. 42160
 Teléfono: 52 (771) 71 720 00 Ext. 4308,2361,4346,4310
 medicina@uaeh.edu.mx



www.uaeh.edu.mx

Durante el desarrollo de estos estudios, se contó con una beca de manutención otorgada por el Consejo Nacional de Ciencia y Tecnología (CONAHCyT), número de beca 797862 y número de CVU 1146841.

Durante el periodo febrero- abril de 2023 se realizó una estancia nacional de investigación en la Facultad del Departamento de Instrumentación Electromecánica del Instituto Nacional de Cardiología Ignacio Chávez, en Ciudad de México, México.

El principal y más importante agradecimiento de este trabajo es hacía la Institución del CONACHyT, por brindar el apoyo en distintas maneras para la realización del mismo, y sin el cual no habría sido posible la realización y culminación de esta investigación.

Agradezco a la Universidad Autónoma del Estado de Hidalgo por abrirme las puertas a su Institución para cursar la Maestría en Ciencias Biomédicas y de la Salud, y brindarme las herramientas necesarias para poder concluir satisfactoriamente este trabajo.

Agradezco en especial al Director del Centro de Salud de Jalalpa, en la Alcaldía Álvaro Obregón, en la Ciudad de México, el Doctor José Humberto Badillo, por su muy valiosa información aportada para la realización de este reporte de investigación.

Durante mi estancia en la maestría, se realizó una publicación titulada “Models for Individualized COVID-19 diagnostic prediction” en la revista Mexican Journal or Medical Research ICSa, Biannual Publicación, Vol. 10, No. 20 (2022) 51-58.

Agradecimiento

Quiero dar las gracias a mis perritos y a mi michi, ellos son siempre mi motor.

Agradezco a la Doctora Claudia y al Doctor Abel por su guía y trabajo, fue muy agradable y enriquecedor trabajar con ustedes, este proyecto no habría sido posible sin su ayuda.

Gracias a mi Comité de evaluación por su valioso tiempo y trabajo.

A mis profesores que me han ayudado a llegar hasta aquí, me inspiraron y me brindaron sus conocimientos, respeto y amistad, espero algún día poder ser como ustedes.

ÍNDICE GENERAL

ÍNDICE GENERAL

ÍNDICE DE FIGURAS	8
ÍNDICE DE TABLAS	9
ABREVIATURAS	10
RESUMEN	11
ABSTRACT.....	12
I INTRODUCCIÓN	13
II ANTECEDENTES.....	15
<i>2.1 Generalidades.....</i>	<i>15</i>
2.1.1 COVID-19.....	15
2.1.2 Etiología y fisiopatología del COVID-19	15
2.1.3 Transmisión del COVID-19	15
2.1.4 Tipos de variantes del SARS-COV-2 y sus características.....	16
2.1.5 Sintomatología	17
2.1.6 Cepa original	18
2.1.7 Variante Alfa.....	19
2.1.8 Variante Beta.....	20
2.1.9 Variante Gamma	20
2.1.10 Variante Delta	21
2.1.11 Variante Ómicron.....	22
2.1.12 Pacientes susceptibles al COVID-19	23
2.1.13 Variación de la sintomatología dependiendo de las características de los pacientes	23
2.1.14 Sector más vulnerable de la población mexicana frente al SARS-COV-2.....	25
2.1.15 Contexto del COVID-19 en México	26
2.1.16 Sistemas de salud mexicanos y COVID-19	26
2.1.17 Manejo de la pandemia en CDMX.....	27

2.1.18	Centros de salud COVID-19 en Alcaldía Álvaro Obregón.....	27
2.1.19	Métodos estadísticos en modelos de predicción de prueba individualizada por COVID-19	28
2.1.20	Regresión logística binaria.....	28
2.1.21	Otros métodos para desarrollar modelos para prueba positiva por COVID-19.....	30
2.2	<i>Antecedentes del problema</i>	31
	Sistemas de predicción de prueba positiva por COVID-19.....	31
2.2.1	Modelo predictivo basado en los síntomas del olfato y el gusto para el diagnóstico de COVID-19.....	31
2.2.2	Predicción individualizada de riesgo para prueba positiva de la enfermedad de Coronavirus 2019.....	32
2.2.3	Más allá de predecir el número de infecciones: Predicción de quién es probable que sea COVID negativo o positivo	33
2.2.4	Modelos de predicción para el diagnóstico y pronóstico de COVID-19: revisión sistemática y evaluación crítica	33
2.2.5	Estimación temprana de los factores de riesgo para hospitalización y mortalidad por COVID-19 en México	35
2.2.6	Desarrollo de un modelo individualizado de predicción de riesgos para COVID-19 utilizando datos de registros electrónicos de salud.....	35
2.2.7	Detección de COVID-19: factores del paciente que predicen una prueba PCR positiva	36
2.2.8	Predicción temprana e identificación para casos graves de pacientes durante la pandemia de COVID-19: Un modelo de riesgo severo COVID-19 construido por análisis de regresión logística multivariable.....	37
2.2.9	Modelos de predicción COVID-19 y datos no explotados	38
2.3	<i>Planteamiento de problema</i>	38
2.4	Pregunta de investigación	40
III	JUSTIFICACIÓN.....	40
IV	HIPÓTESIS.....	41
V	OBJETIVOS.....	41
5.1	<i>Objetivo general</i>	41
5.2	<i>Objetivos específicos</i>	41

VI	MATERIALES Y MÉTODOS	42
6.1	<i>Contexto de la investigación.....</i>	42
	<i>Tipo y diseño de estudio.....</i>	42
6.2	<i>Selección de la población, criterios de inclusión y exclusión</i>	42
6.3	<i>Tamaño de muestra y muestreo</i>	43
6.4	<i>Diagrama de diseño experimental</i>	45
6.5	<i>Definición de variables (dependientes, independientes)</i>	46
6.5.1	<i>Prueba de regresión logística binaria</i>	48
	Prueba principio y procedimiento	
6.6	<i>Análisis Estadístico.....</i>	50
VII	RESULTADOS	52
7.1	<i>Análisis del total de la base de datos</i>	52
7.2	<i>Análisis de la muestra.....</i>	60
7.3	<i>Fase 2: Verificación de los modelos de predicción</i>	67
VIII.	DISCUSIÓN	70
8.1	<i>Desarrollo del diseño del modelo de predicción.....</i>	70
8.2	<i>Confiabilidad del modelo de predicción.....</i>	73
8.3	<i>Limitaciones</i>	76
IX.	CONCLUSIONES	77
X	PERSPECTIVAS	78
XI	REFERENCIAS.....	80
XII	ANEXOS.....	84
12.1	<i>Documento de aprobación del Comité de Ética e Investigación</i>	84
12.2	<i>Instrumentos</i>	85
12.3	<i>Publicaciones y estancia.....</i>	88

ÍNDICE DE FIGURAS

Figura 1: Casos confirmados y defunciones por COVID-19 en la Ciudad de México (2020-2022)	45
Figura 2: Diseño del estudio	46

ÍNDICE DE TABLAS

Tabla 1: Características de los datos por cada una de las 5 olas de la pandemia de COVID-19 y en general.....	45
Tabla 2: Descripción de variable dependiente y variables independientes	47
Tabla 3: Características sociodemográficas y morbilidades de la población de estudios ..	52
Tabla 4: Características sociodemográficas y morbilidades de la población de estudios entre positivos y negativos	54
Tabla 5: Prevalencia de los síntomas entre positivos y negativos	55
Tabla 6: Análisis de regresión logística binaria para predecir prueba individual de SARS-COV-2.....	57
Tabla 7: Resumen del desempeño de los modelos de predicción de prueba individual de SARS-Cov2.....	60
Tabla 8: Características sociodemográficas y morbilidades de la población de estudios de la muestra.....	61
Tabla 9: Características sociodemográficas y morbilidades de la población de estudios entre positivos y negativos de la muestra.....	62
Tabla 10: Prevalencia de los síntomas entre positivos y negativos de la muestra.....	63
Tabla 11: Análisis de regresión logística binaria para predecir prueba individual de SARS-COV-2 de la muestra.....	65
Tabla 12: Resumen del desempeño de los modelos de predicción de prueba individual de SARS-Cov2 de la muestra.....	67
Tabla 13: Valores dados por la fórmula de la regresión logística binaria en cada uno de los modelos.....	69
Tabla 14: Cuadro comparativo de modelos de predicción de COVID-19.....	75

ABREVIATURAS

C.I: Índice de confianza

CDC: Centros De Control para Enfermedades de EUA

CDMX: Ciudad de México.

CHARMS: Evaluación crítica y extracción de datos para revisiones sistemáticas de estudios de modelación de predicción

CONACHyT: Consejo Nacional de Ciencia, Humanidades y Tecnología

EHR: Registros de Salud Electrónicos

EUA: Estados Unidos de América

IA: Inteligencia Artificial

ICSa: Instituto de Ciencias de Salud

InDRE: Instituto de Diagnóstico y Referencia Epidemiológicos

OMS: Organización Mundial de la Salud

OR: Odds ratio

PCR: Siglas en inglés, Reacción en Cadena de Polimerasa.

PROBAST: Herramienta evaluadora de riesgo de sesgo de modelo de predicción

R²: Coeficiente de determinación

RL: Regresión Logística

ROC: Curva característica operativa del receptor

SARS-Cov-2: Síndrome Respiratorio Agudo Severo

SINAVE: Sistema Nacional de Vigilancia Epidemiológica

SISVER: Sistema para el registro de información sobre casos probables de Enfermedades Respiratorias Virales

UI: Unidad de cuidados Intensivos

Resumen

Antecedentes: Después de COVID-19, se desarrollaron métodos alternativos a las pruebas de laboratorio para la predicción del SARS-CoV-2. El objetivo de este estudio fue desarrollar y validar un modelo para la predicción individualizada de la infección por SARS-CoV-2. Nuestra pregunta de investigación fue: ¿Cuál es el desempeño de un modelo desarrollado para predecir el diagnóstico de COVID-19 en pacientes de los Centros de Salud de Álvaro Obregón, con un método remoto de recolección de datos, tales como la sintomatología, la demografía, contacto con caso sospechoso, y que sea desarrollado a partir de un modelo estadístico de regresión logística binaria?

Metodología: El estudio incluyó pacientes de los Centros de Salud de la Alcaldía Álvaro Obregón de 2021 a 2023 abarcando las 6 olas de contagio registradas en el Sistema de Vigilancia Epidemiológica de las Enfermedades Respiratorias Virales (SINAVE). Con muestras de 36,949 pacientes para la primera fase (desarrollo del modelo, primeras 5 olas) y 2,957 para la segunda fase (validación, ola 6). Las variables significativas entre positivos y negativos del virus se usaron para generar modelos de regresión logística binaria. Para la validación, los mejores modelos se aplicaron en datos de la ola 6.

Resultados: Los porcentajes promedio fueron de 73% de eficacia (63%-83%), sensibilidad de 67% (18-86%) y especificidad de 74% (59-91%). El modelo de la ola 5 tuvo el mejor desempeño (eficacia de 83%). En la validación con datos de la ola 6, la ecuación del modelo de la ola 5 tuvo 100% de sensibilidad, 0% de especificidad y 39% de eficacia y la ecuación del modelo general tuvo 66% de sensibilidad, 43% de especificidad, y 52% de eficacia.

Conclusión: Los modelos de predicción individualizada de prueba positiva de COVID-19 basados en datos del SISVER-SINAVE obtuvieron buen desempeño de predicción en la fase de desarrollo y un desempeño menor en la fase de validación. Estos modelos de predicción podrían aplicarse en estrategias de identificación temprana de casos, especialmente en población vulnerable.

Palabras clave: COVID-19; modelado estadístico; regresión logística; SISVER; factores de riesgo; síntomas.

Abstract

Background: After COVID-19, alternative methods to laboratory testing were developed for the prediction of SARS-CoV-2. The aim of this study was to develop and validate a model for the individualized prediction of SARS-CoV-2 infection. Our research question was: What is the performance of a model developed to predict the diagnosis of COVID-19 in patients of the Health Centers of Álvaro Obregón, with a remote method of data collection, such as symptomatology, demography, contact with suspicious case, and that it be developed from a statistical model of binary logistic regression?

Methodology: The study included patients from the Health Centers of the Mayor's Office Álvaro Obregón from 2021 to 2023 covering the 6 waves of contagion registered in the Epidemiological Surveillance System of Viral Respiratory Diseases (SINAVE). With samples of 36,949 patients for the first phase (model development, first 5 waves) and 2,957 for the second phase (validation, wave 6). Significant variables between positive and negative of the virus were used to generate binary logistic regression models. For validation, the best models were applied in wave 6 data.

Results: The models obtained general percentages of 73% of effective (63-83%), sensitivity of 67% (18-86%) and specificity of 74% (59-91%). The wave 5 model had the best performance (efficiency of 83%). In wave 6 data validation, the wave 5 model equation had 100% sensitivity, 0% specificity and 39% efficacy, while the general model equation had 66% sensitivity, 43% specificity, and 52% efficacy.

Conclusion: Individualized prediction models of positive COVID-19 test based on SISVER-SINAVE data obtained good prediction performance in the development phase and lower performance in the validation phase. These prediction models could be applied in early case identification strategies, especially in vulnerable populations.

Keywords: *COVID-19; statistical modeling; logistic regression; SISVER; risk factors; symptoms.*

I Introducción

La pandemia del COVID-19 surgió en el año de 2020 en Wuhan, China. El mundo enfrentó la peor crisis sanitaria en los últimos 100 años, y los sistemas de salud se vieron rebasados por los números tan grandes de contagios en lapsos cortos de tiempo, por lo que los insumos y el personal de salud se volvieron escasos, sobre todo en los países más pobres, incluyendo a México (1). Esta afectación fue mayor el punto más alto de contagios antes de que se desarrollaran las vacunas y comenzaran a aplicarse los esquemas de vacunación, comenzando por la población de mayor riesgo, que eran las personas mayores de 60 años (2). Debido a la falta de pruebas para COVID-19 (PCR y antígenos), y a su elevado precio (2) no hubo una detección adecuada de casos antes de que estos continuaran contagiando a otras personas, por lo que los contagios aumentaron de manera exponencial (2). En consecuencia, se volvió un objetivo urgente el desarrollar métodos de detección del virus, alternativos a los ya existentes, pero a más bajo costo y accesible para una gran parte de la población, a fin de disminuir contagios y evitar más decesos. Después de 6 olas de contagios, la pandemia se declaró terminada, y con ella la urgencia del desarrollo de métodos de detección, dejando al mismo tiempo la oportunidad de entender el fenómeno ocurrido y buscar estrategias para estar mejor preparados para futuros eventos epidemiológicos similares. En esta investigación se desarrolló un modelo de predicción de prueba individualizada positiva por COVID-19 a partir de regresión logística binaria, con datos obtenidos de pacientes que acudieron a los Centros de Salud de la Alcaldía Álvaro Obregón y fueron registrados en el Sistema Nacional de Vigilancia Epidemiológica de México (SINAVE).

El presente reporte se divide en 5 secciones. En la primera sección se aborda una revisión actualizada sobre los modelos de predicción de COVID-19 que se han publicado, en su mayoría desarrollados en Estados Unidos, los cuales utilizan la regresión logística binaria por las variables utilizadas (dicotómicas) o la inteligencia artificial aplicada a las bases de datos disponibles en cada región o país. El presente reporte es el único modelo de predicción de prueba individualizada COVID-19 desarrollado en México. En la segunda sección se describe la justificación, hipótesis y objetivos planteados en este reporte, el objetivo general que es desarrollar un modelo confiable de predicción de prueba positiva COVID-19 con datos de pacientes de los Centros de Salud de la Alcaldía Álvaro Obregón

en la Ciudad de México, esto determinando las variables, desarrollando el modelo y realizando una prueba de verificación. En la tercera sección se detalla el tipo de estudio, la metodología utilizada en la presente investigación, la decisión para el tamaño de muestra, la definición de las variables para la regresión logística binaria, los cuales se resumen en: la fase 1 que es el desarrollo del modelo estadístico de predicción y la fase 2 que es la validación del modelo de predicción obtenido previamente. En la fase 1 se aborda el desarrollo de 6 modelos realizados con el total de la base de datos, y otros 6 modelos con los datos de la muestra seleccionada, uno por cada una de las 5 olas de COVID-19 presentadas hasta ahora, y uno más el cual engloba los datos de todos los pacientes (ola 1-5). Las variables utilizadas incluyen datos del número de la ola, sociodemográficos, comorbilidades, sintomatología, tabaquismo, contacto con casos positivos y resultado de prueba PCR. Se recopilaron los datos de las tres semanas con más casos de cada una de las olas (picos de las olas reportados por el SINAVE), con excepción de la ola 1, en la cual se amplió la selección a 5 semanas debido a la falta de casos negativos. Los modelos se realizaron a partir de distintas pruebas estadísticas y aplicando regresión logística binaria, para posteriormente calcular un porcentaje general de predicción global con los resultados de los R^2 . Los datos obtenidos se muestran en 5 tablas por cada grupo de modelos (base de datos completa y muestra): (i) la tabla 1 muestra las características sociodemográficas y morbilidades del total de pacientes, (ii) la tabla 2 muestra los mismos datos, pero separando casos positivos y negativos, (iii) la tabla 3 muestra la prevalencia de los síntomas por ola en positivos y negativos, (iv) la tabla 4 muestra el análisis de regresión logística binaria para predecir prueba individual de SARS-CoV-2, y (v) en la tabla 5 el resumen de las regresiones logísticas binarias (en cada tabla se marcaron los valores que fueron significativos para p). En la segunda fase se realizó una verificación de los modelos de predicción previamente realizados, esto con nuevos datos obtenidos de la sexta y actual ola, y el tipo de análisis estadístico de acuerdo a los objetivos previamente planteados. En la cuarta sección se describen los resultados de los modelos, la discusión derivada de los datos obtenidos, las conclusiones finales y las recomendaciones acerca de los modelos desarrollados en este reporte y los resultados arrojados. En la quinta y última sección se enumeran las referencias utilizadas y en la sección de anexos se encuentran los formatos de

aprobación del Comité de Ética e Investigación de ICESA, el consentimiento informado, la hoja de estudio de casos con los datos clínicos, y el método de las pruebas PCR.

II ANTECEDENTES

2.1 Generalidades

2.1.1 COVID-19

El SARS-COV-2 es un tipo de coronavirus que provoca la enfermedad del COVID-19, el cual tiene diversos síntomas, y en la mayoría de los pacientes contagiados no es mortal, sobre todo si cuentan con los esquemas de vacunación, sin embargo, dichos síntomas varían de persona a persona y de acuerdo con sus características (1- 3). En México, a principios del 2020, alrededor del 24% de los pacientes necesitan ser hospitalizados, hoy en día esta tasa ha disminuido considerablemente en la población vacunada, y se estima que menos del 10% presenta complicaciones graves por el virus (3).

2.1.2 Etiología y fisiopatología del COVID-19

Los coronavirus son descritos como virus envueltos en ARN de sentido positivo no segmentado, que están presentes tanto en los humanos como en animales mamíferos, y dependiendo del tipo y la gravedad que desarrolle el individuo infectado, es si sus síntomas irán de leves a mortales (3). Con respecto al SARS-COV-2, hasta el día de hoy, hay 5 tipos de variantes de interés de acuerdo a la Organización Mundial de la Salud (OMS), derivados de la cepa original: Alfa, Beta, Gamma, Delta y Ómicron (cada una presenta subtipos), de los que se sabe que Alfa y Beta son los que más afectan a los seres humanos (2, 4).

2.1.3 Transmisión del COVID-19

La transmisión del COVID-19 va de persona a persona, el virus de tipo zoonótico se propaga por pequeñas partículas líquidas, de las personas infectadas, que se expulsan por las vías respiratorias y causa enfermedades del tipo respiratorias. Son distintas, las gotas respiratorias grandes y los aerosoles. El virus, como muchos otros, tiene la habilidad de mutar y adaptarse a distintos organismos y así ha logrado infectar a los seres humanos, se sabe ahora que está en constante evolución y un punto importante para su propagación es

por medio de los pacientes asintomáticos o que no muestran ningún síntoma, lo que nos ha demostrado desde que surgió, que es una alta amenaza para la salud a nivel mundial (5, 6).

Las partículas pueden llegar a las personas que están a una corta distancia, y pueden entrar al cuerpo humano por la nariz, boca o los ojos. Es más fácil que el virus se propague en espacios cerrados o con grandes aglomeraciones, ya que estas partículas se quedan suspendidas en el aire y son capaces de viajar hasta un metro de distancia (2, 5).

Otro de los riesgos de transmisión es tocarse la cara con las manos sucias, es decir, que se haya tocado alguna superficie contaminada y no lavarse las manos. Se ha demostrado que uno de los factores con mayor riesgo de contagio es el contacto con otra persona que está infectada, por lo que se deben de tomar las precauciones correctas en estos casos (1).

Hoy en día se tiene más información acerca del virus y su sintomatología, además de los factores de riesgo, aunque siguen siendo estudiados, dependiendo de la persona y de la variante (5, 6).

2.1.4 Tipos de variantes del SARS-COV-2 y sus características

Los virus, como el SARS-COV-19, cambian constantemente a través de mutaciones, una variante de un virus puede tener una o hasta varias mutaciones que le da características especiales en comparación con las otras, por lo que se debe tener un constante monitoreo para identificar a cada una de ellas, sus características, su sintomatología y cómo reaccionan a las vacunas existentes, para saber qué tan peligrosas son sus variaciones para el ser humano (7, 8). Hasta el momento se han identificado 5 variantes de interés del COVID-19, las cuales tiene la característica de tener más agresividad y ser mucho más contagiosas que la cepa original, los síntomas en general son parecidos a los reportados al inicio de la pandemia, sin embargo, la mortalidad y la probabilidad de agravarse si se contagia, son mayores, sobre todo en personas no vacunadas (7, 8). Desde el surgimiento del virus del COVID-19, los investigadores en todo el mundo se han dedicado a estudiarlo y desarrollar vacunas, sin embargo, se han enfrentado al reto de la constante mutación del virus, por lo que las vacunas tienen que ir modificando sus características a la par de la enfermedad (9).

2.1.5 Sintomatología

Se sabe que los síntomas del COVID varían de persona a persona, dependiendo de las características y antecedentes de los pacientes, y ahora también, que varían de acuerdo con la variante del virus con que se esté contagiado (2, 10, 11). Estos síntomas pueden ir de leves a moderados, y en los casos más graves, pueden incluso llevar a los pacientes a la muerte, de aquí la importancia de la vacunación, las cuales pueden prevenir que los síntomas en general se agraven, dependiendo del tipo de vacuna, pero en general con un aproximado de 80% de prevención de hospitalización. Aunque el espectro de la enfermedad es variado, los síntomas reportados más frecuentes son la fiebre, la tos, la disnea y el dolor de cabeza, pero también es posible que no se tengan estos síntomas en absoluto, ya que hay personas que, a pesar de estar contagiadas, son asintomáticas, o bien, sus síntomas se pueden confundir con síntomas comunes o de resfriado, como estornudos, secreción nasal, dolor de garganta, dolor de cabeza, y problemas en el tracto gastrointestinal. La disfunción de órganos inducida por el SARS-CoV-2, puede ser explicada, en teoría, por una combinación de varios síntomas, toxicidad viral directa, lesión isquémica por vasculitis, trombosis, o inflamación de trombo, entre otros (2, 10, 11).

Otros de los síntomas más frecuentes y a veces los primeros son la pérdida del sentido olfato y del gusto, en mayor o menor medida, además de que también se han reportado dolor de cuerpo, problemas gastrointestinales, así como reportes de disnea en los casos en que se ha presentado incluso el deceso del paciente, y que esta iba agravándose junto con el estado de salud del contagiado. En los casos más graves, se presenta baja saturación de oxígeno, con casos extremos de hasta 30%, problemas para respirar, fiebre de más de 40 grados, sonidos extraños al respirar, como silbidos, diarreas, náuseas y vómitos, además de fatiga crónica o extrema (10, 12).

Un estudio reportó una mediana en la edad de 52 años, además de que los síntomas más frecuentes fueron la fiebre con 67%, la tos con 37%, el malestar corporal y la neumonía con 11%, y el dolor de garganta con 8%. También reportó menos de 1% de pacientes con enfermedades menos comunes como problemas del miocardio, insuficiencia renal y cardíaca, shock séptico. Uno de los síntomas más graves y frecuentes fue el síndrome de dificultad respiratoria. La mayoría de pacientes no presentan sintomatología grave, con menos del 1% del total de los estudiados, mientras que los pacientes que

presentaron defunción, tenía algún tipo de problema en el corazón, neumonía, fiebre alta, o tos (13).

Con respecto a las variantes del virus, diversos estudios han concluido que las cepas más agresivas causan síntomas más graves y agresivos, que aparecen en un menor tiempo de incubación, y pasan de leves a graves en menos tiempo (7- 14). Los síntomas son los mencionados previamente, pero aumentan a otros más complicados, como derrames cerebrales y riesgos cardiacos, entre otros, ya que aún se están descubriendo nuevas consecuencias a raíz del virus, así como el llamado '*Long COVID*', que son los síntomas persistentes después de la recuperación del paciente, en especial a pacientes que se enfermaron antes de que se comenzara con los esquemas de vacunación (14). Más recientemente se ha vuelto predominante la variante Ómicron, de la cual los síntomas son más parecidos al resfrío, y con menor número de personas hospitalizadas, derivado del gran número de vacunas que se han aplicado, además de la llamada inmunidad de rebaño, lograda por el número de contagios que se han presentado. Sin embargo, es de suma importancia que se sigan aplicando las medidas sanitarias adecuadas para seguir mitigando la pandemia y evitar más contagios que podrían derivar en nuevas variantes más agresivas o resistentes a las vacunas (15).

2.1.6 Cepa original

El virus del SARS-COV-2 fue detectado por primera vez en la provincia de Wuhan, China en 2019. Desde que se empezaron a descubrir más y más casos, se ha reportado que los síntomas más comunes son la fiebre, la tos, el dolor de cabeza, la disnea, el dolor de garganta y la pérdida del olfato y del gusto, y en los casos más graves, la dificultad para respirar provocada por la baja saturación de oxígeno, que, sin embargo, podrían no estar presentes, por lo que la afirmación de que un caso es positivo al COVID-19, puede ser compleja si no se realizan las pruebas necesarias (8, 16). Además de los síntomas ya mencionados, se han notificado también síntomas gastrointestinales, y pérdida completa del olfato o gusto en los casos más graves, o incluso afectaciones en otros órganos, así como que la disnea fue más frecuente en los casos graves y mortales. La mayoría de los pacientes infectados por el virus experimentaron síntomas respiratorios leves, y lograron recuperarse sin haber llevado un tratamiento específico. Sin embargo, una de las principales

problemáticas de la aparición de la enfermedad es que aún no se contaban con vacunas para combatirla, por lo que las personas más vulnerables sufrieron con la gravedad de los síntomas. El virus del SARS-CoV-2 ha presentado las llamadas variantes desde su aparición, el virus ha cambiado a lo largo del tiempo que ha infectado a las personas y a esto los expertos le llaman “variantes”, y al virus original se le llama “tipo salvaje”, y cada una de las variantes tiene sus propias características (8, 16).

2.1.7 Variante Alfa

Esta variante fue detectada originalmente en Reino Unido a principios del 2021, se le considera una variante de propagación porque es la más transmisible, se contagia en promedio durante más de 10 días que la original y se expandió rápidamente. Europa es la que presenta más casos de esta variable, y tuvo un riesgo de 3.8 veces mayor de muerte, posibilidades de contagio y de tener que ser internados en cuidados intensivos (2, 8). De acuerdo con un estudio, las mujeres tenían un mayor número de contagios que los hombres, y que la prevalencia de condiciones de comorbilidad no fueron muy diferentes con la de la cepa original (17).

Los pacientes infectados con esta cepa, a pesar de tener sintomatología similar a la original (molestias y dolores, dolor de garganta, diarrea, conjuntivitis, dolor de cabeza, pérdida del olfato y del gusto, erupciones cutáneas), tuvieron una mayor probabilidad de que se complique su condición, lo que podría sugerir un aumento de la virulencia relacionada con la variante Alfa. Se sabe que como con la cepa original, los pacientes infectados con esta variante y que llegaron a ser hospitalizados, en un gran porcentaje tenían enfermedades previas. Diversos estudios sugieren que esta variante tiene características especiales con respecto a las proteínas fuera de la proteína de pico, lo que se puede traducir como que solo algunas horas de haber infectado a la persona, el virus puede suprimir la defensa que crea la respuesta rápida que el cuerpo humano siempre aplica contra patógenos invasores, y así logra lidiar con las “inmunidad innata” o incluso llega a deshacerse de ella, lo que es importante para la transmisión. (8, 18, 19).

2.1.8 Variante Beta

La variante B.1.351, variante Beta o GH501Y.V2, tiene múltiples mutaciones de picos, dio lugar a la segunda ola de COVID-19, apareció en los Estados Unidos a principios de 2021. Tiene mayor riesgo de transmisión y menor neutralización por terapia con anticuerpos monoclonales, sueros convalecientes y vacunación, lo que quiere decir que tiene una mayor capacidad para evadir ciertos tipos de vacunas. Según algunos estudios, se cree que puede llegar hasta reducir hasta 12 veces la efectividad de la vacuna que fue creada por Moderna, gracias a diversos anticuerpos monoclonales que no fueron capaces de neutralizar a este tipo de variante del virus (8, 11, 18- 20). También se obtuvieron datos de que tenían menos capacidad de neutralización de los sueros de las personas infectadas con esta variante que estaban convalecientes, esto hasta en 9 veces, y las personas que ya habían recibido la inmunización con la vacuna, mostraron tener de 10 a 12 veces menos respuesta inmune, aunque por otro lado, a las personas que ya habían recibido una segunda dosis de esta misma vacuna, vieron su respuesta inmune mejorada (8, 11, 18- 20). Esta variante fue detectada originalmente en Sudáfrica, los síntomas más habituales son los mismos que con la variante Alfa: fiebre, tos seca y cansancio y apareció con la tercera ola de contagios del virus del COVID-19 (8, 11, 18- 20).

2.1.9 Variante Gamma

Esta variante fue detectada originalmente en Brasil y en Japón, también se le llama linaje P.1, tienen 17 cambios en sus aminoácidos, de los que 3 son denominados de interés o preocupación (19, 20). Se realizaron una serie de estudios epidemiológicos genómicos de fundación Oswaldo Cruz de Manaus en los cuales se estudiaron 250 tipos de genomas del virus del SARS-CoV-2, de distintas regiones de Brasil, esto entre las fechas de marzo de 2020 y enero del año 2021. Gracias a este estudio se conoció que esta primera fase de contagios y por lo tanto, aumentos exponenciales de esta variable fueron correspondientes a la variable B.1.195, que luego fue desplazada por la B.1.1.28 (19, 20). Los síntomas son parecidos a los de las variantes Alfa y Beta, los cuales incluyen síntomas correspondientes a una enfermedad respiratoria, como dificultad al respirar, tos, anosmia, además de fiebre, escalofríos, dolor de cabeza, cambios en el sentido del olfato y del gusto, e incluso en los casos más graves, problemas en otros órganos (2, 8).

Tiene un notable mayor riesgo de transmisión que la cepa original, pero menos que las otras variantes, una menor neutralización por terapia con anticuerpos monoclonales, sueros convalecientes y vacunación, esta variante evolucionó a finales de noviembre del 2020 y finalmente el 11 de enero del 2021 la OMS la catalogó como variante de preocupación (11).

2.1.10 Variante Delta

Denominada como variante de preocupación por la OMS, posee dos mutaciones, E484Q y L452R, que le permiten mayor unión al ACE2 y capacidad de evadir el sistema inmunológico (21). Se sugiere que esta variante puede ser hasta 60% más transmisible que la variante Alfa, por lo que rápidamente fue catalogada por la OMS como variante de preocupación, sin embargo, demostró que su letalidad es más baja que la de sus predecesoras, ya que para el momento de su aparición, muchas personas ya habían recibido al menos una dosis de vacunación en contra de esta enfermedad. Detectada originalmente en India en noviembre del 2020, es la más letal en personas no vacunadas, tiene síntomas parecidos al resfriado común o fuerte (estornudos, secreción nasal, dolor de garganta, y en casos más graves, problemas en el tracto gastrointestinal, hepatobiliar, cardiovascular, renal y sistema nervioso central. Gracias a su alto nivel de propagación, desató una nueva ola de contagios a poco tiempo de que apareció, por lo que la OMS tuvo especial atención en sus mutaciones, ya que para cuando alcanzó su índice más alto de contagios, ya ocupaba tres cuartas partes del total de los mismos (21). Según el Centro de Control para Enfermedades (CDC) de Estados Unidos, las personas que si bien ya habían sido vacunadas, aún podían contagiarse y seguir contagiando a más personas, sin embargo, dichas personas presentaban un bajo índice de casos con sintomatología grave y pese a que se han reducido las hospitalizaciones y defunciones. También se ha reportado que si una persona que se haya contagiado con esta variante y que ya esté vacunada no se aísla, sigue siendo contagiosa, pero en promedio menos tiempo que las personas que no han recibido ninguna dosis de vacunación. Es la variante predominante en el mundo hacia agosto de 2021 y finales de este mismo año (2, 8, 21). Según un estudio realizado en 2021, la severidad de esta variable fue de 11.1% y cada aumento de años o de bilirrubina total presentaron ser factores de riesgo para los casos de pacientes contagiados que se tornaban graves. Así mismo, conforme

aumento la edad de las personas contagiadas, el riesgo de volverse casos graves aumento en 13.4 veces y 3.9 cuando la edad que presentaron era mayor de 58.5 años. Hasta principios del 2023, esta variante ha sido la más grave de entre las que han surgido (23).

2.1.11 Variante Ómicron

La variante de interés más reciente, llamada B.1.1.529, detectada por primera vez el 24 de noviembre del 2021 (2). Apenas a un día de haber sido descubierta, fue designada como variante de preocupación por la OMS, ya que presentaba una gran cantidad de mutaciones, varias de ellas de preocupación, en comparación con el primer virus que había surgido en China en 2020, de la cual se ha reportado que se tiene el más alto porcentaje de contagio y reinfección, por lo que los números de pacientes con el virus se ha incrementado con los niveles más altos registrados desde el inicio de la pandemia (2). Este tipo de variante ha mutado y se ha propagado internacionalmente, sin embargo, los síntomas son menos graves en comparación con las otras variantes, y en la mayoría de los casos no ameritan hospitalización, esto sumado a que hoy en día en México, de acuerdo a cifras del Gobierno, más del 90% de la población ha sido vacunada (1), aunque se ha reportado según la OMS, las personas que han sido vacunadas o las que ya se habían contagiado previamente del virus, pueden llegar a volver a infectarse con la variante Ómicron. Al haber surgido durante las fiestas de fin de año, tuvo una circulación mucho más rápida a otras variables, por lo que rápidamente se convirtió en la variante predominante desde las primeras semanas del 2022 y actualmente a inicios del 2023, abarcando un total de 3 olas de contagio de las 6 que se han presentado desde el inicio de la pandemia (21, 24). Esta variante puede presentar síntomas parecidos a los de las otras anteriores. Sin embargo, en general presenta síntomas más parecidos a los de un resfriado, igualmente debido a que actualmente, la mayoría de la población de los países con acceso a vacunas, ya han recibido al menos una vacuna, y en los mejores casos, el esquema completo, y se ha demostrado que las vacunas actuales pueden ser eficaces en contra de este tipo de contagios (8). Las secuencias genéticas que se han reportado para esta variante mostraron 50 mutaciones con respecto al virus original, muchas de ellas de preocupación, y correspondientes a un posible aumento de transmisibilidad. Sin embargo, esta variante continúa desarrollando subtipos, los cuales hasta inicios del 2023 siguen causando miles de contagios cada día (21, 25).

2.1.12 Pacientes susceptibles al COVID-19

El nivel de la afectación va dependiendo de características y de antecedentes de los pacientes, en general, las personas con enfermedades previas (co-morbilidades), suelen enfermar de una manera más agresiva, así como las personas con obesidad, malos hábitos de alimentación, con sedentarismo, que realizan poca actividad física, o que presentan problemas pulmonares de algún tipo, aunque también se han conocido casos de personas saludables que han muerto a causa del virus (13, 14, 20). Se debe recalcar que, en promedio, a las personas que más afecta el virus y los síntomas, son las personas mayores de 60 años, y han llegado a afectar hasta al 66% de las de estos adultos en América hasta antes de las vacunas, esto debido a que sus sistemas inmunes están más debilitados y que padecen enfermedades propias de la edad. Según estudios recientes, a medida que envejecemos, los pulmones van perdiendo elasticidad y resistencia, y combinados con otros problemas de salud, hacen perder funciones pulmonares y respiratorias (26).

Derivado de la vacunación que se ha aplicado desde 2021, se han detectado que las personas más afectadas en los últimos meses por el COVID-19, son aquellas que no han sido vacunadas ni siquiera una vez, se reporta que son hasta un 97% de los casos totales hospitalizados en México. Mientras que las personas que han sido vacunadas ocupan menos de 3% del total de las hospitalizaciones, por lo que podemos deducir que los sectores más susceptibles de la población frente al COVID-19, irán cambiando a medida que se sigan aplicando las dosis de vacunación. Sin embargo, teniendo en cuenta que los factores de riesgo siguen afectando aún a las personas vacunadas, ya que la vacunación no previene el contagio, sino la severidad de la enfermedad (1, 27).

2.1.13 Variación de la sintomatología dependiendo de las características de los pacientes

Los síntomas determinados hasta el momento siguen siendo estudiados y actualizados, puesto que durante su tiempo de prevalencia, sigue presentando variantes con las cuales también varía la sintomatología y el grado en que esta se presenta en los pacientes contagiados, aunque si presentan en general los síntomas más característicos mencionados anteriormente. Cabe mencionar que la variante Delta, es la más letal hasta el momento, sin embargo, actualmente la que presenta un mayor números de contagios es la variante Ómicron, según datos de los CDC (8).

Los problemas que presenta un país frente al COVID-19 tiene que ver mucho con las características de su población, ya que hay países con más proporción de personas sanas que otros, o que su población en general es más longeva que en otros países, o su cultura en general para alimentarse, hacer ejercicio o cuidar de su salud (10). Países como México, que tiene el primero lugar mundial en obesidad infantil y el segundo lugar en obesidad en adultos, este es un factor negativo y que hará que el virus afecte más a la población en general (1, 10). La sintomatología general del virus a nivel global, independientemente de las características del país, son: fiebre alta, tos, dolor de cabeza, pérdida del sentido del olfato y del gusto, y en los casos más graves disminución de la saturación en la oxigenación, lo que conduce a dificultad para respirar, sin embargo, la nueva variante dominante Ómicron, tiene síntomas más parecidos a un resfriado común (5, 12, 16, 21). La sintomatología descrita por grupos es la siguiente:

Adultos mayores de 60 años: tienen mayor riesgo a desarrollar síntomas más graves, esto relacionado a las afecciones que ya padecen por su edad, los principales síntomas son: tos seca, fiebre alta, dolor de cabeza o garganta, y fatiga aun mayor que en los casos de contagios jóvenes.

Adultos de 40 a 60 años: síntomas generales ya mencionados.

Adultos de 30 a 40 años: síntomas generales ya mencionados.

Adultos jóvenes de 18 a 30 años: síntomas generales ya mencionados.

Infantes: síntomas generales ya mencionados.

Personas con alguna enfermedad previa: las personas que ya padecen alguna enfermedad corren un mayor riesgo de que se complique el cuadro que presenten a raíz del COVID-19, como la obesidad, enfermedades crónicas, y sobre todo si estas enfermedades tienen que ver con el sistema respiratorio (tales como asma, neumonía, bronquitis). En general los pacientes sintomáticos leves pueden llegar a presentar los síntomas generales (7).

Es importante recalcar que existe el llamado síndrome del COVID persistente, con el cual, los síntomas que se presentaron durante la enfermedad pueden seguir por un periodo aún no determinado de tiempo. Al ser un síndrome nuevo aún se están realizando investigaciones, pero se estima que más del 80% de los pacientes recuperados del COVID-19, presentan prevalencia de los síntomas tiempo después de haber dado negativo en las pruebas, sobre todo en pacientes que se enfermaron en los primeros meses de la pandemia

cuando aún no se contaba con una vacuna, ya que sus cuerpos no contaban con ningún tipo de protección. La mayoría de los pacientes que presentaron gravedad durante este primer periodo eran personas mayores de 50 años, con enfermedades previas, co-morbilidades o problemas pulmonares de algún tipo (28). Se realizó una revisión sistemática en la cual los síntomas persistentes más frecuentes fueron la fatiga, la falta de aire, artralgia, dificultades para dormir, dolor en pecho y síntomas psicológicos, además de otras afecciones de largo plazo, como los son principalmente afectaciones cutáneas, cardiovasculares, respiratorias, salud mental, músculo-esquelético, renal, cardíaca, neurológica, aparecidos en casos de pacientes que presentaron gravedad durante la enfermedad (29).

2.1.14 Sector más vulnerable de la población mexicana frente al SARS-COV-2

Desde el inicio de la pandemia se determinó que el grupo de personas más vulnerable es el de los adultos mayores de 60 años y las personas con enfermedades co-mórbidas, ya que al inicio de la pandemia y sin comenzar con los esquemas de vacunación, se han contabilizado más de 124 000 muertes de este sector, presentaba un porcentaje de 63% del total de las defunciones, mientras que los adultos jóvenes presentaban una menor cantidad con cerca de 70 000 en México a agosto de 2021, y que en las personas con antecedentes de enfermedades de otro tipo presentan también un alto nivel de probabilidad de que complicaciones, y posterior hospitalización, y en casos extremos, la defunción (1).

En fechas recientes, muchos países han comenzado a aplicar una gran cantidad de vacunas, con planes de vacunación que en primera instancia se enfocó en el personal de salud por ser el sector más expuesto al virus, y presentar un gran número de defunciones durante los primeros meses de la pandemia, y actualmente cualquier sector puede aplicarse los refuerzos siempre y cuando haya vacunas disponibles (1, 25). En México, el plan de vacunación contempló, en el segundo sector a vacunar fue el de los adultos mayores de 60 años, ya que estos presentaban un mayor porcentaje de mortalidad, y por lo tanto, eran el sector más vulnerable, y así se siguió un orden en cuánto a la vacunación para cada dosis de vacunación, incluyendo los refuerzos. Derivado de los esquemas de vacunación, la mayoría de casos de mortalidad se presentan principalmente en personas que no han sido vacunadas en contra del virus, y en contraste, las personas mayores que si han sido vacunadas presentan menor número de casos graves y defunciones. Se reporta que al menos el 95% de

la población mayor de edad en México, al menos con una dosis, ha sido inmunizada contra el COVID-19, y que del total de hospitalizaciones, el 97% de los casos, son de personas que no han recibido ninguna dosis de vacunación, mientras que los que sí, representan menos de un 3% (1, 24).

2.1.15 Contexto del COVID-19 en México

El primer caso de COVID-19 en México, llegó el 27 de febrero del 2020. A partir de ello, el número de casos comenzó a aumentar exponencialmente, ya que las personas continuaron teniendo contacto estrecho unas con otras y no se extremaron precauciones para, en primera, no aumentar los casos, y en segunda, para seguir los casos positivos a los posibles contagios (1- 24). En México, tanto el sector público como el privado, hicieron frente a la pandemia a pesar del gran número de contagios, y derivado de esto, el Gobierno de México ha implementado medidas sanitarias conforme al desarrollo de la pandemia, y en función a los semáforos epidemiológicos (1, 24). Desafortunadamente, los mexicanos han demostrado no tener una gran cultura de cuidado en cuanto a la salud, esto sumado a que la baja escolaridad y problemas para seguir las recomendaciones de las autoridades sanitarias, han incrementado tanto los contagios hasta llegar al punto de saturar a los sistemas de salud de todo el país. Actualmente, los números de casos mortales disminuyeron en gran medida por la vacunación y la llamada “inmunidad de rebaño” que se desarrolló a partir del gran número de contagios, ya que a pesar de que aún se reportan miles de casos diarios, los casos graves y fatales no son tantos como antes (2). A inicios de febrero del 2023, cifras oficiales reportan 7.38 millones de casos de contagios y 332 mil defunciones, con más de 9 mil casos por día y un promedio de 4 mil a la semana en todo el territorio mexicano (1, 24).

2.1.16 Sistemas de salud mexicanos y COVID-19

Desde la llegada del COVID-19 a México en 2020, los sistemas de salud han enfrentado su mayor reto en los últimos 100 años, sobre todo por la falta de recursos y el gran número de contagios que se presentan por día, llegando a acumular en su pico más alto hasta 28 000 casos por día en la ciudad de México (24). Los hospitales y Centros de Salud donde se atienden a los pacientes COVID en CDMX alcanzaron su saturación solo 3 meses después del registro del primer contagio en México, estos centros están equipados para tratar los

síntomas más severos que provoca el virus, como son los niveles de oxigenación extremadamente bajos y otras complicaciones (1). De los registros que presentan dichos centros de atención, se estima el número de contagios y de mortalidad, además del semáforo en el que se encontraran las entidades mexicanas, dependiendo del número de camas ocupadas o disponibles. Sin embargo, en muchas ocasiones se han visto rebasados por los abrumantes números de contagios presentados en las diferentes olas. A inicios del 2023 los hospitales y centros COVID-19 ya no presentan tanta saturación debido a la disminución de casos graves (1, 2).

2.1.17 Manejo de la pandemia en CDMX

A partir del surgimiento de la pandemia del SARS-COV-2, el Gobierno de la Ciudad de México comenzó a tomar medidas sanitarias que han sido publicadas y seguidas en la página web oficial del mismo. Cabe destacar que la Ciudad de México es el estado más afectado por el COVID-19, con el mayor número de casos positivos de contagios, con un 25% de los casos totales nacionales, presentando saturación de camas y ventiladores en prácticamente todos los hospitales. A pesar de los esfuerzos del sector salud que llegó a aplicar miles de pruebas diarias, mismas que se reportaban en la página oficial del Gobierno (24, 30). La página web oficial presenta informes oficiales diarios de las cifras y datos obtenidos de los servicios ofrecidos a la población, de información de atención en casa, de apoyos a pacientes COVID-19, videos, boletines, comunicados, reglas básicas para el cuidado y prevención, información, datos y servicios COVID y otros recursos informativos (1).

2.1.18 Centros de salud COVID-19 en Alcaldía Álvaro Obregón

Dentro de la Ciudad de México, la zona que se ha visto más afectada, y la que ha presentado el mayor número de casos de contagios, y de acuerdo a datos del CONACHyT, en julio del 2022, fue el epicentro de contagios del país, con casi 9 mil al día, y un total aproximado a septiembre de este año de 140 mil casos totales confirmados (24). El gobierno de la Ciudad de México brinda información de los Centros de Salud COVID disponibles para el tratamiento a personas infectadas, el porcentaje de ocupación, y los lugares en donde se pueden realizar las pruebas de diagnóstico PCR. La alcaldía de Álvaro

Obregón cuenta con 17 Centros de Salud y 1 hospital público para la atención de la pandemia. La Alcaldía cuenta con una página web oficial, para poner a disposición de las personas de la misma, información acerca del COVID-19 y de las acciones y programas que se están realizando para enfrentarlo, en Álvaro Obregón y en toda CDMX. Así como información acerca de la pandemia, los cuidados, solicitudes de información, medidas de prevención, formas de contagio, mitos y realidades, y preguntas frecuentes. También contiene un apartado de estadísticas con el número de casos registrados actualmente en la Ciudad de México y en la Alcaldía Álvaro Obregón, con fuente en la página oficial de CDMX. Al 2 de febrero del 2023 reporta 235 mil casos confirmados, 377 mil negativos, 12 mil sospechosos, casi 4 mil defunciones, 226 mil recuperados y 674 activos (31).

2.1.19 Métodos estadísticos en modelos de predicción de prueba individualizada por COVID-19

Un modelo predictivo es aquel que, mediante el análisis de diversos datos y métodos que serán las entradas del mismo, y a partir de ellos, produce una salida, lo que significa que predice el comportamiento de un individuo o fenómeno, En este reporte nos enfocamos en los modelos de predicción individualizada de COVID-19 que son alternativos a las pruebas de laboratorio que se aplican y son las más confiables (PCR y antígenos) (32).

2.1.20 Regresión logística binaria

La regresión logística binaria es parte de un conjunto de métodos estadísticos, que tratan de explicar las relaciones causales o las hipótesis cuando se tiene que las variables dependientes son nominales. La regresión logística binaria explica cómo se modifica la variante dependiente (la cual es la parte más importante de dichos modelos, ya que es la característica o dato a predecir, y es una variable binaria, esto significa que tiene únicamente dos categorías, también se les llama dicotómicas) cuando se cambian las variables independientes. Otra parte importante son las variables continuas, que son las más utilizadas. “Lo que caracteriza en principio a las distintas clases de modelos de regresión es la naturaleza de la variable dependiente, con variables continuas”. La regresión logística se ha utilizado ampliamente para análisis epidemiológicos y clínicos por ser un instrumento estadístico muy versátil. De manera resumida, podríamos decir que con las variables

independientes se va a tratar de hacer una predicción de probabilidades de que algo ocurra o de que este algo no ocurra. Estos tipos de modelos son muy útiles en el campo médico, ya que nos permiten hacer análisis de los datos obtenidos de los resultados, ya sean derivados de algún estudio o de los pacientes como tal, y este método nos puede dar, al mismo tiempo, términos explicativos, pero también términos predictivos (33).

En el caso de las investigaciones que han sido publicados hasta el día de hoy, febrero del 2023, en el campo de la predicción de prueba positiva por COVID-19, la mayoría de ellos aplican la regresión logística binaria combinada con otras pruebas para poder desarrollar modelos confiables, esto debido a que los factores de riesgo y la sintomatología en general son variables dicotómicas, así como el resultado de la prueba PCR que es la variable independiente a predecir, en general estos modelos alcanzan más del 70% de porcentajes de predicción (34- 40).

Fórmula de la regresión logística binaria

La fórmula general para calcular el valor de P(Y) que es la probabilidad de que Y tomé el valor de 1 o la presencia de la característica, que en este caso sería la probabilidad de que un caso tomé el valor de prueba individual positiva por COVID-19, es de la siguiente forma (34- 40):

$$\text{logit } [p(x)] = \log \left[\frac{p(x)}{1 - p(x)} \right] = a + b_1x_1 + b_2x_2 + b_3x_3 \dots$$

Simplificando la fórmula, toma la siguiente forma:

$$p = \frac{\exp^{(a+b_1x_1+b_2x_2+b_3x_3\dots)}}{1 + \exp^{(a+b_1x_1+b_2x_2+b_3x_3\dots)}}$$

Donde:

P: probabilidad de la presencia de la característica

e: número de Euler=2.71828 aprox.

a: valor de la constante

bn: valor de B de la variable

xn: conjunto de la covariable

2.1.21 Otros métodos para desarrollar modelos para prueba positiva por COVID-19

A pesar de que la mayoría de modelos relacionados con la predicción de COVID-19 fueron desarrollados con la regresión logística binaria, varios otros han aplicado la inteligencia artificial (IA) para predecir, algunos otros como el “inmunoensayo” y la “identificación basada en biosensores”, algunos incluso aplicados por medio de dispositivos móviles como el Apple watch, Bioharness o anillos inteligentes, a partir de las lecturas de los biosensores. Algunos son utilizados como un sistema de detección de pacientes asintomáticos de COVID-19 (41). Otro modelo es el de los “biosensores plasmáticos” que se basan en resonancias plasmóticas superficiales localizadas a partir de nanomateriales aplicando la ingeniería nanométrica. Los “biosensores electromecánicos” utilizan electrodos junto con sensores que monitorean las señales biológicas del cuerpo humano para detectar la presencia de alguna anomalía. También existe la “detección de COVID-19 a través de radiografía” haciendo radiografías de pecho, usando estas imágenes junto con inteligencia artificial, también tomando en cuenta otro tipo de estudios e información proporcionada por el paciente. El “método de detección del virus usando la tos y los sonidos de la respiración” es otro tipo de detección del COVID-19, a partir de la examinación a los pacientes, como la tos, la respiración y la voz. Otro tipo de método es el “Rastreador de síntomas de COVID-19” el cual detecta y muestra patrones de comportamiento e infección en los pacientes comparándolos con información previa de pacientes positivos (41).

Se ha desarrollado un modelo de aprendizaje automático con síntomas intensivos, el cual trabaja identificando los síntomas más importantes o relevantes que puedan indicar la existencia del virus del COVID-19 en los pacientes. Esto utilizando varios conjuntos de datos clínicos, en especial la sintomatología más reportada de los pacientes positivos a la enfermedad, también de laboratorio, registros de datos de epidemiología, síntomas de los pacientes o incluso si se cuenta con estudios de laboratorio, son utilizados. Toma en cuenta la región en donde vive el paciente. El modelo presentó hasta un 97% de precisión (42).

Estos métodos trabajan a partir de bases de datos. Estas investigaciones intentan predecir los resultados de las pruebas de los pacientes sospechosos, la severidad de los síntomas de los pacientes confirmados, la mortalidad, la hospitalización y la progresión del virus. Sin embargo este tipo de investigaciones en general requieren una gran cantidad de recursos, tiempo, y son más complicados. Lo que se busca es lograr que sean métodos

alternativos y eficientes. Estos modelos han demostrado ser confiables y han reportado más del 80% de predicción de resultado por COVID-19 (37, 44, 45).

2.2 Antecedentes del problema

Sistemas de predicción de prueba positiva por COVID-19

2.2.1 Modelo predictivo basado en los síntomas del olfato y el gusto para el diagnóstico de COVID-19

El COVID-19 comparte varios síntomas con la gripe común que todos conocemos, y a pesar del gran número de estudios que se han realizado por la apremiante necesidad de conocer al virus y sus síntomas, se sigue estudiando cuáles son los síntomas del mismo, ya que, con el paso del tiempo y el surgimiento de las variantes, los síntomas han ido variando y han aparecido nuevos, sin embargo, se tiene un censo común de los más frecuentes (34). Este trabajo se realizó en California, E.U.A., y para este estudio, se les dio a los pacientes una encuesta electrónica a través para que pudieran contestarla, pero las encuestas fueron anónimas y se realizaron por medio de redes sociales, para consultar con los pacientes que tenían su prueba positiva de COVID-19, y saber cuál había sido la sintomatología más común que presentaron. La encuesta consistió de 10 de los síntomas más frecuentes, su información demográfica, comorbilidades y resultados de su prueba (34). El método de análisis de los datos se realizó por medio de regresión logística por etapas, que se utilizó para identificar los predictores para la positividad del virus de COVID-19 en los pacientes. Se realizaron evaluaciones de los clasificadores que fueron seleccionados para realizar las predicciones, utilizando la característica operativa de la curva del análisis del receptor (ROC). Los resultados del estudio, con total de 145 pacientes con COVID-19 positivos y 157 pacientes con resultados negativos, todos en una edad media de 39 años, y con un 72% del total eran hombres, indicaron que los síntomas más frecuentes fueron el cambio del olfato de del gusto, la fiebre y el dolor corporal fueron los principales síntomas asociados al COVID-19, mientras que la falta de aliento y el dolor de garganta fueron los síntomas más frecuentemente asociados con los pacientes con resultados negativos de la prueba. Un modelo desarrollado para las pruebas de predicción, que incluyó 5 síntomas diagnósticos, tuvo una precisión del 82% en la capacidad de predicción para prueba COVID-19 positiva (34). Para poder maximizar la sensibilidad y aún conservar una exactitud en el

diagnóstico, se usó una combinación de 2 síntomas, que fueron los cambios en el gusto y el olfato y la fiebre, lo que mostró una sensibilidad del 70% y una precisión de discriminación del 75%. Se concluyó que el predictor determinante más importante para el resultado de COVID-19 positivo fueron los cambios en el olfato y el gusto, y sumado a la fiebre, se predice de forma correcta en un 75% (34).

2.2.2 Predicción individualizada de riesgo para prueba positiva de la enfermedad de Coronavirus 2019 (35)

El siguiente estudio describe predicciones de COVID-19. Desde el surgimiento del COVID-19, el mundo entero se ha visto afectado por este virus, y debido a que se ha trabajado a marchas forzadas por conocer las características del virus, y también desarrollar las vacunas eficaces en contra de él, aún faltan conocimientos prácticos para poder hacer frente a esta pandemia (35). La predicción individualizada consistió en el desarrollo de un registro prospectivo de todos los pacientes que fueron evaluados en Cleveland Clinic. Los casos que se tomaron para la probabilidad fueron lo que contaban con pruebas faríngeas o nasales positivas a COVID-19. Se realizó una mínima reducción y se desarrolló un algoritmo con un selector operacional de regresión logística para quitar las variables que no contribuían al índice de concordancia de validación cruzada del modelo (35). Después se le dio validación externa en una cohorte en diferente tiempo y lugar. El modelo estadístico para la predicción se determinó como un nomograma y se desarrolló en un calculador de riesgo en línea. En la cohorte del desarrollo, se tuvieron 11,672 pacientes que cumplieron los criterios requeridos para el estudio (datos de edad, sexo, síntomas, socio economía), y de ellos 818 (7%) dieron positivo en sus pruebas para COVID-19. En la cohorte de validación se tuvieron 2,295 pacientes que cumplieron con todos los criterios, y de ellos, 290 dieron positivo a COVID-19 en su prueba (35). Se encontró que se tenían un mayor riesgo de ser positivo a COVID-19 los siguientes factores: hombre, de ascendencia africana, persona mayor y los que tuvieron una exposición al virus, y un menor riesgo los que tenían polisacárido neumococal o contaban con su vacuna antigripal (influenza) o que tomaban melatonina, paroxetina o carvedilol. El modelo desarrollado en esta investigación tuvo una discriminación favorable (de 0.863 en la cohorte de desarrollo y un 0.840 en la de validación) y en la calibración. (35).

2.2.3 Más allá de predecir el número de infecciones: Predicción de quién es probable que sea COVID negativo o positivo (36)

Se describe el anterior trabajo citado, el cual tiene como objetivo el predecir la probabilidad de que una persona sea COVID positiva o negativa antes de realizar las pruebas de laboratorio, lo que podría ayudar a los sistemas de salud para aliviar la carga de los recursos y del personal, relacionados con la aplicación de pruebas. Todo esto mediante la identificación de los factores de riesgo de infecciones COVID-19, utilizando información de fácil acceso, lo cual permite prevenir y controlar este virus y en general enfermedades infecciosas, de una forma más específica, en el caso de que en el país exista escasez de recursos médicos, tales como las pruebas COVID-19 (36). El estudio hizo una encuesta primaria en 521 adultos en abril de 2020 en Irán, con un 3% que afirmó ser COVID-19 positivo y con un 15% que no estaban seguros de tener el virus, con ello se realizó el nivel de significancia de 5%, de lo cual se pudo determinar que los encuestados que hicieron ejercicio tuvieron menos probabilidades de ser positivos en COVID-19, y con cada hora más de ejercicio que hicieron por día, se predijo un aumento al 78% de ser negativo en el virus, así también se determinó que los adultos con problemas de salud crónicos presentaron un 48% más de probabilidades de ser COVID-19 positivos que los que no las tenían, se encontró que las personas reclusas en casa o que realizan trabajo en casa son los menos probables en ser COVID-19 positivos, y los adultos empleados en organizaciones más grandes tuvieron menos probabilidades de ser positivos con la enfermedad (36). En este estudio se determinó que en general, para poder prevenir y controlar enfermedades infecciosas, es de gran importancia determinar cuáles son los mayores factores de riesgo, en este caso, la información de los del COVID-19 se han vuelto de muy fácil acceso, y estos mismos factores pueden ayudar a seguir desarrollando la investigación acerca de los modelos de predicción del virus (36).

2.2.4 Modelos de predicción para el diagnóstico y pronóstico de COVID-19: revisión sistemática y evaluación crítica (37)

A continuación se describe una revisión sistemática que tiene como objetivo los modelos de predicción para el diagnóstico y pronóstico, fue evaluar la validez y utilidad de los informes publicados para el diagnóstico del COVID-19 en pacientes que tengan la sospecha de

infección del virus, esto derivado de la presencia de la sintomatología, las predicciones se utilizarán para el pronóstico de pacientes y para la detección en personas de la población general que esté en mayor riesgo de contraer el virus por diversos factores, o los que son ingresados en los hospitales por padecer la enfermedad (37). Se realizó una revisión sistemática y una evaluación crítica en el grupo con presencia de COVID-19, en el año de 2020. El modelo que se desarrolló fue uno de predicción multivariable de COVID-19, la recolección de datos se realizó por dos investigadores diferentes que utilizaron el “CHARMS (evaluación crítica y extracción de datos para revisiones sistemáticas de estudios de modelación de predicción) lista de verificación; el riesgo de sesgo se evaluó utilizando PROBAST (herramienta de evaluación del riesgo de sesgo del modelo de predicción)”. Los resultados obtenidos fueron 169 estudios que describen 232 modelos de predicción, se identificaron 7 modelos de predicción de riesgo en la población general, 118 para diagnóstico de detección de COVID-19 (75 basados en imágenes médicas, 10 para diagnosticar la gravedad) y 107 de pronósticos para predecir el riesgo de mortalidad, progresión a enfermedad grave, ingreso en la UI, ventilación, intubación o duración de ingreso hospitalario (37). Los mayores predictores en los modelos de predicción son: los signos vitales, la edad, y las enfermedades previas. Se encontró que los síntomas son parecidos a la gripe fueron con frecuencia predictivos en los modelos de diagnóstico, así como el sexo, la proteína C reactiva y los recuentos de linfocitos fueron factores frecuentes en los de pronósticos. Los modelos se evaluaron de acuerdo a su alto o poco claro riesgo de sesgo, esto debido a la selección no representativa de los pacientes de control y de exclusión de los que no habían tenido el evento esperado del estudio, muchos de los modelos resultantes no incluyeron la descripción de la población objetivo y solo 11 fueron validados por un gráfico de calibración, se encontró que los modelos diagnósticos Qcovid, la puntuación PRIEST, Carr, el ISARIC4C, de deterioro, y Xie tuvieron mejor predicción y bajos riesgos de sesgo, por ejemplo: predictores como morbilidades, edad y masa corporal para hospitalización y mortalidad, y el PRIEST 262 predijo muerte o riesgo de soporte con las puntuaciones vitales presentadas por los pacientes (frecuencia respiratoria, saturación, temperatura, etc.) (37).

2.2.5 Estimación temprana de los factores de riesgo para hospitalización y mortalidad por COVID-19 en México (38)

A continuación se describe otro estudio en donde se considera que, debido al gran número de enfermedades crónicas no degenerativas en el país, el COVID-19 ha tenido tan altas probabilidades de complicación en los pacientes contagiados, así como los altos porcentajes de mortalidad que se han presentado en la población mexicana (38). El objetivo de la investigación es el de estimar los factores de riesgo para la hospitalización y mortalidad de los mexicanos infectados con el virus. Los datos usados en la investigación son públicos y fueron obtenidos del Sistema de Vigilancia Epidemiológica para las Enfermedades respiratorias Virales de la Secretaría de Salud de México, de los registros de casos positivos de SARS-COV-2. A partir de ellos se realizaron dos regresiones logísticas múltiples y se ajustaron estos modelos para determinar la asociación entre la hospitalización y la mortalidad, dentro del cual se incluyeron otras covariables (38). El estudio incluyó a 10,544 sujetos, de los cuales el 57.68% fueron hombres, con una media de edad de 46.47 años. Los hombres resultaron ser aproximadamente 1.54 veces más propensos que las mujeres para hospitalización, mayor mortalidad OR=1.53, así como las personas de 50 a 74 años con respecto a las de 25 a 49 años OR 2.05 y 3.84. También se encontró que si padecían hipertensión, obesidad y diabetes tenían más probabilidades de ser hospitalizados que los que no las tienen y de morir OR=2.10. Mientras que los mayores factores de mortalidad fueron ser hospitalizados, intubados o tener neumonía (OR 5.02, 4.27 y 2.57 respectivamente). En conclusión, las mujeres presentan mayor riesgo a hospitalización y mortalidad por el virus de COVID-19, así como las personas mayores con respecto a los jóvenes. Las personas con enfermedades co-mórbidas presentan el mayor riesgo junto con los ancianos (38).

2.2.6 Desarrollo de un modelo individualizado de predicción de riesgos para COVID-19 utilizando datos de registros electrónicos de salud (39)

Se describe un estudio el cuál se centró en el desarrollo de un modelo de predicción de riesgo individualizado para infección de COVID-19, que fuera preciso e interpretable, el cual podría servir para clasificar las pruebas ya existentes y los recursos médicos en general, de manera más eficiente y así prevenir posible escasez de estos (39). Para el

sistema se desarrolló una calculadora de riesgo interpretable, la cual usa los registros electrónicos de salud provenientes de la Universidad de Alabama en el área de informática para la integración biológica de Birmingham (UAB-i2b2) del depósito de COVID-19. Se generaron clasificadores de riesgo análogos a los clasificadores que han sido más comúnmente usados en los estudios acerca de mayores riesgos de infección para COVID-19. De acuerdo al diseño del modelo, estos puntajes de riesgo pudieron ser calculados de manera sencilla. Los predictores de riesgo que se utilizaron, implementaron un modelado Credit Scorecard en datos longitudinales EHR, proveniente de 7 262 pacientes inscritos en el Sistema de Salud de la Universidad, y fueron evaluados para COVID-19 entre enero y junio de 2020. Se consideró el momento en que aparecieron los síntomas y la gravedad, y después se probaron los efectos producidos al aplicar distintas técnicas para seleccionar las variables (39). Se encontró que en promedio dos semanas antes de que se diagnosticará a los pacientes con COVID-19 positivo, los predictores más comunes fueron los problemas en vías respiratorias (tos, falta de aliento), dolor de garganta y pecho, fumadores y con trastorno depresivo. Se extendió el tiempo considerado y se incluyeron predictivos relacionados a condiciones crónicas previas. El procesamiento de datos, modelado y calculador de riesgos, se pueden aplicar a cualquier paquete de datos EHR, siempre que se siga el formato de datos correspondiente, y se pueden emplear para estimar el riesgo de COVID-19 y así poder optimizar los recursos hospitalarios (39).

2.2.7 Detección de COVID-19: factores del paciente que predicen una prueba PCR positiva (40)

A continuación se describe un estudio que evaluó las características de pacientes positivos a COVID-19, para poder determinar cuáles eran los predictores para una prueba positiva. Se encontró que las personas que habían viajado o que tuvieron exposición con alguna persona contagiada con el virus, tuvieron más probabilidades de dar positivo a la prueba de PCR de COVID-19 (40). Otros de los factores de predicción fueron la sintomatología que presentaron los sujetos del estudio. Se realizó en una clínica en Minnesota, y se examinó a gran escala a pacientes positivos a COVID-19, a los cuales se les proporcionaron cuestionarios antes de realizarse las pruebas, este contenía preguntas acerca de los síntomas que presentaba (se incluyeron los más comunes informados en otros estudios, como fiebre y

dolor de garganta), así como si habían realizado viajes recientemente o estado expuestos a casos confirmados. Se examinó a 48 sujetos positivos y 98 negativos, de los primeros pacientes presentados en el centro y los positivos de manera aleatoria de edad, sexo, fecha en que se presentaron y lugar de la prueba (Minnesota, Wisconsin o Arizona) y examinados en marzo de 2020 (40). Los cuestionarios se manejaron de manera manual por los médicos, para poder identificar las características de cada uno, y poder predecir la detección de la positividad antes de la fecha de su prueba, los viajes resultaron ser factores de gran importancia. Los datos electrónicos se recopilaron y analizaron con en la misma clínica, y el análisis estadístico, las T-pruebas y regresión logística se realizaron con la versión 14 del programa. Las estadísticas descriptivas, las T-pruebas y el análisis de regresión logística se realizaron utilizando JMP® Version 14 (SAS Institute Inc., Cary, NC) (40). La edad media que presentaron los pacientes en la cohorte fue de 46 años, con un poco más hombres que mujeres, se utilizó una estrategia para controles negativos, por lo que no hubo diferencia significativa en las estadísticas de los dos grupos, ambos presentaron similitud en fiebre y tos, y pocas comorbilidades, mientras que el factor de diferencia más grande fueron los viajes a lugares con alta densidad poblacional y el contacto con pacientes confirmados positivos a COVID-19 en las últimas dos semanas. Se desarrolló un modelo con regresión logística multivariable para predicción positiva de COVID-19 con estos dos factores, y se encontró que se tenía 17 veces más de probabilidad de ser positivos al virus si se había tenido contacto con alguna persona infectada (IC 95% 4,6-88,4), y 4.5 veces más por viajes (IC del 95 %: 1,9-12,7) (40).

2.2.8 Predicción temprana e identificación para casos graves de pacientes durante la pandemia de COVID-19: Un modelo de riesgo severo COVID-19 construido por análisis de regresión logística multivariable (43)

El estudio que se describe a continuación se realizó con el objetivo de que se pudiera tener una identificación temprana del virus y que pudieran recibir el tratamiento oportuno, más en específico los pacientes graves, esto para que se pueda reducir la mortalidad del virus del COVID-19 (43). Se tomaron en cuenta las características clínicas y los predictores tempranos de virus en pacientes graves, esto para desarrollar un modelo de predicción para la identificación y clasificación de los pacientes. Se tomaron en cuenta a los pacientes

ingresados al Affiliated Hospital of Air Force Medical University, en un estudio retrospectivo y no intervencionista. Se dividieron en grupos leves y graves, comparando sus datos clínicos, con análisis de regresión logística multivariable para identificar los predictores tempranos de los casos graves, con la curva de características operativas del receptor (ROC) para evaluar los valores predictivos (43). De un total de 40 pacientes, 19 fueron leves y 21 graves, se obtuvieron predictores estadísticamente significativos, como edad, comorbilidades e hipertensión, fiebre, síntomas respiratorios. Con una sensibilidad y especificidad para predecir COVID-19 positivo y negativo de 90,5% y 84,2%, valor predictivo total del 85% (43).

2.2.9 Modelos de predicción COVID-19 y datos no explotados (44)

En la revisión sistemática que se describe a continuación se utilizaron modelos de aprendizaje automático que se basan en técnicas de ajuste de curvas, modelos SEIR/SIR y otras herramientas estadísticas utilizadas en este tipo de trabajos (44). Se buscaba predecir factores como casos confirmados, casos hospitalizados, severidad, capacidad hospitalaria, tasa de contagios, demografía, sectores vulnerables y pobreza, a corto y largo plazo. Se utilizaron modelos matemáticos que pueden ajustar parámetros de manera dinámica y automática en cualquier momento, se utilizaron datos de la Ciudad de New York al 14 de mayo del 2020: 27,567 decesos y 340,661 casos confirmados y se predijeron 2 millones de muertes en Estados Unidos, sin embargo, el modelo no puede ser utilizado para otras zonas, ya que sus características son muy específicas en cada Ciudad (44).

2.3 Planteamiento de problema

Dentro del contexto de la pandemia de COVID-19, en la que todo el mundo ha estado envuelto a finales del 2019, los servicios de salud se han visto rebasados por el gran número de contagios diarios en cada país, incluyendo México, lo que ha propiciado la escasez de insumos y falta de mitigación de contagios, algo que probablemente se podría controlar y mejorar con más pruebas de COVID-19, ya que a pesar de que hay muchos métodos para realizar las pruebas, son muy costosas y poco accesibles. La prueba de laboratorio para detección de SARS-CoV-2, con más eficacia es de PCR, y derivado de la contingencia, su precio aumento de manera desproporcionada, en México a principios de

2023, los precios van de los 950 MX hasta los 1500 MX, mientras que con las pruebas de anticuerpos su precio va de 280-500 MX, sin mencionar que dichas pruebas solo pueden realizarse en las ciudades o lugares que cuenten con los laboratorios equipados (46,47).

De acuerdo a la página oficial de gobierno de CDMX y CONACHyT, a febrero del 2023, se han realizado un total aproximado de 15.5 millones de pruebas y hubo de 156 mil casos sospechosos a nivel Nacional. En la Alcaldía Álvaro Obregón se hicieron más de medio millón de pruebas y hubo más de 12 mil casos sospechosos (1), con la finalidad de disminuir el número de contagios y mitigar la pandemia, se debe aumentar el número de pruebas realizadas con los primeros síntomas que se presenten. Si bien hoy en día los casos graves que requieren hospitalización han disminuido mucho en gran medida por los esquemas de vacunación aplicados, aún se reportan miles de casos de contagios diarios, además de nuevos subgrupos de la variante de Ómicron que han mutado derivados de estos ellos, por lo que sigue siendo una tarea fundamental tratar de frenar los contagios para evitar nuevos brotes y variantes.

En este estudio se desarrolló un modelo de predicción individualizada de prueba positiva para COVID-19, con información de los pacientes que se presentaron a los servicios de los Centros de Salud de la CDMX, que se encuentran en la Alcaldía Álvaro Obregón, registrados en el Sistema Nacional de Vigilancia Epidemiológica (SINAVE) de la Secretaría de Salud de México.

Con la intención de que este modelo tenga un alto intervalo de confianza de predicción y sea accesible para el mayor número de personas de la población mexicana. El modelo se realizó a través de la regresión logística binaria, debido a que nuestra variable dependiente es una variable dicotómica (resultado de la prueba PCR) que se predecirá integrando las variables independientes, tales como: número de ola, la sintomatología (tomando en cuenta también la sintomatología de las variantes de interés, porque el resultado final se centra en la positividad del virus COVID-19, independientemente de la variante con la que se esté contagiado), datos demográficos, co-morbilidades, tabaquismo y contacto con casos conocidos positivos del virus. No se tuvo en cuenta si el paciente está vacunado o no, ya que la vacunación no previene el contagio, sino la hospitalización y la gravedad de los síntomas, lo que se tomó en cuenta es la presencia o ausencia de la sintomatología o de las características ya mencionadas, por tanto, los pacientes que se

presentaron a realizar la prueba en los Centros de Salud, estuvieron conscientes de la presencia leve o moderada de al menos alguno de los síntomas.

2.4 Pregunta de investigación

¿Cuál es el desempeño de un modelo desarrollado para predecir el diagnóstico de COVID-19 en pacientes de los Centros de Salud de Álvaro Obregón, con un método remoto de recolección de datos, tales como la sintomatología, la demografía, contacto con caso sospechoso, y que sea desarrollado a partir de un modelo estadístico de regresión logística binaria?

III Justificación

A inicios del 2023, en México, páginas oficiales reportan cifras de más de 7.38 millones de casos de contagios y 332 mil defunciones, con más de 9 mil casos por día y un promedio de 4 mil casos nacionales a la semana por lo que es vital que los recursos con los que se cuenta sean utilizados de la mejor manera, no obstante la mayoría de la población mexicana ya ha sido vacunada por lo menos una vez, el reporte de casos positivos sigue aumentando, además la Ciudad de México fue declarada en semáforo rojo a inicios del mes de febrero de 2023 (1). Se ha determinado que mientras más pronto se detecte el virus en las personas infectadas, más pronto se les dará una atención médica adecuada, además de que se puede evitar que los contagios se propaguen, al saberse infectados pueden seguir las medidas necesarias, por lo que un modelo de detección de prueba positiva COVID-19 con un amplio porcentaje de eficacia, de bajo costo en comparación de las pruebas comunes y que no necesite ir a un laboratorio, con aplicación remota y de fácil acceso para la población mexicana, tiene implicaciones en el gasto de recursos de la población y del sistema de salud en México, disminuyendo el número de decesos con un diagnóstico y atención médica pronta.

Este modelo de predicción muestra ventajas frente a otros, como, por ejemplo, es el único en Latinoamérica, emplea métodos de recolección y análisis extraídos de una base de datos. Permite que los usuarios registren sus datos de manera fácil y rápida a través del personal de salud en el SINAVE, y que se conozca el resultado de manera más rápida que

las pruebas de laboratorio. Este modelo buscó un índice global de predicción igual o mayor al de otros modelos internacionales parecidos, más de un 80%.

Se pretende que este modelo pueda tener un fuerte impacto social, aplicado al campo de la salud pública y la educación de la población, al dar a los pacientes un método fácil y rápido de predicción de COVID-19, que les brinde seguridad y soporte emocional frente a la posibilidad de contagio y todo el proceso que conlleva.

IV Hipótesis

El modelo de predicción individual de prueba positiva de COVID-19, obtenido a partir de regresión logística binaria y con datos recolectados mediante un cuestionario aplicado a distancia, presenta una capacidad de predicción mayor a 80% para los pacientes de los Centros de Salud de Álvaro Obregón.

V Objetivos

5.1 Objetivo general

Desarrollar y evaluar el desempeño de un modelo para predicción de prueba individual positiva para COVID-19, para los pacientes que se presenten para realizarse la prueba PCR en los Centros de Salud de la Alcaldía Álvaro Obregón, perteneciente a la Ciudad de México, basado en la sintomatología, variables demográficas y contacto con caso sospechoso.

5.2 Objetivos específicos

1. Determinar la sintomatología y los factores predictivos principales determinantes en los pacientes diagnosticados con el virus SARS-COV-2, y después identificarlos en las bases de datos del SINAVE, de los pacientes que acudieron a los Centros de Salud en la Alcaldía Álvaro Obregón en la Ciudad de México.
2. Desarrollar un modelo de predicción logística binaria a partir de la una base de datos creada con la muestra de los Centros de Salud de Álvaro Obregón.
3. Realizar una fase de verificación aplicando los modelos generados en una nueva base de datos de pacientes COVID-19, con la finalidad de evaluar su efectividad.

VI Materiales y métodos

6.1 Contexto de la investigación

Esta investigación se realizó en los Centros de Salud de la Secretaría de Salud en la Alcaldía Álvaro Obregón, en la Ciudad de México. Los pacientes que acudieron a los Centros de Salud para realizarse pruebas confirmatorias de PCR para COVID-19, entre el 11 de mayo del 2020 y 14 de enero del 2023 (olas 1 a 6) abarcado primera y segunda fase, los cuales registraron sus datos en el Sistema de Vigilancia Epidemiológica Nacional (SINAVE).

Tipo y diseño de estudio

Este reporte de investigación tiene un enfoque cuantitativo, no experimental (observacional) y ambispectivo, ya que tendrá una etapa retrospectiva y una etapa prospectiva. Es analítico porque se obtendrán modelos estadísticos de predicción.

6.2 Selección de la población, criterios de inclusión y exclusión

La Alcaldía Álvaro Obregón cuenta con una población aproximada de 749,982 personas, la cual es y ha sido actualmente la más afectada de toda la Ciudad de México, de acuerdo con datos del CONACHyT (2021-2023), con el mayor número de casos y contagios de las Alcaldías de la Ciudad de México y acumulando a febrero del 2023, un total aproximado de 140 mil casos positivos, y en julio del 2022 se registró como el epicentro nacional de contagios, con casi 9 mil por día (1). Se seleccionaron a los pacientes en las bases de datos que acudieron a los Centros de Salud de la Alcaldía Álvaro Obregón para realizarse la prueba de PCR para COVID-19.

Criterios de Inclusión:

- Pacientes que solicitaron la prueba PCR por sospecha de COVID-19 en los Centros de Salud de la Alcaldía Álvaro Obregón.
- Pacientes mayores de 18 años.
- Pacientes que proporcionen la información completa para la muestra.

Criterios de exclusión:

- Pacientes que presenten resultados inconclusos en su prueba PCR.

6.3 Tamaño de muestra y muestreo

El tamaño de la muestra

Se realizaron dos tipos de modelos de predicción, el primer tipo se realizó con el total de la base de datos a la cual se tuvo acceso (SISVER-SINAVE). En el primer tipo de modelos, solo se tomó en cuenta los pacientes seleccionados de las bases de datos totales, como se muestra a continuación: para evaluar el valor predictivo del modelo estadístico se calculó la sensibilidad y la especificidad en 7 muestras, 6 correspondientes a cada una de las olas de COVID-19 y una general. Para el cálculo de sensibilidad y especificidad se tomó en cuenta un estudio previo con un modelo estadístico para predicción de prueba personalizada de COVID donde reportaron valores de sensibilidad = 74.1%, especificidad = 71.4% durante la fase de validación del modelo (34). Para calcular el tamaño de la muestra utilizamos la fórmula para estimación de una proporción en una variable (48).

Estimación de una proporción (Variable cualitativa)

$$N = \frac{Z\alpha^2 P(1 - P)}{i^2}$$

N: número de sujetos necesarios para el estudio.

Z α : valor de Z que corresponde al riesgo de α fijado.

P: es el valor de la proporción que existe en la población de estudio.

i: es la precisión con la que se quiere estimar el parámetro (2*i es la amplitud del intervalo de confianza).

Para evaluar el valor predictivo del modelo estadístico se calculará la sensibilidad y la especificidad, conforme a lo siguiente: Sensibilidad = número de casos positivos en el modelo/ número de verdaderos positivos por resultado de laboratorio. Especificidad = número de casos negativos en el modelo / número de verdaderos negativos por resultado de laboratorio. Estos resultados se expresan como porcentaje (de 0 a 100%).

Para calcular el tamaño de la muestra utilizamos la fórmula para estimación de una proporción en una variable cualitativa (48): considerando un error alfa del 5% con un valor de Z correspondiente = 1.96, y con un nivel de precisión i= 5%, calculamos el número de

sujetos necesarios para cada una de las proporciones correspondientes a los valores reportados por Roland *et al.* 2020.

Para sensibilidad = 74.1% ($p = 0.741$) se estima un número de sujetos necesarios $N = 295$. Para especificidad = 71.4% ($p = 0.714$) se estima un número de sujetos necesarios $N = 314$. Por lo anterior se requieren al menos 314 sujetos o pruebas realizadas. Ajustando por un máximo de 20% de pérdidas, el tamaño de muestra para la validación será de 377 pruebas por cada muestra.

Para el segundo tipo de modelos, se realizaron con el total de la base de datos, se seleccionaron la semana con más contagios en cada una de las olas, y a partir de esta, una semana adelante y una atrás, esto de acuerdo a las cifras reportadas en la Ciudad de México (Figura 2). En la primera ola se obtuvieron fechas del 11 al 31 de mayo del 2020 de acuerdo al criterio de selección estipulado, sin embargo, se obtuvieron nuevas fechas debido a la falta de casos negativos, del 19 de abril al 21 de junio del 2020, esto debido a que en los primeros meses de pandemia solo se realizaban pruebas por COVID-19 a los pacientes que presentaban los síntomas correspondientes a la enfermedad, por lo que prácticamente la totalidad de casos eran positivos. Para las olas 2- 6 se tomó la semana con mayor número de contagios y de esta, una semana adelante y una atrás.

Tabla 1: Características de los datos por cada una de las 5 olas de la pandemia de COVID-19 y en general

	Ola 1	Ola 2	Ola 3	Ola 4	Ola 5	Ola 6
Rango de fechas	19 abril-21 junio 2020	4- 24 enero 2021	19 julio- 8 agosto 2021	3- 23 enero 2022	20 junio- 10 julio 2022	27 noviembre 2022- 14 de enero 2023
Variante principal	<i>cepa original</i>	<i>Alpha, Beta y Gamma</i>	<i>Delta</i>	<i>Ómicron</i>	<i>Ómicron (sublinajes)</i>	<i>Ómicron (sublinajes)</i>
Total de pacientes	2081	13535	15368	6159	1177	3332
No cumplen criterios de inclusión (<18 años)	71	1196	2072	822	167	375
No. Final de pacientes	2010	12339	13296	5337	1010	2957

Figura 1. Casos COVID-19 en México de acuerdo al SINAVE

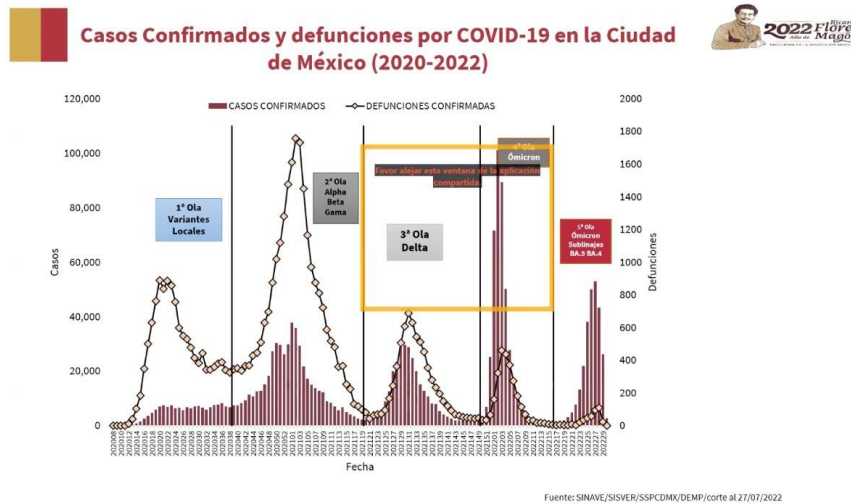


Figura 1: Casos confirmados y defunciones por COVID-19 en la Ciudad de México (2020-2022). Fuente: SINAVE/SISVER/SSPCDMX/DEMP/Corte al 27/07/2022

Para la fase 1 (desarrollo del modelo) se tomaron en cuenta dos bases de datos, una con el total de los pacientes determinados como previamente se mencionó y una muestra aleatoria simple de cada una de las olas de la pandemia de COVID-19 (ya que se ha reportado que en cada ola ha sido predominantes distintas variables del virus), para las consideraciones pertinentes en cuanto a las diferentes variantes, en total fueron 6 muestras (con 377 pacientes cada una) y una general con 2232 pacientes.

Para la Fase 2: validación del modelo de predicción, se determinaron las fechas de selección de la nueva base de datos del 27 de noviembre al 14 de enero del 2023 (ola 6), se aplicaron los modelos a la base de datos con un total de 2957 pacientes.

6.4 Diagrama de diseño de estudio

En el diagrama de la Figura 1 se describen las fases del estudio, Fase 1: Desarrollo del modelo estadístico de predicción y Fase 2: Validación del modelo de predicción.

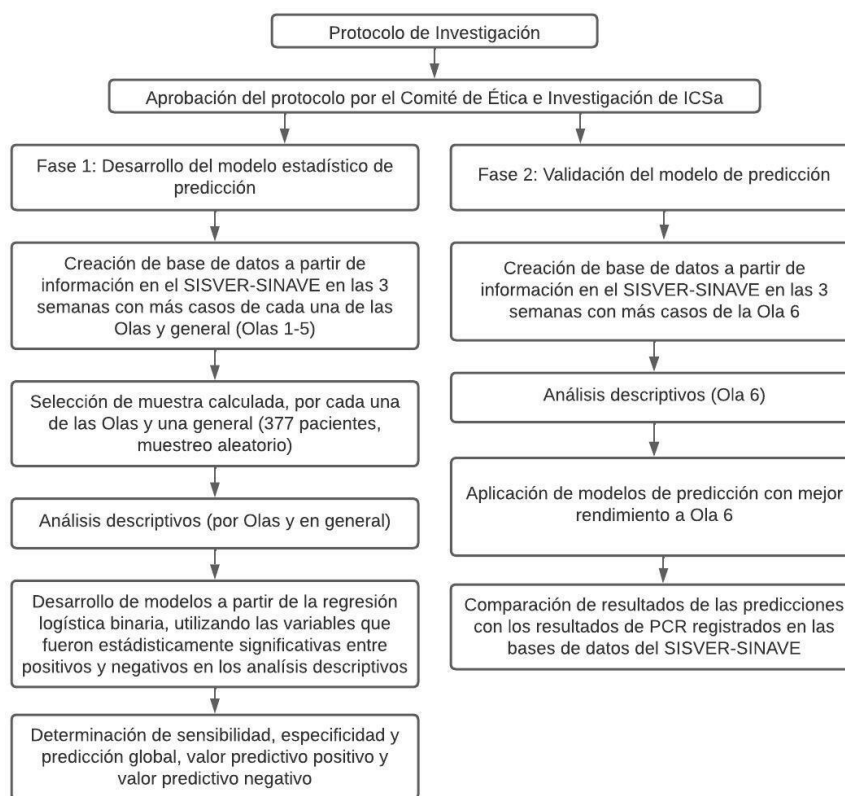


Figura 2. Diagrama de flujo del diseño del estudio, identificando las fases del mismo.

6.5 Definición de variables (dependientes, independientes)

Tabla 2: Descripción de variable dependiente y variables independientes

Variable	Tipo de variable	Definición de variable	Medición
Resultado de la prueba PCR para COVID-19	Dependiente, cualitativa	Prueba de diagnóstico para detectar el COVID-19	Cualitativa, dicotómica: positiva, negativa
Número de ola	Independiente, cuantitativa	Ola a la que pertenece el paciente (1-6)	Cuantitativa: número
Edad del paciente	Independiente, cuantitativa	Tiempo de vida en años desde el nacimiento	Cuantitativa, continua: años
Sexo del paciente	Independiente, cualitativa	Sexo biológico con el que nace el paciente	Cualitativa discreta: femenino, masculino
Dolores torácico	Independiente,	Molestia en el tórax	Cualitativa,

	cualitativa	no proveniente de alguna condición conocida, en cualquier nivel	dicotómica: si, no
Fiebre	Independiente, cualitativa	Temperatura corporal mayor a la normal (36.5°)	Cualitativa, dicotómica: si, no
Escalofríos	Independiente, cualitativa	Temblores por frío sin estar a baja temperatura	Cualitativa, dicotómica: si, no
Disnea	Independiente, cualitativa	Dificultades de cualquier nivel para inhalar o exhalar	Cualitativa, dicotómica: si, no
Cefalea	Independiente, cualitativa	Dolores de cabeza recurrentes sin razón aparente	Cualitativa, dicotómica: si, no
Mialgia	Independiente, cualitativa	Dolores musculares sin razón aparente	Cualitativa, dicotómica: si, no
Artralgia	Independiente, cualitativa	Dolores en las articulaciones sin razón aparente	Cualitativa, dicotómica: si, no
Rinorrea	Independiente, cualitativa	Nariz congestionada o tapada	Cualitativa, dicotómica: si, no
Polipnea	Independiente, cualitativa	Respiración que muestra ser superficial o muy rápida	Cualitativa, dicotómica: si, no
Vómito	Independiente, cualitativa	Expulsión sin quererlo del contenido del estómago	Cualitativa, dicotómica: si, no
Dolor abdominal	Independiente, cualitativa	Dolores en el abdomen o en los músculos	Cualitativa, dicotómica: si, no
Conjuntivitis	Independiente, cualitativa	Problemas de infección o inflamación en el ojo o parpado	Cualitativa, dicotómica: si, no
Cianosis	Independiente, cualitativa	Piel con color distinto al normal (azul o gris)	Cualitativa, dicotómica: si, no
Anosmia	Independiente, cualitativa	Perdida del olfato	Cualitativa, dicotómica: si, no
Disgeusia	Independiente,	Algún tipo de	Cualitativa,

	cualitativa	alteración en el sentido del gusto	dicotómica: si, no
Diabetes	Independiente, cualitativa	Enfermedad crónica que afecta la conversión de alimentos	Cualitativa, dicotómica: si, no
Hipertensión	Independiente, cualitativa	Enfermedad que afecta la presión en la sangre	Cualitativa, dicotómica: si, no
Obesidad	Independiente, cualitativa	Masa corporal mayor a la normal	Cualitativa, dicotómica: si, no
Tabaquismo	Independiente, cualitativa	Adicción al tabaco	Cualitativa, dicotómica: si, no
Contacto con caso conocido	Independiente, cualitativa	Contacto con personas contagiadas de COVID-19	Cualitativa, dicotómica: si, no
Tos	Independiente, cualitativa	Expulsiones bruscas de aire por afección en los pulmones	Cualitativa, dicotómica: si, no
Odinofagia	Independiente, cualitativa	Dolor por deglución	Cualitativa, dicotómica: si, no
Diarrea	Independiente, cualitativa	Evacuaciones líquidas más de 3 veces por día	Cualitativa, dicotómica: si, no
Irritabilidad	Independiente, cualitativa	Sensación enojo o frustración sin razón aparente	Cualitativa, dicotómica: si, no

Fuente: Diccionario médico de la Real Academia Española; 2023; <https://dtme.ranm.es/index.aspx>, elaboración propia.

6.5.1 Prueba de regresión logística binaria

Principio

Se determinó utilizar la regresión logística binaria debido a su amplio uso en la epidemiología, ya que se usa cuando la variable dependiente cualitativa dicotómica se quiere predecir o se desea saber la relación con ella a partir de las variables independientes cualitativas o cuantitativas dicotómicas. Otra de las razones para elegir la regresión logística es una función de esta misma para realizar una serie de combinaciones de distintas formas, para encontrar las variables que predicen mejor a la variable dependiente, en este caso se utilizó la función de “Adelante condicional” en 30 iteraciones en SPSS.

Se compararon las variables entre grupos de pacientes positivos y negativos con los métodos descritos en la sección de análisis estadístico.

Las variables con diferencias significativas entre el grupo con prueba positiva y el grupo con prueba negativa se incluyeron en el análisis de regresión logística binaria adelante condicional por pasos (realizada ya que se tiene una variable dependiente: resultado de la prueba PCR, que se quiere predecir a partir de variables independientes: síntomas, morbilidades y datos socio-economicos), para la creación de modelos de predicción individual de prueba positiva de COVID, considerando el resultado de la prueba PCR como variable dependiente y las demás características del paciente como variables independientes. Para cada variable independiente se obtuvo la razón de momios u “odds ratio” (O.R.) con intervalo de confianza al 95% y el valor de p . Se utilizó el método de solución de los modelos de regresión logística de “Adelante condicional” y se obtuvo el valor del coeficiente de determinación (R^2) como parámetro de bondad de ajuste del modelo final.

Con los modelos obtenidos se calcularon los valores de predicción del modelo (sensibilidad, especificidad, porcentaje global, valor predictivo positivo, valor predictivo negativo) comparados contra el resultado de la prueba de COVID obtenido con PCR (estándar de oro). Los análisis estadísticos se realizaron con SPSS versión 21.0 y Excel versión Office 2010, se consideró significativo un valor de $p < 0.05$ y $p < 0.01$, dependiendo de la tabla.

Procedimiento

Este reporte fue aprobado por el Comité de Ética e Investigación de ICSa, con el número de folio Comiteei.icsa ICSA «017» / 2023.

Las muestras de cada uno de los pacientes se tomaron en los centros COVID designados especialmente para este tipo de atención, por medio de exudados faríngeos y nasofaríngeos; fueron enviadas al InDRE y posteriormente los resultados fueron cargados en la base de datos del SINAVE.⁸ Los resultados se registraron como “Resultado definitivo” y se clasificaron como: “SARS-Cov-2”, “Negativo”, “Rechazada”, “No adecuado”, “No subtipificado”, “Positivo” y campos vacíos, se tomaron como positivos: “SARS-Cov-2” y “Positivo”, como negativo: “Negativo” que fueron recodificados en 1 y 0

respectivamente para el análisis y se eliminaron los demás criterios al no ser adecuados para el análisis (Anexo 2).

Las variables incluidas en el análisis de regresión fueron seleccionadas a partir de una extensa revisión de literatura de las principales características que presentaron los pacientes diagnosticados como positivos con COVID-19 en cada una de las olas, dichas características cambiaron con respecto al subtipo del virus de SARS-Cov-2. Para los modelos se incluyeron las principales características de cada uno de los subtipos, debido a que las pruebas aplicadas por el SINAVE solo daba como resultado “POSITIVO” o “NEGATIVO”, no del subtipo específico si es que había contagio.

Se incluyó la información de los siguientes campos: sexo, edad, comorbilidades (diabetes, hipertensión y obesidad)⁹, tabaquismo, contacto con caso sospechoso y presencia de síntomas (fiebre, tos, odinofagia, disnea, irritabilidad, diarrea, dolor torácico, escalofríos, cefalea, mialgia, artralgia, rinorrea, polipnea, vómito, dolor abdominal, conjuntivitis, cianosis, anosmia, disgeusia). Todas las variables excepto edad y sexo se clasificaron como “SI” y “NO” y fueron recodificados en 1 y 0 respectivamente. Se excluyeron los casos con datos capturados como “SE IGNORA” y campos vacíos.

Se consultó la base de datos del SISVER/SINAVE para obtener todos los casos registrados para los Centros de Salud de la alcaldía Álvaro Obregón durante cada pico de las primeras 6 olas de la pandemia (Tabla 1). Se identificaron las fechas con mayor número de casos de cada ola. Para las olas 2, 3, 4, 5 y 6 se seleccionó un intervalo de una semana previa y una semana posterior a la semana con más casos (periodo total de muestreo de 3 semanas por ola). Durante la ola 1 había poca disponibilidad de pruebas y los criterios para definir casos sospechosos fueron más estrictos, por lo que se registró una tasa baja de casos negativos. Por lo tanto, para la muestra de la ola 1 se seleccionó un intervalo de dos semanas previas y dos semanas posteriores a la semana con más casos (periodo total de muestreo de 5 semanas).

6.6.1 Análisis Estadístico

La edad no tuvo distribución normal (prueba de Kolmogorov-Smirnov, $p < 0.01$), por lo que se describe como mediana (percentil 25 - percentil 75) y se comparó entre olas por medio de Kruskal Wallis y entre resultados positivo o negativo de cada ola mediante prueba

U de Mann-Whitney. Las variables nominales se describen como recuento y porcentaje y se compararon entre olas y entre resultado de la prueba (positivo o negativo) con prueba Chi-cuadrada.

Para la predicción del resultado de la prueba de SARS-CoV-2 se realizó un análisis de regresión logística binaria, obteniendo un modelo de regresión por cada una de las olas y se incluyeron solo las variables que mostraron tener significancia entre los grupos de positivos y negativos de cada ola. Se utilizó el método de adelante condicional con la bondad de ajuste Hosmer-Lemeshow. Para las variables significativas en cada modelo se reporta la razón de momios con un intervalo de confianza (I.C.) de 95%. El análisis estadístico y la regresión logística se realizaron con el programa SPSS, versión 21.0 y Excel versión Office 2010. Se consideró como significativo los valores de $p < 0.01$.

Para la fase de verificación del modelo, se realizaron los análisis descriptivos de la base de datos de la ola 6, seleccionada a partir de la semana con más casos y una semana adelante y una atrás (27 noviembre 2022-14 enero 2023), con un total de 3332 pacientes, a los cuales se les aplicaron los criterios de eliminación, quedando con un total final de 3957 pacientes.

Finalmente, los resultados para prueba positiva por COVID-19, calculada a partir de la fórmula general para la regresión logística binaria se muestran en la tabla 13. Para cada uno de los modelos realizados de las olas 1 a 5, la general y la ola 6. Se muestra el número de ola, y las dos secciones correspondientes a los modelos de la base de datos total y de la muestra, posteriormente a partir de la aplicación de la fórmula y de la comparación con los resultados reales de la prueba PCR de los pacientes, se calcularon los verdaderos negativos (VN), falsos negativos (FN), falsos positivos (FP) y verdaderos negativos (VN). Después se realizó nuevamente un conteo total de los casos, ya que Excel no realizó los ajustes en el caso de valores perdidos como lo hace SPSS, esto para calcular correctamente los porcentajes obtenidos, se determinó la N completa, la N de los clasificados, el total de negativos y el total de positivos de cada uno de los modelos. Posteriormente, se calcularon los porcentajes de los correctos negativos, correctos positivos, el porcentaje global correcto, la sensibilidad, la especificidad, el valor predictivo positivo y el valor predictivo negativo.

VII Resultados

7.1 Análisis del total de la base de datos

La Tabla 3 muestra las características generales de los participantes del estudio por cada una de las olas, haciendo una comparación de estos resultados entre todas las olas. Los símbolos de asterisco (*) muestran que la ola 1 comparada con las otras olas (2-6), tuvo mayores cifras en edad, mujeres, fumadores, hipertensión, obesidad, diabetes y contacto con casos sospechosos. Comparado con la ola 2 (símbolo &), la edad fue menor que en las olas 3 y 4 pero mayor en la 5 y 6, mayor proporción de fumadores (comparado con olas 3, 5 y 6), más hipertensión y diabetes (comparado con 3 y 4), menos obesidad (comparado con la ola 4) y menor contacto con casos sospechosos (3, 4, 5, 6), menor proporción de comorbilidades (hipertensión, obesidad, diabetes) y de contacto con casos sospechosos que la ola 1. Comparando con la ola 3 (símbolo #), las olas 4, 5 y 6 tuvieron diferencias significativas en edad, número de fumadores, proporción de pacientes con hipertensión, obesidad (sólo en la ola 4), diabetes y casos sospechosos (olas 5 y 6). La ola 6 perteneciente a la fase de validación (símbolo ≠) mostró tener diferencias estadísticamente significativas con las otras olas en todas las variables.

Tabla 3. Características socio-demográficas y morbilidades de la población de estudios.

	Ola 1	Ola 2	Ola 3	Ola 4	Ola 5	Ola 6
Variable	(n = 2009)	(n = 12339)	(n = 13296)	(n = 5337)	(n = 1010)	(n=2957)
Edad (años)	46	40 *	35 * ^{&}	38 * ^{&#}	43 * ^{&#S}	43 * ^{&#S} ≠
KW	(35 – 57)	(29 – 52)	(26 – 48)	(29 – 51)	(31 – 58)	(30 – 55)
	*	*	*	*	*	*

Sexo		*	*	*	*&\$	*&#\$
Femenino	1006 (50%)	6668 (54%)	7306 (55%)	2854 (53%)	590 (58%)	1754 (59.3%)
Masculino	1003 (50%)	5671 (46%)	5990 (45%)	2483 (47%)	420 (42%)	1203 (40.7%)
Fumador	240 (12%)	1234 (10%)	1036 (8%) * ^{&}	560 (11%) [#]	41 (4%) * ^{&#\$}	180 (6.1%) * ^{&#\$} ≠
Hipertensión	367 (18%)	1124 (9%) *	818 (6%) * ^{&}	267 (5%) * ^{&#}	112 (11%) * [#] \$	255 (8.6%) * ^{&#\$} ≠
Obesidad	355 (18%)	672 (5%) *	721 (5%) *	110 (2%) * ^{&#}	41 (4%) * ^{\$}	180 (6.1%) * ^{&#\$} ≠
Diabetes mellitus	301 (15%)	941 (8%) *	606 (5%) * ^{&}	212 (4%) * ^{&}	73 (7%) * [#] \$	176 (6.0%) * ^{&#\$} ≠
Contacto con caso sospechoso	833 (62%)	7202 (59%)	5198 (39%) * ^{&}	2125 (41%) * ^{&}	236 (24%) * ^{&#\$}	615 (20.8%) * ^{&#\$} ≠

Tabla 3: Características sociodemográficas y co-morbilidades de la población de estudios, muestra recuento y porcentajes o mediana (percentil 25 – percentil 75). U de Mann-Whitney y chi-cuadrada. * $p < 0.01$ vs ola 1, [&] $p < 0.01$ vs ola 2, [#] $p < 0.01$ vs ola 3, ^{\$} $p < 0.01$ vs ola 4, [≠] $p < 0.01$ vs ola 5. KW: Kruskal Wallis que muestra ser estadísticamente significativa en comparación con las otras olas. Elaboración propia.

La Tabla 4 muestra los datos socio-demográficos y morbilidades de los participantes comparados respecto al resultado de la prueba PCR (positivo o negativo), los asteriscos indican diferencias estadísticamente significativas entre positivos y negativos dentro de una misma ola ($p < 0.01$). En la ola 1 las variables diferentes entre positivos y negativos (indicadas con *) son: edad, sexo, hipertensión y diabetes. En la ola 2 las variables que contrastaron son: edad, fumador, hipertensión, obesidad, diabetes y contacto con caso sospechoso. En la ola 3 los positivos distintos a los negativos en: edad, sexo, obesidad, diabetes y contacto con caso sospechoso. Con respecto a la ola 4 positivos y negativos se diferencian en: hipertensión, obesidad y contacto con caso sospechoso,

mientras que en la ola 5 las variables con diferencia entre positivos y negativos fueron: edad, fumador, hipertensión, obesidad, diabetes y contacto con caso sospechoso. En la ola 6 la edad, la diabetes mellitus y el contacto con caso sospechoso fueron estadísticamente significativos comparando positivos vs negativos.

Tabla 4. Características socio-demográficas y morbilidades de la población de estudios entre positivos y negativos

Variable	Ola 1		Ola 2		Ola 3		Ola 4		Ola 5		Ola 6	
	Positivo (n = 1119)	Negativo (n = 874)	Positivo (n = 6372)	Negativo (n = 5363)	Positivo (n = 5435)	Negativo (n = 7658)	Positivo (n = 1150)	Negativo (n = 3229)	Positivo (n = 298)	Negativo (n = 616)	Positivo (n = 961)	Negativo (n = 1532)
Edad (años) \bar{x}	48 (37 – 60)	43 * (34 - 55)	40 (30 – 52)	39 * (28 – 51)	34 (25 – 46)	36 * (26 – 49)	39 (29 – 51)	38 (29 – 50)	49 (33 – 66)	40 * (28.5 – 55)	45 (33-58)	40* (30-55)
Sexo		*				*						
Femenino	45.0%	56.4%	54.2%	53.8%	52.5%	56.8%	53.0%	53.7%	57.4%	59.3%	59.9%	59.7%
Masculino	55.0%	43.6%	45.8%	46.2%	47.5%	43.2%	47.0%	46.3%	42.6%	40.7%	40.1%	40.3%
Fumador	11.9%	12.2%	11.2%	8.5% *	8.4%	7.4%	11.6%	11.0%	7.1%	2.3% *	5.8% %	6.3%
Hipertensión	21.6%	14.4%*	10.3%	7.5% *	6.5%	5.8%	7.0%	4.9% *	20.3%	5.5% *	9.4%	8.7%
Obesidad	19.0%	16.1%	7.1%	3.1% *	6.6%	4.6% *	3.6%	2.0% *	7.8%	2.3% *	5.8%	6.1%
Diabetes mellitus	20.0%	8.9% *	8.6%	6.3% *	5.3%	3.9% *	5.3%	3.9%	11.9%	3.9% *	5.8%	6.9%*
Contacto con caso sospechoso	58.0%	65.0%	54.7%	62.7% *	42.5%	37.1% *	44.3%	37.1% *	37.5%	18.5% *	25.3%	19.9%*

Tabla 4: * $p < 0.01$ comparado con positivos de la misma ola. Las significancias se calcularon a partir de Chi-cuadrada o U de Mann Whitney. En la tabla se muestra comparación entre positivos y negativos de cada una de las olas, como porcentajes o mediana (percentil 25 – percentil 15). Elaboración propia.

En la Tabla 5 se compara el porcentaje de presencia de síntomas en los pacientes con prueba positiva VS negativa dentro de cada ola y en general. Al realizar las comparaciones por olas de contagio (1-6) se hallaron diferencias estadísticamente significativas en todas las variables: tos, rinorrea, disnea, polipnea, cianosis, odinofagia, anosmia, disgeusia, diarrea, dolor abdominal, vómito, dolor de cabeza, mialgia, artralgia, dolor torácico, fiebre, escalofríos, irritabilidad y conjuntivitis, y de manera específica la ola 1 no las tuvo en: cianosis, odinofagia, diarrea, dolor abdominal, vómito, irritabilidad y conjuntivitis. La ola 4 no tuvo diferencias en: disnea, polipnea, cianosis, anosmia, disgeusia, vómito y conjuntivitis. La ola 5 no presentó diferencias en: cianosis, anosmia y vómito. En la ola 6 excepto en polipnea, cianosis y dolor abdominal.

Tabla 5: Prevalencia de los síntomas entre positivos y negativos

Variable	Ola 1		Ola 2		Ola 3		Ola 4		Ola 5		General		Ola 6	
	Positivo (n = 1119)	Negativo (n = 874)	Positivo (n = 6372)	Negativo (n = 5363)	Positivo (n = 5435)	Negativo (n = 7658)	Positivo (n = 1150)	Negativo (n = 3229)	Positivo (n = 298)	Negativo (n = 616)	Positivo (n = 14374)	Negativo (n = 17740)	Positivo (n = 961)	Negativo (n = 1532)
Tos	80.8	61.8 *	51.4	22.4 *	62.5	81.1 *	19.3	21.3 *	48.7	14.8 *	58.3	22.4 *	60.3	53.9*
Rinorrea	34.4	27.0 *	27.8	9.0 *	26.7	37.4 *	5.4	6.5 *	28.3	11.0 *	28.1	9.0 *	41.3	35.7*
Disnea	46.7	26.2 *	9.9	5.1 *	8.5	20.9 *	4.5	3.9	2.5	1.7 *	11.9	5.0 *	9.8	14.6*
Polipnea	21.4	12.1 *	3.2	1.6 *	3.1	2.7 *	0.5	1.1	0.1	0.1 *	4.3	1.6 *	3.2	4.9
Cianosis	6.9	4.2	1.9	0.7 *	1.4	0.7 *	0.2	0.5	0.1	0.1	1.9	0.7 *	0.9	1.9
Odinofagia	46.8	41.4	28.6	11.2 *	29.0	37.2 *	7.6	6.9 *	37.6	11.4 *	30.1	10.7 *	41.2	39.2*
Anosmia	17.1	8.9 *	19.0	4.8 *	14.3	2.1 *	0.7	1.6	2.0	1.1	15.4	2.7 *	4.1	2.5*
Disgeusia	17.9	8.1 *	16.8	4.6 *	12.5	2.8 *	0.5	1.6	1.7	0.9 *	13.7	2.6 *	3.4	2.5*
Diarrea	23.6	20.9	9.6	4.6 *	11.3	10.4 *	4.7	6.7 *	5.7	2.8 *	11.1	6.0 *	10.9	9.9*

Dolor abdominal	13.9	12.3	5.3	1.8 *	7.3	17.2 *	3.4	3.2 *	3.3	1.5 *	6.8	2.9 *	5.9	7.3
Vómito	9.9	7.1	3.3	1.4 *	4.3	5.4 *	1.9	1.5	2.0	1.1	4.1	1.7 *	3.0	6.3*
Dolor de cabeza	75.2	70.0 *	44.4	19.7 *	45.6	65.0 *	17.0	15.5 *	32.0	12.3 *	46.7	18.9 *	48.1	46.2*
Mialgia	59.7	43.8 *	25.3	8.2 *	26.2	48.1 *	10.7	4.7 *	22.5	6.4 *	28.5	8.2 *	34.1	31.3*
Artralgia	57.7	41.3 *	21.4	6.2 *	20.8	30.0 *	5.9	3.6 *	13.9	3.6 *	23.4	6.3 *	27.5	23.6*
Dolor torácico	31.4	22.0 *	14.3	6.0 *	15.4	30.3 *	8.6	6.2 *	7.0	2.5 *	15.8	6.3 *	16.8	16.9*
Fiebre	71.8	45.2 *	33.4	13.0 *	42.2	52.7 *	9.3	10.6 *	18.3	4.2 *	38.9	11.8 *	35.3	31.3*
Escalofríos	43.8	29.1 *	21.8	6.5 *	27.5	40.7 *	6.7	7.2 *	22.0	5.7 *	26.0	7.7 *	28.6	26.6*
Irritabilidad	19.4	16.0	8.7	5.1 *	10.1	13.1 *	5.4	5.9 *	2.9	0.7 *	9.7	5.2 *	11.6	11.9*
Conjuntivitis	10.9	11.3	7.3	2.2 *	8.5	15.5 *	2.3	2.7	1.6	1.1 *	7.8	2.7 *	7.8	10.8*

Tabla 5: * $p < 0.01$ comparado con positivos de su grupo en cada una de las olas y general. Se muestran porcentajes. Las significancias se calcularon a partir de Chi-cuadrada. Elaboración propia.

La Tabla 6 presenta los resultados del análisis de regresión logística con modelos para predecir el resultado de la prueba de SARS-CoV-2 (positivo o negativo) como variable dependiente. En cada ola se calculó un modelo distinto que incluyó como variables independientes aquellas que tuvieron diferencias significativas entre positivos y negativos (Tablas 4 y 5). Por ejemplo, el modelo inicial de la ola 1 incluyó edad, sexo femenino, hipertensión, diabetes, tos, rinorrea, disnea, polipnea, anosmia, disgeusia, cefalea, mialgia, artralgia, dolor torácico, fiebre y escalofríos. Las variables que no se incluyeron inicialmente en el modelo de la ola 1, indicadas como “No incluido”, son: fumador, obesidad, contacto con caso sospechoso, cianosis, odinofagia, diarrea, dolor abdominal, vómito, irritabilidad y conjuntivitis. Al aplicar el análisis de regresión logística en el modelo 1 con el método adelante condicional, se eliminaron las variables que no tuvieron contribución significativa a la variable dependiente, indicadas como “Eliminado”, en las variables de edad, hipertensión, rinorrea, polipnea, anosmia, mialgia, artralgia, dolor torácico y escalofríos. El modelo final de la ola 1 tuvo las siguientes variables con razón de momios significativa ($p < 0.05$): sexo femenino, diabetes, tos, disnea, disgeusia, cefalea y fiebre.

Aplicando la misma estrategia de análisis de regresión logística, los modelos finales de las olas subsecuentes tuvieron diferentes variables predictoras (Tabla 5). El modelo final de la ola 2 tuvo como variables predictoras a la edad, fumador, obesidad, contacto con caso sospechoso, tos, rinorrea, disnea, polipnea, odinofagia, anosmia, disgeusia, vómito, cefalea, artralgia, dolor torácico, fiebre, escalofríos, irritabilidad, y conjuntivitis. El modelo final de la ola 3 tuvo como variables predictoras a la edad, sexo femenino, tos, rinorrea, cianosis, odinofagia, anosmia, disgeusia, diarrea, dolor abdominal, cefalea, mialgia, artralgia, fiebre, escalofríos e irritabilidad. El modelo final de la ola 4 tuvo como variables predictoras finales a tos, odinofagia, mialgia, fiebre, escalofríos e irritabilidad. El modelo final de la ola 5 tuvo como variables predictoras la edad, tos, rinorrea, cefalea, fiebre y escalofríos. Y la ola 6 incluyó las variables de edad, contacto con caso sospechoso, tos, disnea, vómito, artralgia y conjuntivitis, excluyendo o eliminando a las otras variables. La ola 6 incluyó las variables de edad, contacto con caso sospechoso, tos, disnea, vómito, artralgia y conjuntivitis, excluyendo o eliminando a las otras variables.

Con la misma estrategia de análisis de regresión se calculó un modelo general con los datos de las 5 primeras olas (Tabla 5). El modelo final general tuvo las siguientes variables: sexo femenino, hipertensión, diabetes, tos, rinorrea, disnea, polipnea, cianosis, odinofagia, anosmia, disgeusia, diarrea, dolor abdominal, cefalea, mialgia, artralgia, dolor torácico, fiebre, escalofríos e irritabilidad.

Tabla 6: Análisis de regresión logística binaria para predecir prueba individual de SARS-COV-2

Variable	Ola 1	Ola 2	Ola 3	Ola 4	Ola 5	General	Ola 6
Edad (años)	Eliminado	1.010 (1.007 – 1.013)	1.005 (1.002 – 1.008)	No incluido	1.022 (1.011 – 1.033)	No incluido	1.014(1.008 – 1.019)
Femenino	0.695 (0.542 – 0.891)	No incluido	0.749 (0.687 – 0.817)	No incluido	No incluido	0.818 (0.776 – 0.863)	No incluido
Fumador	No incluido	1.173 (1.017 – 1.353)	No incluido	No incluido	Eliminado	Eliminado	No incluido
Hipertensión	Eliminado	Eliminado	No incluido	Eliminado	Eliminado	1.184 (1.066 –	No incluido

						1.316)	
Obesidad	No incluido	1.475 (1.199 – 1.815)	Eliminado	Eliminado	Eliminado	Eliminado	No incluido
Diabetes mellitus	1.761 (1.199 – 2.585)	Eliminado	Eliminado	No incluido	Eliminado	1.312 (1.166 – 1.475)	Eliminado
Contacto con caso sospechoso	No incluido	0.718 (0.660 – 0.782)	Eliminado	Eliminado	Eliminado	Eliminado	1.315(1.077 – 1.606)
Tos	1.725 (1.280 – 2.324)	2.045 (1.853 – 2.256)	3.155 (2.875 – 3.464)	2.970 (2.479 – 3.558)	4.558 (2.882 – 7.209)	2.455 (2.311 – 2.608)	1.480(1.229 – 1.782)
Rinorrea	Eliminado	2.210 (1.951 – 2.503)	2.057 (1.799 – 2.352)	Eliminado	3.658 (2.211 – 6.052)	1.888 (1.747 – 2.041)	Eliminado
Disnea	1.548 (1.161 – 2.063)	0.806 (0.669 – 0.972)	Eliminado	No incluido	Eliminado	0.865 (0.767 – 0.976)	0.530(0.399 – 0.704)
Polipnea	Eliminado	0.565 (0.403 – 0.792)	Eliminado	No incluido	No incluido	0.727 (0.588 – 0.899)	No incluido
Cianosis	No incluido	Eliminado	0.529 (0.316 – 0.886)	No incluido	No incluido	0.649 (0.473 – 0.889)	No incluido
Odinofagia	No incluido	1.498 (1.328 – 1.689)	2.267 (1.987 – 2.586)	2.231 (1.841 – 2.704)	Eliminado	1.627 (1.508 – 1.756)	Eliminado
Anosmia	Eliminado	2.528 (2.034 – 3.142)	3.805 (2.825 – 5.124)	No incluido	No incluido	2.964 (2.525 – 3.478)	Eliminado
Disgeusia	2.259 (1.520 – 3.359)	1.413 (1.127 – 1.771)	1.709 (1.255 – 2.327)	No incluido	No incluido	1.663 (1.407 – 1.965)	Eliminado
Diarrea	No incluido	Eliminado	0.725 (0.616 – 0.854)	Eliminado	No incluido	0.726 (0.650 – 0.811)	Eliminado
Dolor abdominal	No incluido	Eliminado	0.777 (0.626 – 0.963)	Eliminado	Eliminado	0.826 (0.714 – 0.957)	No incluido

Vómito	No incluido	0.582 (0.413 – 0.820)	Eliminado	No incluido	No incluido	Eliminado	0.412 (0.249 – 0.682)
Cefalea	0.717 (0.528 – 0.966)	1.590 (1.434 – 1.762)	1.700 (1.531 – 1.887)	Eliminado	1.964 (1.283 – 3.007)	1.531 (1.433 – 1.636)	Eliminado
Mialgia	Eliminado	Eliminado	1.865 (1.570 – 2.215)	1.292 (1.014 – 1.647)	Eliminado	1.311 (1.185 – 1.449)	Eliminado
Artralgia	Eliminado	1.552 (1.334 – 1.806)	1.387 (1.144 – 1.682)	Eliminado	Eliminado	1.285 (1.151 – 1.434)	1.241(1.012 – 1.521)
Dolor torácico	Eliminado	1.271 (1.079 – 1.497)	Eliminado	Eliminado	Eliminado	1.144 (1.034 – 1.266)	Eliminado
Fiebre	2.060 (1.575 – 2.695)	1.864 (1.664 – 2.089)	2.978 (2.670 – 3.322)	1.631 (1.240 – 2.146)	3.654 (2.340 – 5.704)	2.393 (2.229 – 2.570)	Eliminado
Escalofríos	Eliminado	1.485 (1.275 – 1.729)	1.668 (1.454 – 1.913)	1.518 (1.176 – 1.959)	1.870 (1.134 – 3.083)	1.492 (1.365 – 1.632)	Eliminado
Irritabilidad	No incluido	0.760 (0.628 – 0.920)	0.802 (0.676 – 0.950)	2.005 (1.089 – 3.691)	Eliminado	0.857 (0.761 – 0.965)	Eliminado
Conjuntivitis	No incluido	1.309 (1.025 – 1.670)	Eliminado	No incluido	Eliminado	Eliminado	0.705(0.513 – 0.9

Tabla 6: No incluido= variable no incluida en la regresión, Eliminado: Variable eliminada por la regresión, %, los datos numéricos representan EXP(B) (inferior- superior) con IC 95%. La tabla muestra los análisis de regresión binaria para cada una de las olas y el modelo general, los datos se presentan como razón de momios con un intervalo de confianza al 95%. Elaboración propia.

La Tabla 7 muestra los resultados del desempeño de los modelos de la base completa. Los coeficientes de determinación (R^2) de los modelos variaron entre 0.049 a 0.537. Las predicciones globales también variaron de acuerdo con la ola, con una predicción acertada en promedio de 73% de eficacia (rango entre 63% y 83.1%), con sensibilidad promedio de 67% (rango entre 18 y 86%),

especificidad promedio de 74% (rango entre 66 y 91%), valor predictivo positivo promedio de 62% (rangos entre 34 y 74%) y valor predictivo negativo promedio de 77% (rango entre 58 y 91%).

Tabla 7. Resumen del desempeño de los modelos de predicción de prueba individual de SARS-Cov-2.

Variable	<i>Ola 1</i>	<i>Ola 2</i>	<i>Ola 3</i>	<i>Ola 4</i>	<i>Ola 5</i>	<i>General</i>	<i>Ola 6</i>
R²	0.151	0.267	0.405	0.201	0.537	0.313	0.049
Predicción global (%)	66.2%	71.1%	76.9%	76.6%	83.1%	73.2%	62.5%
Sensibilidad	65.8%	67.4%	77.4%	79.6%	86.4%	73.0%	18.1%
Especificidad	66.4%	74.4%	75.8%	59.3%	74.9%	73.2%	91.1%
Valor predictivo positivo	73.7%	71.0%	63.7%	34.1%	71.0%	60.8%	56.9%
Valor predictivo negativo	57.7%	71.1%	86.0%	91.7%	88.9%	82.8%	63.2%

Tabla 7: Resumen de las regresiones logísticas de las olas 1-6 y general, incluyendo R² y predicción global que incluye positivos y negativos. Elaboración propia.

7.2 Análisis de una muestra aleatoria de la base de datos

La Tabla 8 muestra las características de la muestra seleccionada de los participantes del estudio. Los símbolos de asterisco (*) muestran que la ola 1 tuvo la mayor edad, mayor porcentaje de hombres, más proporción de fumadores, hipertensión, obesidad, diabetes y contacto con caso sospechoso comparado con las olas subsecuentes (2 a la 6). Asimismo, la ola 2 tuvo mayor edad que las olas, mayor proporción de mujeres y de diabetes y contacto con caso sospechoso que las olas 3-5, mientras que las otras

variables fueron menores. Comparado con la ola 3 (símbolo #), tuvo menor edad, mayor cantidad de proporción de mujeres, de fumadores, de obesidad, y de contacto con caso sospechoso, mientras que la ola 4 (símbolo &) tuvo menor edad, mayor proporción de hombres, y de caso sospechoso que la ola 5. Mientras que la ola 5 (símbolo \$) tuvo menor cantidad de fumadores, y de contacto con caso sospechoso que las demás olas. La ola 6 (símbolo ≠) mostró tener diferencias estadísticamente significativas con todas las otras olas en todas las variables.

Tabla 8. Características socio-demográficas y morbilidades de la población de estudios de la muestra.

	Ola 1	Ola 2	Ola 3	Ola 4	Ola 5	Ola 6
Variable	(N = 377)	(N = 377)	(N = 377)	(N = 377)	(N = 377)	(N = 377)
Edad (años)	46	41 *	36 * ^{&}	38 *	40 * [#]	42 * ^{&#} ≠
KW	(34 – 56)	(30 – 52)	(27 – 48)	(29 – 50)	(28 – 55)	(30 – 53)
	*	*	*	*	*	*
Sexo						^{&#}
Femenino	191 (50.7%)	195 (51.7%)	218 (57.8%)	200 (53.1%)	203 (53.8%)	* ^{&#}
Masculino	186 (49.3%)	182 (48.3%)	159 (42.2%)	177 (46.9%)	174 (46.2%)	204 (59.1%)
						173 (44.9%)
Fumador	41 (10.9%)	29 (7.7%)	37 (9.8%)	27 (7.3%)	14 (3.8%) * ^{&#}	22 (5.9%) * ^{&#} ≠
Hipertensión	65 (17.3%)	36 (9.9%) *	28 (7.4%) *	25 (6.8%) *	39 (10.4%) *	32 (9.6%) * ^{&#} ≠
Obesidad	66 (17.6%)	19 (5.0%) *	28 (7.4%) *	6 (1.6%) * ^{&#}	15 (4.0%) *	23 (6.1%) * ^{&#} ≠
Diabetes mellitus	61 (16.3%)	32 (8.5%) *	20 (5.3%) *	17 (4.6%) * ^{&}	29 (7.8%) *	26 (6.8%) * ^{&#}

Contacto con caso sospechoso	167 (64.5%)	214 (56.9%)	148 (39.6%) ^{&} *	139 (38.4%) ^{&} *	93 (24.9%) ^{&#&\$} *	79 (20.9%) ^{&#&\$} * [≠]
-------------------------------------	-------------	-------------	--------------------------------	--------------------------------	---------------------------------------	--

Tabla 8: Características sociodemográficas y co-morbilidades de la población de estudios, muestra recuento y porcentajes o mediana (percentil 25 – percentil 75). U de Mann-Whitney y chi-cuadrada. * $p < 0.05$ vs ola 1, [&] $p < 0.01$ vs ola 2, [#] $p < 0.01$ vs ola 3, ^{\$} $p < 0.01$ vs ola 4, [≠] $p < 0.01$ vs ola 5. KW: Kruskal Wallis que muestra ser estadísticamente significativa en comparación con las otras olas. Elaboración propia

La Tabla 9 muestra los datos socio-demográficos y morbilidades de los participantes de la muestra comparados respecto al resultado de la prueba (positivo o negativo), los asteriscos indican diferencias estadísticamente significativas entre positivos y negativos dentro de una misma ola ($p < 0.01$). En la ola 1 las variables distintas entre positivos y negativos (indicadas con *) son: edad, sexo, hipertensión y diabetes. En la ola 5 las variables con diferencias estadísticamente significativas fueron hipertensión, diabetes y contacto con caso sospechoso. En la ola 6 se mostró diferencias entre edad, diabetes mellitus y contacto con caso sospechoso.

Tabla 9. Características socio-demográficas y morbilidades de la población de estudios entre positivos y negativos de la muestra

Variable	Ola 1		Ola 2		Ola 3		Ola 4		Ola 5		Ola 6	
	Positivo (N = 213)	Negativo (N = 164)	Positivo (N = 195)	Negativo (N = 182)	Positivo (N = 151)	Negativo (N = 226)	Positivo (N = 115)	Negativo (N = 262)	Positivo (N = 131)	Negativo (N = 246)	Positivo (N = 94)	Negativo (N = 283)
Edad (años)	47 (37 – 59)	43 * (33 – 55)	41 (31 – 54)	41 (28 – 50)	35 (27 – 47)	36 (27 – 49)	37 (28 – 51)	38 (29 – 50)	44 (27 – 64)	39 (28 – 53)	44 (28-56)	39* (30-53)
Sexo	*											
Femenino	44.1%	59.1%	55.1%	47.4%	53.7%	60.3%	44.6%	56.5%	51.3%	55.3%	59.1%	58.9%
Masculino	55.9%	40.9%	44.9%	52.6%	46.3%	39.7%	55.4%	43.5%	48.7%	44.7%	42.1%	43.0%
Fumador	11.4%	10.4%	7.6%	7.5%	12.1%	8.5%	8.4%	7.4%	5.5%	2.6%	5.6%	6.4%

Hipertensión	22.4%	11.0%*	11.4%	7.5%	7.4%	6.7%	7.2%	6.1%	20.0%	3.9%*	9.2%	8.3%
Obesidad	18.1%	16.0%	7.0%	2.9%	7.4%	7.6%	1.2%	2.2%	7.3%	2.6%	6.0%	6.5%
Diabetes mellitus	21.4%	9.8%*	10.3%	6.4%	4.7%	5.4%	6.0%	3.9%	14.5%	3.9%*	5.6%	7.0%*
Contacto con caso sospechoso	58.1%	70.3%	54.9%	59.5%	43.9%	36.6%	39.8%	38.0%	38.7%	19.5%*	26.2%	20.6%*

Tabla 9: Características sociodemográficas y co-morbilidades de la población de estudios, muestra recuento y porcentajes o mediana (percentil 25 – percentil 75). U de Mann-Whitney y chi-cuadrada. * $p < 0.05$ comparado con positivos de la misma ola. Elaboración propia

En la Tabla 10 se compara el porcentaje de presencia de síntomas en los pacientes con prueba positiva VS negativa dentro de cada ola y en general de la muestra. Se marcan con asteriscos y negritas las prevalencias con significancia entre positivos y negativos.

Tabla 10: Prevalencia de los síntomas entre positivos y negativos de la muestra.

Variable	Ola 1		Ola 2		Ola 3		Ola 4		Ola 5		General		Ola 6	
	Positivo (N = 213)	Negativo (N = 164)	Positivo (N = 195)	Negativo (N = 182)	Positivo (N = 151)	Negativo (N = 226)	Positivo (N = 115)	Negativo (N = 262)	Positivo (N = 131)	Negativo (N = 246)	Positivo (N = 805)	Negativo (N = 1080)	Positivo (N = 94)	Negativo (N = 283)
Tos	83.4	65.2*	48.1	19.1*	67.6	23.7*	48.2	16.1*	78.6	17.5*	66.7	26.5*	62.1	53.1*
Rinorrea	39.6	35.0	24.3	8.1*	34.2	5.4*	31.3	13.5*	41.1	6.1*	34.0	12.5*	40.0	37.6*
Disnea	52.1	25.0*	11.9	7.5	12.1	4.9*	3.6	3.0	17.9	4.4*	23.4	8.0*	10.8	14.4*
Polipnea	16.4	13.1	4.3	1.7	0.7	0.9	0.0	0.0	3.6	0.4	6.4	2.7*	4.3	5.4
Cianosis	7.7	3.1	2.7	0.6	0.0	0.0	0.0	0.0	0.9	0.4	3.0	0.7*	1.2	3.1
Odinofagia	50.7	48.8	27.6	11.0*	28.2	7.1*	39.8	14.3*	36.0	8.8*	36.9	16.4*	43.0	41.2*
Anosmia	21.9	11.3*	21.5	5.9*	17.2	1.8*	1.3	1.4	1.9	0.5	14.7	3.2*	5.2	2.3*
Disgeusia	21.1	7.5*	17.9	5.3*	15.2	2.2*	0.0	0.9	1.9	0.9	12.8	2.8*	3.1	2.7*

Diarrea	23.3	24.4	12.4	3.5 *	8.7	4.9	7.2	2.2	13.4	5.3	14.3	7.3 *	12.6	9.3*
Dolor abdominal	13.5	14.4	4.3	2.3	4.0	4.0	3.6	2.6	21.4	3.9 *	9.4	5.0 *	5.4	6.2
Vómito	8.7	9.4	3.8	2.9	2.7	1.8	1.2	0.9	2.7	2.2	4.5	3.1	3.6	5.9*
Dolor de cabeza	75.4	74.4	45.4	20.8 *	45.6	14.7 *	31.3	14.3 *	65.2	17.5 *	55.4	25.9 *	51.4	49.6*
Mialgia	58.9	48.1 *	21.1	7.5 *	24.8	4.0 *	30.1	7.0 *	41.1	12.7 *	36.5	14.2 *	35.1	30.0*
Artralgia	53.1	39.1 *	16.8	6.4 *	18.8	1.8 *	14.5	4.3 *	28.6	5.3 *	28.9	9.9 *	25.9	21.6*
Dolor torácico	35.3	27.5	11.4	5.8	12.8	8.5	12.0	2.2 *	21.4	7.9 *	20.0	9.5 *	15.9	16.1*
Fiebre	73.0	45.1 *	30.3	12.1 *	38.3	14.7 *	13.3	5.2 *	52.7	9.6 *	45.5	15.9 *	37.2	34.3*
Escalofríos	46.9	33.8 *	22.2	5.2 *	20.8	9.4 *	20.5	5.2 *	37.5	5.3 *	31.0	10.6 *	29.3	26.6*
Irritabilidad	28.6	15.2 *	7.6	6.4	10.1	7.1	1.2	0.4	11.6	6.1	13.9	6.6 *	13.8	11.8*
Conjuntivitis	10.6	12.5	5.9	2.9	8.1	2.7 *	2.4	2.2	14.3	2.2	8.6	4.0 *	8.2	12.8*

Tabla 10: Prevalencia de los síntomas (positivos y negativos de cada ola y general) mostrada en porcentajes. Chi-cuadrada. * $p < 0.05$ comparando con positivos de su grupo. Elaboración propia

La Tabla 11 presenta los resultados del análisis de regresión logística con modelos para predecir el resultado de la prueba de SARS-CoV-2 (positivo o negativo) como variable dependiente. En cada ola se calculó un modelo distinto que incluyó como variables independientes aquellas que tuvieron diferencias significativas entre positivos y negativos (Tablas 9 y 10). Se consideraron las variables que no se incluyeron inicialmente en el modelo indicadas como “No incluido”, así como se eliminaron las variables que no tuvieron contribución significativa a la variable dependiente, indicadas como “Eliminado”. Se consideró la razón de momios significativa ($p < 0.05$). Estos criterios fueron aplicados a cada uno de los modelos de las olas 1-6 y en la general.

Tabla 11: Análisis de regresión logística binaria para predecir prueba individual de SARS-COV-2 de la muestra.

Variable	Ola 1	Ola 2	Ola 3	Ola 4	Ola 5	General	Ola 6
Edad (años)	Eliminado	No incluido	No incluido	No incluido	No incluido	1.012 (1.005 – 1.020)	1.010(1.006 – 1.015)

Femenino	Eliminado	No incluido	No incluido	No incluido	No incluido	Eliminado	No incluido
Fumador	No incluido	No incluido	No incluido	No incluido	No incluido	No incluido	No incluido
Hipertensión	2.454 (1.066 – 5.647)	No incluido	No incluido	No incluido	No incluido	Eliminado	No incluido
Obesidad	No incluido	No incluido	No incluido	No incluido	No incluido	Eliminado	No incluido
Diabetes mellitus	Eliminado	No incluido	No incluido	No incluido	Eliminado	Eliminado	1.1012(1.013 – 1.203)
Contacto con caso sospechoso	No incluido	No incluido	No incluido	No incluido	0.352 (0.158 – 0.785)	Eliminado	1.300(1.067 – 1.506)
Tos	Eliminado	2.687 (1.605 – 4.499)	4.898 (2.866 – 8.371)	2.996 (1.605 – 5.593)	7.641 (3.524 – 16.568)	2.875 (2.175 – 3.800)	1.480(1.229 – 1.782)
Rinorrea	No incluido	2.230 (1.105 – 4.501)	5.993 (2.749 – 13.064)	Eliminado	4.105 (1.863 – 9.043)	2.013 (1.468 – 2.760)	Eliminado
Disnea	2.162 (1.104 – 4.231)	No incluido	Eliminado	No incluido	Eliminado	Eliminado	Eliminado
Polipnea	No incluido	No incluido	No incluido	No incluido	No incluido	Eliminado	No incluido
Cianosis	No incluido	No incluido	No incluido	No incluido	No incluido	Eliminado	No incluido
Odinofagia	No incluido	Eliminado	Eliminado	2.115 (1.092 – 4.097)	Eliminado	Eliminado	1.030(1.011 – 1.058)
Anosmia	Eliminado	3.296 (1.513 – 7.179)	5.962 (1.729 – 20.562)	No incluido	No incluido	3.986 (2.400 – 6.620)	Eliminado
Disgeusia	2.833 (1.106 – 7.257)	Eliminado	Eliminado	No incluido	No incluido	Eliminado	Eliminado
Diarrea	No incluido	Eliminado	No incluido	No incluido	No incluido	Eliminado	Eliminado
Dolor abdominal	No incluido	No incluido	No incluido	No incluido	Eliminado	Eliminado	No incluido
Vómito	No incluido	No incluido	No incluido	No incluido	No incluido	No incluido	Eliminado
Cefalea	No incluido	Eliminado	2.051 (1.121 – 3.750)	Eliminado	2.773 (1.349 – 5.699)	1.393 (1.050 – 1.849)	1.591(1.359 – 1.690)

Mialgia	Eliminado	Eliminado	No incluido	2.358 (1.052 – 5.283)	Eliminado	Eliminado	Eliminado
Artralgia	Eliminado	Eliminado	3.618 (0.962 – 13.754)	Eliminado	Eliminado	Eliminado	1.241(1.012 – 1.521)
Dolor torácico	No incluido	No incluido	No incluido	Eliminado	Eliminado	Eliminado	Eliminado
Fiebre	3.008 (1.656 – 5.461)	Eliminado	Eliminado	Eliminado	3.944 (1.898 – 8.198)	1.915 (1.406 – 2.608)	1.136(0.953– 1.212)
Escalofríos	Eliminado	3.167 (1.352 – 7.417)	Eliminado	Eliminado	Eliminado	1.706 (1.187 – 2.452)	1.509(1.254– 1.704)
Irritabilidad	Eliminado	No incluido	No incluido	No incluido	No incluido	Eliminado	Eliminado
Conjuntivitis	No incluido	No incluido	Eliminado	No incluido	Eliminado	Eliminado	Eliminado

Tabla 11: Regresión logística de cada una de las olas y general. No incluido= variable no incluida en la regresión, Eliminado: Variable eliminada por la regresión. Se muestran las regresiones de cada una de las olas y en general, se presentan como razón de momios con un intervalo de confianza al 95%. Elaboración propia

La Tabla 12 muestra los resultados del desempeño de los modelos de la muestra. Los coeficientes de determinación (R^2) de los modelos variaron entre 0.045 a 0.544. Las predicciones globales también variaron de acuerdo con la ola, con una predicción acertada en promedio de 72.1% de eficacia (rango entre 61.9 y 82.0%), con sensibilidad promedio de 62.6% (rango entre 18.6 y 79.60%), especificidad promedio de 76.9% (rango entre 67.05 y 90.2%), valor predictivo positivo promedio de 62.45% (rangos entre 33.7 y 75.4%) y valor predictivo negativo promedio de 79.9% (rango entre 54.3 y 92.2%).

Tabla 12. Resumen del desempeño de los modelos de predicción de prueba individual de SARS-Cov-2 de la muestra.

Variable	Ola 1	Ola 2	Ola 3	Ola 4	Ola 5	General	Ola 6
R²	0.227	0.215	0.430	0.194	0.544	0.297	0.045
Predicción global (%)	65.3%	70.2%	75.9%	76.7%	82.0%	73.1%	61.9%
Sensibilidad	64.17	71.92	76.92	60.86	76.60	69.00	18.6%
Especificidad	67.05	67.40	75.47	79.40	83.95	74.97	90.2%
Valor predictivo positivo	75.4	68.7	55.6	33.7	63.9	56.1	57.3%
Valor predictivo negativo	54.3%	71.8	89.1	92.2	90.7	83.9	61.9%

Tabla 12: Resumen de las regresiones logísticas de las olas 1-5 y general, incluyendo R² y predicción global que incluye positivos y negativos. Elaboración propia.

7.3 Fase 2: Verificación de los modelos de predicción

La Tabla 13 muestra los resultados para prueba positiva por COVID-19 como resultado de las predicciones de los modelos desarrollados. El modelo de la ola 5 aplicado a los datos de la base de datos completa de la ola 6 se obtuvo un 100% de porcentaje de predicción correcta positiva, 0% en predicción correcta negativa, y un porcentaje global de acierto de 39%, el modelo de la ola general obtuvo un 66% de porcentaje de predicción correcta positiva, 43% en predicción correcta negativa, y un porcentaje global de acierto de 52%, el modelo de la ola 6 aplicado a su misma base de datos obtuvo un 92% de porcentaje de predicción correcta positiva, 18% en predicción correcta negativa, y un porcentaje global de acierto de 64% mientras que el modelo de la ola 6 aplicado a su misma base de datos obtuvo un 92% de porcentaje de predicción correcta positiva, 18% en predicción correcta negativa, y un

porcentaje global de acierto de 64%. Se tomó en cuenta el valor de corte establecido para la fórmula de la regresión logística binaria en >0.5 como predicción positiva. Se aplicaron los mismos modelos de la misma forma en los datos de la muestra, y se obtuvieron los mismos resultados para todas las olas y modelos.

Tabla 13: Valores dados por la fórmula de la regresión logística binaria en cada uno de los modelos desarrollados.

<i>Modelo</i>		Resultado final de la prueba PCR por modelo															
		Base de datos completa	VN	FN	FP	VP	Porcentaje correcto negativo	Porcentaje correcto positivo	Porcentaje correcto global	Sensibilidad	Especificidad	Valor predictivo positivo	Valor predictivo negativo	N COMPLETOS	N CLASIFICADOS	TOTAL NEGATIVOS	TOTAL POSITIVOS
Desarrollo	Ola 1	475	251	399	868	54%	78%	67%	78%	54%	69%	65%	1993	1993	874	1119	
	Ola 2	3799	1860	1564	4512	71%	71%	71%	71%	71%	74%	67%	11735	11735	5363	6372	
	Ola 3	6567	1966	1091	3469	86%	64%	77%	64%	86%	76%	77%	13093	13093	7658	5435	
	Ola 4	2959	757	270	393	92%	34%	77%	34%	92%	59%	80%	4379	4379	3229	1150	
	Ola 5	542	83	68	203	88%	68.12%	84%	87%	88%	75%	89%	914	896	616	298	
	General	14398	5473	3343	8901	81%	62%	73%	62%	81%	73%	72%	32115	32115	17741	14374	
Validación en ola 6	Ola 5	0	0	1532	961	0%	100%	39%	100%	0%	39%	0%	2493	961	1532	961	
	General	652	328	880	633	43%	66%	52%	66%	43%	42%	67%	2493	1285	1532	961	
	Ola 6	1410	787	122	174	92%	18%	64%	18%	92%	59%	64%	2493	1584	1532	961	
Muestra																	
Desarrollo	Ola 1	82	53	82	158	50%	75%	64%	75%	50%	66%	61%	375	375	164	211	
	Ola 2	124	58	49	127	72%	69%	70%	69%	72%	72%	68%	358	358	173	185	
	Ola 3	200	69	24	80	89%	54%	75%	54%	89%	77%	74%	373	373	224	149	
	Ola 4	212	55	18	28	92%	34%	77%	34%	92%	61%	79%	313	313	230	83	
	Ola 5	206	43	22	70	90%	62%	81%	62%	90%	76%	83%	341	341	228	113	
	General	845	373	174	367	83%	50%	69%	50%	83%	68%	69%	1760	1759	1019	741	
Validación en	Ola 5	0	0	283	375	0%	100%	39%	100%	0%	39%	0%	377	375	283	94	

General	43	2	58	42	43%	66%	52%	66%	43%	42%	67%	377	375	283	94
Ola 6	93	53	79	11	92%	18%	64%	18%	92%	59%	64%	377	375	283	94

Tabla 13. Cálculo de los valores de p dados por la fórmula de la regresión logística binaria, mostrados en recuento y porcentajes. Siglas: VN= verdaderos negativos, FN= falsos negativos, FP= falsos positivos, VP= verdaderos positivos. Elaboración propia

VIII. Discusión

El propósito principal de este estudio fue desarrollar un modelo de predicción estadístico para resultado de prueba por COVID-19, que no necesitara pruebas de laboratorio y tomara menos tiempo de aplicación y de resultados, así como que tuviera un porcentaje de acierto de predicción mayor al 70%, el cual se logró de manera individual en los modelos de las distintas olas, sin embargo, bajo drásticamente su efectividad en la fase de verificación.

La principal contribución del presente reporte es demostrar que es posible el desarrollo de modelos estadísticos para predicción individual de la infección por el virus SARS-CoV2 con predicción global mayor al 80% (con la ola 5) y en general mayor al 50% con las demás olas, esto a partir de datos socio-demográficos y los síntomas que usualmente se recaban en el sistema de vigilancia epidemiológica de México, los cuáles son confiables y tienen bajo costo y pueden ser de fácil acceso a la población mexicana en general, sobre todo al sector más vulnerable. Como población de prueba, el reporte se enfocó en una muestra por censo de todos los casos registrados en la Alcaldía Álvaro Obregón, una de las más pobladas y con gran heterogeneidad socioeconómica, en una de las Ciudades más pobladas del mundo, la Ciudad de México.

8.1 Desarrollo del diseño del modelo de predicción

El modelo de predicción para prueba positiva de COVID-19 presenta similitudes y mejoras con respecto a otros modelos parecidos que se han desarrollado a partir del surgimiento de la pandemia en 2020 mencionados a continuación. Las variables contempladas en los modelos se determinaron revisando la literatura acerca de la sintomatología y las características generales que presentaba el virus a la fecha en la que se tenía acceso en el momento del desarrollo de cada uno de los modelos. Los modelos más antiguos que son los del año de 2020 presentaron como variables los datos de la primera ola de contagios, al inicio de la pandemia no se contaba con información suficiente sobre la enfermedad, tampoco se identificaban las características de los pacientes que la agravaban o aminoraban el problema. Un ejemplo de ello es Roland (34) en 2020 en California, el cual incorporó los síntomas proporcionados a través de una encuesta digital, al igual que en este reporte, con las cuales determinó los diez síntomas más frecuentes, información demográfica y comorbilidades, a través de un modelo de regresión, las variables utilizadas aquí varían con

respecto a la población. Otro estudio en Cleveland (35) incorporó nueva información y de igual manera, a través de la regresión logística binaria y determinó los factores de riesgo para ser positivos, así mismo desarrollaron una validación externa, en nuestro caso fuimos un paso más adelante y estos factores de riesgo los utilizamos para desarrollar modelos de predicción de prueba para COVID-19. Otras investigaciones determinaron factores de riesgo, pero también factores protectores (36), la mayoría de modelos parecidos o relacionados aplicaron encuestas electrónicas o hicieron uso de bases de datos de sus síntomas de salud, aplicando distintos tipos de regresiones logísticas y obteniendo valores predictivos globales, positivos, negativos, sensibilidad, especificidad o precisión tal y como se aplicaron en este reporte.

El modelo diseñado en el presente estudio determinó la sintomatología más reportada para la enfermedad de COVID-19, para el desarrollo de este reporte, ya se contaba con información acerca de las distintas variantes del virus, por lo que se incluyó información de todas ellas, ya que las bases de datos utilizadas no contenían información del tipo de variante con la cual el paciente dio positivo para SARS-CoV-2 (original, alpha, Beta, Gamma, Delta, Ómicron y sublinajes de Ómicron (24)), las cuales fueron: fiebre, tos, odinofagia, disnea, irritabilidad, diarrea, dolor torácico, escalofríos, cefalea, mialgia, artralgia, rinorrea, polipnea, vómito, dolor abdominal, conjuntivitis, cianosis, anosmia y disgeusia. Las características socio-demográficas de los pacientes que pudieran ser variables predictoras de acuerdo a la literatura: sexo, edad, tabaquismo y contacto con caso conocido, y las comorbilidades: diabetes, hipertensión y obesidad. Como control se manejó la variante con el número correspondiente a cada ola, que incluye de la primera a la sexta ola para las dos fases desarrolladas en este reporte, y la variable dependiente fue el resultado de la prueba de PCR para COVID-19. Al obtener la lista de estas variables, se cotejaron con las variables disponibles en las bases de datos del SINAVE, para determinar cuáles debían de ser incluidas y excluidas. Dentro de las comorbilidades se determinó trabajar con las 3 principales que afectan a la población mexicana en la que se realizó el estudio.

En este reporte y en la literatura revisada, las variables determinantes coincidieron con las identificadas previamente en otras poblaciones (1-4,6,7,17). Sin embargo, cabe destacar que no se habían diseñado modelos que incluyeran las cinco olas en forma

individual y en general, por lo que hasta la fecha no hay una comparación plena con otro modelo de predicción.

La regresión logística binaria fue determinada para usar en este modelo por su amplio uso en la epidemiología, y por ser aplicable a las variables determinadas, ya que son dicotómicas y cualitativas con excepción de la edad. Este método de análisis tiene la característica de combinar las variables predictoras con mayor probabilidad para predecir el resultado, que corresponda a la prueba de PCR positiva. El modelo muestra en distintos pasos las diferentes combinaciones de variables, reportando finalmente el paso con mayor porcentaje de predicción, y las variables que lo conforman, junto con sus características del análisis.

En cada una de las olas se utilizaron distintas variables para realizar el modelo de predicción final, que fueron determinadas previamente de las diferencias significativas entre los casos positivos y negativos de cada una de las olas. Como ya se mencionó, en cada uno de los picos de cada ola se presentaron distintas variantes del SARS-CoV-2, por lo que las variables predictoras cambiaron dependiendo de estas características.

Respecto a los modelos de las muestras, en la Tabla 11 se observa que en general cada regresión tuvo menos variables que en los obtenidos de las bases de datos completas, ya que la mayoría no fueron estadísticamente significativas. Por ejemplo, las variables predictoras para cada una de las olas y que fueron aplicadas en los modelos cambiaron, no en todas se obtuvieron los mismos resultados de significancia, por lo que no todas fueron incluidas, y dichas variables cambiaron en su mayoría de acuerdo a la información obtenida por la literatura. Por lo cuál queda demostrada la importancia del análisis de todas las variables y posteriormente la elección de las más adecuadas en cada uno de los casos. La tos, el dolor de cabeza y el contacto con caso positivo fueron las variables predictoras más comunes.

En la Tabla 12 se muestra que la R^2 y la predicción global más altas se presentaron en el modelo de la ola 5 con: 0.544, 82.0% respectivamente, la sensibilidad más alta se obtuvo en la ola 3 con 76.92%, la especificidad más alta se obtuvo en la ola 5 con 83.95%, mientras que al igual que en los anteriores modelos, el mayor valor predictivo positivo, se obtuvo en la ola 1 con 75.4% y el mayor valor predictivo negativo se obtuvo en la ola 4 con

92.2%. En general podemos notar que las variables más repetidas en los modelos fueron: tos, fiebre y cefalea, mientras que los otros no se incluyen en todos los modelos.

Después de obtener los modelos a través de la regresión logística, se obtuvieron los valores incluidos en la ecuación de la regresión logística binaria para obtener el valor de p o de la probabilidad de que un caso pertenezca a una categoría determinada, en este caso a los positivos de prueba PCR para COVID-19. Con cada uno de los modelos se obtuvieron los valores de β y la constante, estos valores fueron sustituidos en la ecuación para obtener los porcentajes de predicción, esto al aplicarlos en las bases de datos con los que fueron desarrollados los modelos, para cada una de las olas y para general. Para la ola 5 se obtuvo el mayor porcentaje global de aciertos de predicción de prueba individual positiva con 84%.

Para la fase de validación se aplicó la fórmula correspondiente a cada uno de los modelos de predicción en Excel para validar los resultados obtenidos con SPSS, los resultados mostraron una ligera mejora aplicando la fórmula de la regresión logística binaria. Los mejores resultados para los modelos fueron los de la ola 5, por lo que esta fórmula fue aplicada a los datos de la ola 6, junto con el del modelo general, mostrando excelentes resultados para el modelo de la ola 5, con más del 80% de predicción global correcta.

8.2 Confiabilidad del modelo de predicción.

Se desarrollaron modelos previos de predicción de pruebas positivas para COVID-19, la mayoría en 2020, ya que al inicio de la pandemia las pruebas de laboratorio para COVID-19 eran muy pocas en comparación con el gran número de personas contagiadas todos los días. La mayoría de modelos de predicción que usaron bases de datos y análisis para predecir los resultados de las pruebas solo tomaron en cuenta los datos que se tenían hasta la primera ola de contagios, ya que el virus era nuevo y no se tenía mucha información. Con el tiempo esto ha ido cambiando, pues se han desarrollado nuevas olas de contagio con nuevas variantes del virus, por lo que las investigaciones y, por lo tanto, la información continúan actualizándose día a día.

Con la urgencia por las pruebas disminuida gracias a las vacunas, se han desarrollado nuevos métodos de detección de COVID-19 además de las ya mencionadas,

como los que usan biomarcadores y aparatos que la gente usa día a día, como los Smartwatch, las pulseras de actividad y los anillos inteligentes, entre otros (41).

Uno de los modelos de predicción que aplica el análisis y las bases de datos, además de regresión logística por etapas es el de Roland en 2020, que aplicó un modelo basado en la sintomatología más frecuente conocida hasta ese entonces, con los 10 síntomas más reportados por los pacientes de una encuesta electrónica en California, en el que se obtuvo un 82% de precisión y una sensibilidad del 70%, además de una predicción global correcta del 75%, sin embargo, este modelo no toma en cuenta las variantes ni las olas, tampoco hace una fase de verificación como en el caso de este reporte, además de que el modelo con mejor predicción global alcanzó en 84%, por lo que a pesar de sus buenos resultados, la falta de información podría bajar su efectividad aplicada a una ola diferente.

Un reporte realizado en Irán de Zhang en 2020 estudió de factores de riesgo en la población, refiere que las personas que hacen ejercicio tienen 78% de porcentaje de probabilidad de ser negativos, mientras que las personas con problemas crónicos de salud, aumentan sus probabilidades de ser positivos hasta en 48%. A pesar de que los factores de riesgo son importantes, se debe aumentar el número de variables de acuerdo a la población de estudio, para complementar los modelos lo más posible, por ejemplo, los modelos aquí presentados, además de determinar el porcentaje de riesgo de cada una de las variables, también hace una combinación de ellas para mejorar las probabilidades de acierto de positividad en la prueba.

Otros tipos de modelos se han desarrollado más recientemente, haciendo uso de los nuevos conocimientos y de las tecnologías, como los que aplican la IA, que a pesar de tener un excelente desempeño, es de difícil acceso y con un costo muy alto, mayor incluso al actual de las pruebas de laboratorio. Actualmente, existen varios modelos de predicción que utilizan distintos tipos de biosensores, si bien muestran buenos resultados, estos son menores a los de las pruebas de laboratorio, con materiales de difícil acceso y procesos más largos para los pacientes, por lo que se vuelven poco adecuados para la población en general. Los dispositivos móviles, como los Smartwatch se han vuelto sumamente útiles en el cuidado de la salud de las personas, y los métodos de predicción de COVID-19 que aplican tienen altos porcentajes de efectividad con aproximadamente 63% de eficacia, sin embargo, al ser de alto costo, solo una pequeña parte de la población mexicana puede

permitirse adquirir uno a varios de estos aparatos. El modelo de predicción a través del uso de radiografías reporta hasta un 98% de acierto, pero el uso de los equipos de radiografía es costoso y tarda hasta varios días en dar el resultado dependiendo del lugar donde se realice, por lo que es poco práctico tomando en cuenta el tiempo y el costo, además de que se tiene conocimiento de pacientes que no presentan problemas pulmonares, lo cual podría ser un problema similar el método que emplea los sonidos de la tos y la respiración, ya que muchos pacientes pueden no reportar estos síntomas, sobre todo dependiendo de la ola en la que se encuentran, puesto que no todas las variantes del virus presentaron este síntoma problemas pulmonares (modelo de radiografías) como factor de riesgo de acuerdo con los modelos desarrollados en este reporte.

Al revisar los modelos ya mencionados se pueden encontrar ventajas sobre ellos, la principal es la fase de verificación y sus excelentes resultados, otras como incluir la mayoría de variables reportadas hasta ahora, el hacer la distinción entre las olas por las variantes del virus y de la pandemia en general, el personalizar algunos factores a la población, en este caso las comorbilidades más frecuentes de la población a la que se aplican los modelos (mexicana), el fácil acceso y bajo costo, además de los resultados casi inmediatos. Finalmente, los resultados obtenidos en general hacen a estos modelos una opción confiable para la predicción de prueba positiva por COVID-19.

Tabla 14: Cuadro comparativo de modelos de predicción de COVID-19

<i>Método de predicción</i>	<i>Lugar de estudio</i>	<i>Método de análisis utilizado</i>	<i>Datos utilizados</i>	<i>Resultados</i>	<i>Fase de verificación</i>	<i>Costos</i>
Modelos de predicción para pacientes COVID-19 en Álvaro Obregón	Ciudad de México, Delegación Álvaro Obregón	Regresión logística	Bases de datos	Hasta 86% de capacidad de predicción	Resultado general del 63% de aciertos de predicción	Mínimo
Inteligencia artificial (41)	Estados Unidos y otros	Inteligencia Artificial	Bases de datos, señales biológicas	99% de capacidad de predicción	No reportada	Elevado
Identificación basada en biosensores (41)	Estados Unidos y otros	Biosensores	Bioseñales	Prometedores	No reportada	Elevado
Dispositivos electrónicos (41)	Estados Unidos y otros	Análisis estadísticos	Bases de datos, señales biológicas	98% de capacidad de predicción	No reportada	Medio
Método por radiografías (41)	Estados Unidos y otros	Imagenología y análisis de laboratorio	Imágenes y datos biológicos	98% de sensibilidad	No reportada	Elevado

Modelo basado en gusto y olfato (34)	Estados Unidos	Regresión logística	Encuestas electrónicas	82% de capacidad de predicción	No reportada	Mínimo
Predicción individualizada de prueba (35)	Estados Unidos	Regresión logística	Base de datos	95% de índice de confianza	No reportada	Mínimo
Predicción de probabilidad de COVID (36)	Irán	Regresión logística	Encuestas	95% de índice de confianza	Prometedora	Mínimo
Modelo de predicción individual (39)	Estados Unidos	Calculadora longitudinal de riesgos	Encuestas electrónicas	Predictores más comunes	No reportada	Mínimo
Predicción de casos graves y positivos (43)	Estados Unidos	Regresión logística multivariable	Registros de pacientes	90% de sensibilidad	No reportada	Mínimo

Elaboración propia.

Los resultados de esta investigación demuestran que es posible crear nuevos métodos de predicción de pruebas por COVID-19, considerando que los contagios aumentarán con nuevas olas en el futuro. El virus sigue cambiando y hoy más que nunca es de vital importancia que la población tenga acceso a métodos alternos a los de laboratorio, que también sean más rápidos. Las implicaciones epidemiológicas surgidas a partir de este reporte de investigación son la posible disminución de contagios en la población mexicana, al brindar nuevas herramientas para cuidar de su salud, de manera más rápida y barata, así como que abren las puertas a que se sigan desarrollando e investigando este tipo de métodos aplicados a la población mexicana. Como se ha mencionado, cada población tiene sus características específicas y con ello cambian las variables consideradas. Dentro de las consideraciones clínicas de los resultados obtenidos podemos apreciar el posible alivio de los servicios médicos que aún siguen prestando servicios a los pacientes COVID-19, y que a pesar de que los esquemas de vacunación han ayudado a la población a enfrentar la pandemia, no se puede ignorar el hecho de que el virus del SARS-CoV-2 llegó para quedarse como enfermedad estacionaria (50).

Los resultados reportados en este estudio coinciden con los reportados en la literatura, llegando a alcanzar hasta un 80% de capacidad de predicción por modelo individual, sin embargo, al no haber estudios parecidos con respecto a la aplicación de modelos de otra ola a nuevos datos, en este caso a la ola 6, no es posible realizar una comparación adecuada. Sin embargo, a raíz de los resultados de la fase 2, se pueden inferir

los cambios o ajustes que serían necesarios para obtener mejores resultados, esto tomando en cuenta las distintas características de las poblaciones y adecuando como tal los modelos.

8.3 Limitaciones

Aunque se encontraron diversas variables que no podían ser controladas debido al gran número de contagios, las variables del virus, el desconocimiento de la variable con la que están contagiados los pacientes, los esquemas de vacunación, movilidad de los pacientes, y a las características diversas de la población de México, en particular de la Ciudad de México, los modelos pudieron ser aplicados de manera correcta con los datos disponibles.

Otra de las limitaciones de este estudio es que no se ajustaron los valores de p para comparaciones múltiples, sin embargo, el valor de p que se consideró fue estricto a 0.5.

IX. Conclusiones

En el presente estudio se desarrollaron 14 modelos de predicción de prueba individual en una muestra de la Ciudad de México que demostraron ser una alternativa confiable a las pruebas de laboratorio para COVID-19, que se caracterizan por tener un buen porcentaje de acierto global de prueba positiva para COVID-19, ya que el modelo con mejor capacidad de predicción tuvo un 83.1%, correspondiente al modelo de la ola 5 de la base de datos completa, con una sensibilidad del 86.4% y una especificidad del 74.9%, además de que los otros modelos en general mostraron resultados de predicción global de 66.2 a 76.9% para la base de datos completa, y del 65.3 al 82.0% para la base de datos que incluía solo la muestra.

En la fase de verificación con la ola 6 se obtuvo un porcentaje de predicción global del 100% (aplicando el corte de acierto en >0.5 para los resultados de la fórmula), indicando que todos los resultados mostraron valores de p superiores a este valor, por lo que se concluye que la verificación obtuvo excelentes resultados para el modelo de la ola 5 y el general de ambas bases de datos.

Los factores de riesgo y las variables cambiaron para cada uno de los modelos debido al momento de la pandemia en el que se encontraban, principalmente por la variante predominante de contagio que se encontraba en México en ese momento, ya que como se

menciona en la literatura, cada una de ellas tiene distintas características, además, se consideran las jornadas de vacunación que disminuyeron el número de hospitalizaciones y casos graves.

Los modelos aquí desarrollados podrían ser de gran utilidad para el cuidado de la salud de la población mexicana, ya que a través de ellos, las personas con acceso a los modelos pueden obtener una predicción del resultado de su prueba por COVID-19, de igual manera el personal médico puede hacer uso de estas predicciones para tomar acciones para con los pacientes mientras se obtienen los resultados de las pruebas de laboratorio de ser el caso. Basados en metodología de bajo costo y con datos del SINAVE, además mostraron muy buen desempeño, por lo que son una nueva herramienta para afrontar nuevas olas de contagio, promoviendo un diagnóstico de riesgo temprano y con ello favorecer medidas sanitarias que disminuyan los contagios.

Dentro de las consideraciones clínicas de los resultados obtenidos podemos apreciar el posible alivio de los servicios médicos que aún siguen prestando servicios a los pacientes con COVID, y que a pesar de que los esquemas de vacunación han ayudado a la población a enfrentar la pandemia, no se puede ignorar el hecho de que el virus del SARS-CoV-2 llegó para quedarse como enfermedad estacionaria.

X. Perspectivas

La perspectiva más importante para los modelos de acuerdo a los resultados es la necesidad de evaluar los datos de la ola 6 en adelante de manera más detallada, con el fin de ajustar los modelos a sus características, ya que se observa que son distintos a los de las otras olas, principalmente por la variable de la vacunación y el acceso más fácil que se tiene de las pruebas por COVID-19. Podemos observar que la mayoría de pacientes presentan un resultado negativo en su prueba PCR, pero los porcentajes de síntomas entre positivos y negativos son muy parecidos, además de que son bajos en ambos. Los ajustes de los modelos son necesarios porque la pandemia y la población han evolucionado desde la primera ola de contagio, por lo que se deben adaptar a las nuevas necesidades y datos que se tienen actualmente.

El siguiente paso para el modelo de predicción es desarrollar un método para darle a acceso a la población mexicana con la finalidad de que ellos mismos puedan aplicar la

predicción sin la necesidad de personal de salud especializado o estudios de laboratorios previos, de manera fácil y rápida en la comodidad de sus hogares, únicamente utilizando un dispositivo electrónico y una conexión a internet para obtener acceso al modelo, para ello se debe desarrollar un método de “traducción” de los modelos para que las personas puedan predecir su prueba de manera clara, y que se logre aplicar con el uso de un cuestionario o una app, a la que tanto la población como el personal de salud tengan la oportunidad de utilizarlo rápida y a bajo o nulo costo.

Es muy probable que estos modelos se puedan complementar con métodos más sofisticados como la inteligencia artificial para elevar la sensibilidad y la especificidad, o con nuevos datos o bases que contengan los resultados de las pruebas PCR por variante u otras variables predictoras importantes como la movilidad reciente o detalles del esquema de vacunación de los pacientes.

El uso de otros métodos como complemento, como las IA, podría mejorar el desempeño del modelo.

Se espera que estos modelos de predicción puedan ayudar a predecir contagios de COVID-19 y ayudar a que las personas busquen la atención médica adecuada a tiempo.

XI Referencias

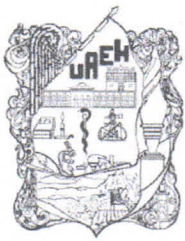
1. Gobierno de México. ¿Qué es el coronavirus? COVID-19. CDMX: Gobierno de México; 2020.
2. World Health Organization. Coronavirus disease (COVID-19). Coronavirus disease (COVID-19) Pandemic. World Health Organization. 2020.
3. Chilamakuri R, Agarwal S. COVID-19: Characteristics and therapeutics, cells. 2021; 10(2), 206.
4. Palacios M, Santos E, Velázquez MA, León M. COVID-19, a worldwide public health emergency, *Rev clín Esp.* 2021; 221(1):55–61.
5. Sharma A, Ahmad Farouk I, Lal SK. COVID-19: A review on the novel coronavirus disease evolution, transmission, detection, control and prevention, *viruses.* 2021; 13(2), 202.
6. Gao Z, Xu Y, Sun C, Wang X, Guo Y, Qiu S. A systematic review of asymptomatic infections with COVID-19, *Journal of microbiology, immunology, and infection* 2021; 54(1), 12–16.
7. Kandeel M, Ibrahim A, Fayez M, Al-Nazawi M. From SARS and MERS CoVs to SARS-CoV-2: Moving toward more biased codon usage in viral structural and nonstructural genes, *J med virol.* 2020; 92(6):660–6.
8. Centros para el control de enfermedades E. Salud Pública en los Estados Unidos y México | CDC. DBC 24/7: Salvamos vidas, EUA. 2020.
9. Tregoning JS, Flight KE, Higham SL, Wang Z, Pierce BF. Progress of the COVID-19 vaccine effort: viruses, vaccines and variants versus efficacy, effectiveness and escape, *Nature reviews.* 2021; 21(10), 626–636.
10. Iser BP, Sliva I, Raymundo V. Poletto M, Trevisol BF. Suspected COVID-19 case definition: a narrative review of the most frequent signs and symptoms among confirmed cases, *epidemiol e serv saude rev do sist unico saude do Bras.* 2020; 29(3):e2020233.
11. Cascella M, Rajnik MA, Scott A, Dulebohn CR. Features, evaluation, and treatment of coronavirus (COVID-19), *PubMed, StatPearls Publishing.* 2021.
12. Khan M, Khan H, Khan S, Nawaz M. Epidemiological and clinical characteristics of coronavirus disease (COVID-19) cases at a screening clinic during the early outbreak period: a single-centre study, *Journal of medical microbiology.* 2020; 69(8), 1114–1123.

13. Tandan M, Acharya Y, Pokharel S, Timilsina M. Discovering symptom patterns of COVID-19 patients using association rule mining. *Computers in biology and medicine*. 2021; 131, 104249.
14. Akbarialiabad H, Taghrir MH, Abdollahi A, Ghahramani N, Kumar M, Paydar S, Razani B, et. al. Long COVID, a comprehensive systematic scoping review, *Infection*. 2021; 49(6), 1163–1186.
15. Chenchula S, Karunakaran P, Sharma S, Chavan M. Current evidence on efficacy of COVID-19 booster dose vaccination against the Omicron variant: A systematic review, *Journal of medical virology*. 2022; 94(7), 2969–2976.
16. Forchette L, Sebastian W, Liu T. A Comprehensive review of COVID-19 virology, vaccines, variants, and therapeutics, *current medical science*. 2021; 41(6), 1037–1051.
17. Vassallo M, Manni S, Klotz C, Fabre R, Pini P, Blanchouin E, et al. Patients admitted for variant alpha COVID-19 have poorer outcomes than those infected with the old strain, *J clin med*. 2021; 10(16):3550.
18. Whelan VX, Bronzovich T, Haider T, Thorne LG, Towers GJ, Jolly C, et al. SARS-CoV-2 Spike evolution influences GBP and IFITM sensitivity, *bioRxiv*. 2022; 03.07.481785.
19. Merad M, Blish CA, Sallusto F, Iwasaki A. The immunology and immunopathology of COVID-19, *Science*. 2022; 375(6585): 1122-1127.
20. Knight SR, Ho A, Pius R, Buchan I, Carson G, Drake T, et al. Risk stratification of patients admitted to hospital with covid-19 using the ISARIC WHO Clinical characterisation protocol: development and validation of the 4C mortality score, *BMJ*. 2020; 370.
21. Ingrassia V. Alfa, Beta, Gamma, Delta y Ómicron: diferencias y letalidad de las variantes del COVID-19 que marcaron la pandemia. *Internacional: Infobae*; 2021.
22. Davido B, Seang S, Bariziend N, Tubiana STP. “Post-COVID-19 chronic symptoms” - Author’s reply, *Clin microbiol infect*. 2021; 27(3):495–6.
23. Hu K, Lin L, Liang Y, Shao X, Hu Z, Luo H. et al. COVID-19: risk factors for severe cases of the Delta variant, *Aging*. 2021; 13(20), 23459–23470.
24. Gobierno de México. Sistema Nacional de Vigilancia Epidemiológica, dirección de vigilancia epidemiológica. Ciudad de México: SINAVE; 2023.

25. Araf Y, Akter F, Tang YD, Fatemi R, Parvez MSA, et al. Omicron variant of SARS-CoV-2: Genomics, transmissibility, and responses to current COVID-19 vaccines, *Journal of medical virology*. 2022; 94(5), 1825–1832.
26. Pietrobon AJ, Teixeira FSM. Immunosenescence and inflammaging: risk factors of severe COVID-19 in older people, *front immunol*. 2020; 11.
27. Hemmer CJ, Löbermann M, Reisinger EC. COVID-19: Epidemiologie und mutationen: ein update [COVID-19: epidemiology and mutations : An update], *Der radiologe*. 2021; 61(10), 880–887.
28. Davido B, Seang S, Barziend N, Tubiana STP. “Post-COVID-19 chronic symptoms” - Author’s reply, *Clin microbiol infect*. 2021; 27(3):495–6.
29. Akbarialiabad H, Taghrir MH, Abdollahi A, Ghahramani N, Kumar M, Paydar S, et al. Long COVID, a comprehensive systematic scoping review, *infection*. 2021; 49(6), 1163–1186.
30. Gobierno de la Ciudad de México. Decretos, normas y acuerdos oficiales. Publicaciones en la Gaceta Oficial de la Ciudad de México, derivadas de la contingencia sanitaria por COVID-19. Ciudad de México: Gobierno de México; 2020.
31. Alcaldía Álvaro Obregón. Alcaldía Álvaro Obregón - Sitio oficial. Alcaldía Álvaro Obregón, Tu alcaldía aliada, COVID-19. Ciudad de México: Gobierno de México; 2020.
32. Espino TC, Martínez FX. “Análisis predictivo: técnicas y modelos utilizados y aplicaciones del mismo - herramientas Open Source que permiten su uso. 2017; (Principio activo y prestación ortoprotésica): 67.
33. Fiuza MD, PJ. La regresión logística: una herramienta versátil, *Nefrología, Publicacion oficial de la Sociedad Espanola Nefrologia*. 2000.
34. Roland LT, Gurrola JG, Loftus PA, Cheung SW, Chang JL. Smell and taste symptom-based predictive model for COVID-19 diagnosis, *Int forum allergy rhinol*. 2020; 10(7):832–8.
35. Jehi L, Ji X, Milinovich A, Erzurum S, Rubin BP, Gordon S, et al. Individualizing risk prediction for positive coronavirus disease 2019 testing: results from 11,672 Patients, *Chest*. 2020; 158(4):1364–75.

36. Zhang SX, Sun S, Afshar JA, Wang Y, Nazarian MA, Li J, Mokhtari DM. Beyond predicting the number of infections: predicting who is likely to be COVID negative or positive, risk management and healthcare policy. 2020; 13, 2811–2818.
37. Wynants L, Van CB, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal, *BMJ*. 2020; 369.
38. Carrillo VMF, Salinas EG, García PC, Gutiérrez RLM, Parra RL. Early estimation of the risk factors for hospitalization and mortality by COVID-19 in Mexico, *PLoS One*. 2020; 15(9).
39. Mamidi T, Tran NK, Melvin RL, Worthey EA. Development of an individualized risk prediction model for COVID-19 using electronic health record data, *front big data*. 2021; 4.
40. Challener D, Challener G, Gow-Lee V, Fida M, Shah A, O'Horo J. Screening for COVID-19: patient factors predicting positive PCR test, *Infect control hosp epidemiol*. 2020; 41(8):968–9.
41. Alyafei K, Ahmed R, Abir F, Chowdhury H, Naji K. A comprehensive review of COVID-19 detection techniques: From laboratory systems to wearable devices. *Computers in biology and medicine*. 2022; 149, 106070.
42. Fang L, Liang X. ISW-LM: An intensive symptom weight learning mechanism for early COVID-19 diagnosis. *Computers in biology and medicine*. 2022; 146, 105615.
43. Hu H, Du H, Li J, Wang Y, Wu X, Wang C, et al. Early prediction and identification for severe patients during the pandemic of COVID-19: A severe COVID-19 risk model constructed by multivariate logistic regression analysis, *Journal of global health*. 2020; 10(2), 020510.
44. Santosh KC. COVID-19 Prediction models and unexploited data, *Journal of medical systems*. 2020; 44(9), 170.
45. Organización Panamericana de la Salud. De información y las tecnologías en el primer nivel de atención. Organización Mundial de la Salud. 2020; 1–4.
46. Salud Digna. ¿Síntomas de COVID? Realízate una prueba. México: Salud Digna; 2023.
47. Laboratorio Médico del CHOPO. COVID. México: CHOPO; 2023.

48. Argimon P, Jimenés J. Métodos de investigación clínica y epidemiológica, Elsevier. 2019; 5ª edición.
49. Rosner B. Fundamentals of biostatistics, Harvard University, Books cole cengage learning. 2010; 7ª edición.
50. Chap T. Introductory biostatistics, Wiley Interscience. 2003; 1ª edición.
51. Fontal A, Bouma MJ, San José A, Lopez L, Pascual M, Rodó X. Climatic signatures in the different COVID-19 pandemic waves across both hemispheres. Nature comput sci. 2021.



San Agustín Tlaxiaca, Hidalgo a 24 de mayo de 2023
Oficio Comiteei.icsa ICSa «017» / 2023

Asunto: DICTAMEN DEL COMITÉ DE ÉTICA E INVESTIGACIÓN.

Victorino Aguilar Mariel
Investigador Principal
Correo: vi244818@uaeh.edu.mx

PRESENTE

Título del Proyecto: Desarrollo y validación epidemiológica de un modelo de predicción para pacientes COVID-19 en Álvaro Obregón de la Ciudad de México.

Le informamos que su proyecto de referencia ha sido evaluado por el Comité de Ética e Investigación del Instituto de Ciencias de la Salud y las opiniones acerca de los documentos presentados se encuentran a continuación:

Decisión
«Aprobado»

Este protocolo tiene vigencia del 24 de mayo de 2023 al 24 de mayo de 2024.

En caso de requerir una ampliación, le rogamos tenga en cuenta que deberá enviar al Comité un reporte de progreso de avance de su proyecto al menos 60 días antes de la fecha de término de su vigencia.

Le rogamos atender las indicaciones realizadas por el revisor, y enviar nuevamente una versión corregida de su protocolo para una nueva evaluación.

Atentamente

Dra. Itzia María Cazares Palacios
Presidenta del Comité



Para la validación de este documento, informe el siguiente código en la sección Validador de documentos del sitio web del Comité de Ética e Investigación del Instituto de Ciencias de la Salud: «CS*h@#=#+j»

<https://sites.google.com/view/comiteei-icsa/validador-de-documentos>



Circuito ex-Hacienda La Concepción s/n
Carretera Pachuca Actopan, San Agustín
Tlaxiaca, Hidalgo, México. C.P. 42160
Teléfono: 52 (771) 71 720 00 Ext. 4306
investigacion_icsa@uaeh.edu.mx

www.uaeh.edu.mx

Estudio epidemiológico de caso sospechoso de enf

DATOS GENERALES

ANEXO 1

Nombre de la unidad:

Fecha de notificación en plataforma: dd/mm/aaaa Folio plataforma:

Apellido Paterno: Apellido Materno: Nombre (s):

Fecha de Nacimiento: Día: Mes: Año: CURP:

Sexo: Hombre: ¿Está embarazada? Si No Meses de embarazo: Se encuentra en periodo de puerperio Si No Dias de puerperio

Nacionalidad: Mexicana: Extranjera: ¿Es migrante? Si No País de nacionalidad: País de origen:

Países en tránsito en los últimos tres meses: 1 2 3 Otro: Fecha de ingreso a México:

País de nacimiento: Entidad federativa de nacimiento:

Entidad de Residencia: Municipio de residencia:

Localidad:

Calle: Número:

Entre qué calles: y

Colonia: C:P: Teléfono:

¿Se reconoce cómo indígena? Si No ¿Habla alguna lengua indígena? Si No

Ocupación:

¿Pertenece a alguna institución educativa?

DATOS CLÍNICOS

Servicio de ingreso: Tipo de paciente: 1=Ambulatorio 2=Hospitalizado

Fecha de ingreso a la unidad: dd/mm/aaaa Fecha de inicio de síntomas: dd/mm/aaaa

A partir de la fecha de inicio de síntomas:

¿Tiene o ha tenido alguno de los siguientes signos y síntomas?

	Si	No
Inicio súbito de los síntomas		
Fiebre		
Tos		
Cefalea		
Disnea		
Irritabilidad		
Dolor torácico		
Escalofríos		
Odinofagia		
Mialgias		
Artralgias		
Anosmia		
Disgeusia		
Rinorrea		
Conjuntivitis		

Co-morbilidad

	Si	No
Diabetes		
EPOC		
Asma		
Inmunosupresión		
Hipertensión		
VIH/SIDA		
Enfermedad cardiovascular		
Obesidad		
Insuficiencia renal crónica		
Tabaquismo		
Otros		

Especifique otros:

Otros síntomas	Si	No
Ataque al estado general		
Diarrea		
Polipnea		
Dolor Abdominal		
Vómito		
Cianosis		

Diagnóstico probable: 1=Enfermedad tipo influenza (ETI)
2=Infección respiratoria aguda grave (IRAG)

*ETI es considerada como Enfermedad respiratoria leve

TRATAMIENTO

¿Desde el inicio de los síntomas ha recibido tratamiento con antipiréticos?

Si	No
<input type="text"/>	<input type="text"/>

¿Desde el inicio de los síntomas ha recibido tratamiento con antivirales?

Si	No
<input type="text"/>	<input type="text"/>

Si la respuesta fue afirmativa:
 Seleccione el antiviral: 1=Amantadina 2=Rimantadina 3=Oseltamivir
 4=Zanamivir 5=Otro, Especifique otro:

¿Cuándo se inicio el tratamiento antiviral? dd/mm/aaaa

En la unidad médica:

¿Se inicia tratamiento con antimicrobianos ?

Si	No
<input type="text"/>	<input type="text"/>

¿Se inicia tratamiento con antivirales?

Si	No
<input type="text"/>	<input type="text"/>

Seleccione el antiviral: 1=Amantadina 2=Rimantadina 3=Oseltamivir
 4=Zanamivir 5=Otro, Especifique otro:

ANTECEDENTES EPIDEMIOLÓGICOS

¿Tuvo contacto con casos con enfermedad respiratoria en las ultimas dos semanas?

Si	No
<input type="text"/>	<input type="text"/>

Durante las semanas previas al inicio de los síntomas tuvo contacto con:

Aves

Si	No
<input type="text"/>	<input type="text"/>

 Otro animal

Cerdos

Si	No
<input type="text"/>	<input type="text"/>

¿Realizó algún viaje 7 días antes del inicio de signos y síntomas?

Si	No
<input type="text"/>	<input type="text"/>

País: Ciudad:

¿Recibió la vacuna contra influenza en último año?

Si	No
<input type="text"/>	<input type="text"/>

Fecha de vacunación: dd/mm/aaaa

LABORATORIO

¿Se le tomó muestra al paciente?

Si	No
<input type="text"/>	<input type="text"/>

Laboratorio al que se solicita el procesamiento de la muestra:

Tipo de muestra: 1=Exudado faríngeo 2=Exudado Nasofaríngeo
 3=Lavado bronquial 4=Biopsia de pulmón

Fecha de toma de muestra: dd/mm/aaaa

Resultado:

EVOLUCIÓN

Evolución: 1=Alta 2=En tratamiento/Referencia/Seguimiento domiciliario/Seguimiento terminado
 3=Caso grave 4=Caso no grave 5=Defunción*

Si el caso se da de alta: Especifique la evolución: 1=Mejoría 2=Curación
 3=Voluntaria 4=Traslado

¿El caso está o estuvo ingresado en la UCI durante la enfermedad?

Si	No
<input type="text"/>	<input type="text"/>

¿El caso está o estuvo intubado en algún momento durante la enfermedad?

Si	No
<input type="text"/>	<input type="text"/>

¿El caso tiene o tuvo diagnóstico de neumonía durante la enfermedad?

Si	No
<input type="text"/>	<input type="text"/>

Fecha de egreso: dd/mm/aaaa

Se trató de un caso positivo a COVID-19 por asociación o dictaminación clínica-epidemiológica:*

* Marque solo una de las siguientes opciones de respuesta con una X

- a. Confirmado a COVID-19 por asociación clínica-epidemiológica
- b. Confirmado a COVID-19 por dictaminación clínica-epidemiológica (aplica solo en defunciones)
- c. No (es decir ninguna de las anteriores)

Defunción: Fecha de defunción: dd/mm/aaaa

Folio de certificado de defunción *Defunción por influenza o COVID-19

Si	No
<input type="text"/>	<input type="text"/>

*Anexar copia de certificado de defunción si cumple con definición operacional de caso sospechoso de enfermedad respiratoria viral

Nombre y cargo de quien elaboró _____ Nombre y cargo de quien autorizó _____ Fecha de elaboración: dd/mm/aaaa

ANEXO 2

Pruebas de diagnóstico COVID-19

La obtención y análisis de las muestras para el diagnóstico de COVID-19 se realizará conforme al Lineamiento Estandarizado para la Vigilancia Epidemiológica y por laboratorio de la Enfermedad Respiratoria Viral emitido por la Secretaría de Salud. Se considerará caso confirmado a la persona que cumpla con la definición operacional de caso sospechoso y que cuente con diagnóstico confirmado por laboratorio de la Red Nacional de Laboratorios de Salud Pública reconocidos por el Instituto de Diagnóstico y Referencia Epidemiológicos (InDRE) (1).

El procedimiento de evaluación clínica y toma de muestra se realizará conforme a las medidas de precaución para el manejo de pacientes con enfermedad respiratoria viral, para el personal de salud (1).

La toma de muestra será de tipo exudado faríngeo o nasofaríngeo conforme al procedimiento descrito en. Como material se utilizará: medio de transporte viral, hisopos de dacrón o rayón con mango de plástico (exudado faríngeo) e hisopos de dacrón o rayón con mango flexible (exudado nasofaríngeo). Se usará temperatura de transporte de 2 a 8°C y de almacenamiento de 2 a 8°C (tiempo menor a 5 días). Se exudado faríngeo o nasofaríngeo se deben colocar en el mismo tubo para incrementar la carga viral. Conforme a la Guía de la OMS sobre regulaciones para el transporte de sustancias infecciosas 2019-2020, las muestras serán transportadas como UN3373, "sustancia biológica, Categoría B". Las muestras serán enviadas de manera inmediata (menos de 24 horas) al Laboratorio Estatal de Salud Pública (LESP) o al Laboratorio de Apoyo a la Vigilancia Epidemiológica (LAVE), avalados como Laboratorios de la Red Nacional de Influenza más cercano (1).

La información clínica del paciente será ingresada al sistema de vigilancia epidemiológica con la creación de un folio para seguimiento del caso y la notificación del resultado de la prueba a través del centro de salud donde se realizó la toma de la muestra y la evaluación del caso. El Anexo 1 muestra el formato con las variables de información clínica utilizado en el sistema de vigilancia epidemiológica (1).

Referencias

1. Secretaría de Salud. Lineamientos vigentes de la Red Nacional de Laboratorios de Salud. Secretaría de Salud; 2020 (citado en May 30, 2020). Disponible online en: <https://www.gob.mx/salud/documentos/lineamientos-vigentes-red-nacional-de-laboratorios-de-salud-publica?state=published>



SALUD
SECRETARÍA DE SALUD



Instituto Nacional de Cardiología Ignacio Chávez
Renacimiento de la Excelencia

Ciudad de México, a 10 de febrero de 2023

Dr. Manuel Sánchez Gutiérrez

Coordinador de la Maestría en Ciencias Biomédicas y de la Salud
Universidad Autónoma del Estado de Hidalgo
Presente

Estimado Dr. Sánchez:

Por este medio hago constar que he acordado con la alumna **Maríel Victorino Aguilar** y su tutor de tesis, *Dr. Abel Lerma Talamantes*, que la alumna realice una estancia de estancia de investigación en mi laboratorio del 13 de febrero al 27 de marzo del año en curso, como parte de sus actividades del 4º semestre de la maestría, en la que está inscrita con número de cuenta 244818 y número de CVU 1146841.

El objetivo principal de la estancia es concluir el análisis de resultados del proyecto de tesis y la redacción de un artículo de investigación, conforme al cronograma adjunto.

Sin más por el momento, quedo a sus órdenes para cualquier duda sobre la estancia y este documento, y aprovecho la ocasión para enviarle un cordial saludo.

Atentamente,

Dra. Claudia Lerma González

Investigador en Ciencias Médicas D

Investigador Nacional (SNI) nivel 2

Departamento de Instrumentación Electromecánica

Instituto Nacional de Cardiología Ignacio Chávez

Teléfono: (55) 5573 2911 extensión 26202 (oficina); (55) 3233 4062 (celular)

Correo electrónico: dr.claudialerma@gmail.com

Models for individualized COVID-19 diagnostic prediction

Modelos para predicción de diagnóstico individualizado de COVID-19

Mariel Victorino- Aguilar^a, Abel Lerma-Talamantes^b, Claudia Lerma^c

Abstract:

Since the COVID-19 pandemic, the world has experienced a large incidence of infections in short periods of time, giving rise to waves of contagion caused by the different variations of SARS-COV-2. Health services, as well as personnel, have been overwhelmed, especially in the poorest countries. Currently and after two years, the pandemic continues and according to experts it is here to stay, which highlights the importance of vaccines and methods of detecting the disease, to curb the number of infections and avoid that the pandemic continues to spread and thus the virus continues to mutate. Detection tests have been scarce and expensive for most of the population, so alternative methods to laboratory ones could be a decisive factor so that people can self-isolate before continuing to infect more people. One of the most effective methods have been statistical predictions of the diagnosis of COVID-19 in a patient, based on certain variables. In this article, it was identified that the most common prediction models were developed from logistic regression and machine-learning, which have shown high percentages of predicting test results for COVID-19. The most important predictor variables in the different models developed in various regions of the world were identified and the opportunities, limitations and perspectives of this prediction method are discussed.

Keywords:

COVID-19, individual prediction, prediction models, symptoms, detection

Resumen:

Desde que la pandemia del COVID-19, el mundo ha vivido el gran número de infecciones en periodos cortos de tiempo, dando lugar a las olas de contagio provocadas por las distintas variaciones del SARS-COV-2. Los servicios de salud, así como el personal, se han visto superados, esto especialmente en los países más pobres. Actualmente y después de dos años, la pandemia continua y según expertos llegó para quedarse de forma estacionaria, por lo que hoy más que nunca la importancia de las vacunas y de los métodos de detección de la enfermedad, para frenar el número de contagios y evitar que la pandemia siga extendiéndose y así el virus siga mutando. Las pruebas de detección han resultado escasas y caras para la mayoría de población, por lo que los métodos alternativos a los de laboratorio podría ser un factor decisivo para que las personas puedan autoaislarse antes de seguir contagiando a más personas. Uno de los métodos más eficaces ha sido lo que involucran predicciones estadísticas de diagnóstico de COVID-19 en un paciente, a partir de ciertas variables. En este artículo se identificaron que los modelos de predicción más comunes se desarrollaron a partir de la regresión logística e inteligencia artificial, el objetivo de este trabajo es demostrar los altos porcentajes de predicción de resultado de prueba por COVID-19 de estos métodos alternativos a las pruebas de laboratorio, para mostrar que son confiables como alternativas a ellas, y aplicarlas a la población como método de control de la pandemia de COVID-19. Se identificaron las variables predictoras más importantes en los distintos modelos desarrollados en varias regiones del mundo y se discuten las oportunidades, limitaciones y perspectivas de este método de predicción.

Palabras Clave:

COVID-19, predicción individual, modelos de predicción, síntomas, detección.

^a Departamento de Instrumentación Electromecánica, Instituto Nacional de Cardiología Ignacio Chávez, <https://orcid.org/0000-0002-1338-3253>, Email: biomedica.ingmva@outlook.com.

^b Departamento de Instrumentación Electromecánica, Instituto Nacional de Cardiología Ignacio Chávez, <https://orcid.org/0000-0001-7212-641X>, Email: aleta_@yahoo.ca.

^c Departamento de Instrumentación Electromecánica, Instituto Nacional de Cardiología Ignacio Chávez, <https://orcid.org/0000-0002-4679-7751>, Email: dr.claudialerma@gmail.com

INTRODUCTION

The COVID-19 disease is caused by the SARS-CoV-2 virus, which began to spread as a pandemic since 2020, affecting all countries in the world with almost 6 million deaths worldwide until this year.^{1,2} This disease has brought a number of subsequent problems, from economic to health problems, revealing how little is known today about viruses, since, although there are already vaccines that are effective in preventing the development of severe COVID-19 by up to 98%, there are still no drugs that cure the disease.³ However, the fact that people are vaccinated with one or more doses, although they do not prevent infection, does significantly lower the chances of being hospitalized, as well as dying.⁴ Here is the importance of continuing to vaccinate as many people as possible. Moreover, although the number of infections has fallen by a large percentage, there are still a large number of infected people worldwide.²

Due to the large number of infections, it is of vital importance that detection tests continue to be done for COVID-19 infection, for people to isolate themselves, and by not having contact with other people, contagion can be avoided, especially with the last variant of interest of COVID-19, which caused the most recent wave of infections, and whose symptoms are more like the common cold, so it is more likely that people will become infected even without knowing it.⁵

All this without mentioning that the long-term sequelae are still being studied and that the vast majority of recovered patients have reported, which approach from headaches, psychological problems, to multiple organ failure (varying in each patient, and depending on many factors).⁶ Due to all this, it is of the utmost importance to prevent more people from getting infected, which could be avoided exponentially by increasing the number of tests performed, but because they have been scarce, as well as the medical personnel who apply them, and having such high costs, new detection methods have been developed, among which statistical methods stand out, such as logistic regression models, which through certain variables, are able to satisfactorily predict the result of an individualized test for COVID-19.^{7,8}

Because the models for individualized prediction of COVID-19 that have been reported so far are heterogeneous in terms of their composition and performance, as well as the applicability in the time and region in which they were developed, the objective of this work was to analyze the characteristics of the individualized prediction models of COVID-19 that have been reported, identify the main predictor variables, the performance of the models, their limitations and strengths.⁹

VARIANTS OF INTEREST OF THE SARS-COV-2 VIRUS

The mutation of the SARS-CoV-2 virus has occurred constantly since its emergence, through the genetic changes it has undergone, also called genetic mutations in which the genome is replicated as there are still a large number of people infected and who continue to infect, the virus acquires new mutations, which are classified according to the World Health Organization (WHO), as variants of interest or not.^{2,10} So far 3

variants of interest have been detected because they can cause serious health problems for humans, and in the most serious cases, even death.¹¹

The variants of interest are: the original strain, delta and omicron, each of which has caused millions of infections and thousands of deaths around the world.¹⁰

Although cases have decreased drastically thanks to vaccination in many countries of the world, there is still a significant lag in terms of the least developed countries, so the number of infections worldwide continues to increase.¹²

WAVES OF CONTAGION IN THE PANDEMIC

Since the beginning of the pandemic, in China, in 2019, four waves of contagion have been detected worldwide, which stand out dramatically for the numbers of infections registered until that moment.¹³ According to information provided by the WHO and the Centers for Disease Control (CDC) of USA, the first wave was recorded in July-August 2020, the second in January 2021, the third in August of the same year, and the fourth and most recent one, in January 2022.^{2,14} Consequently, it is of the utmost importance to continue with sanitary measures and apply a greater number of reliable screening tests, to avoid or mitigate the effects of a possible fifth wave, since health systems are on the verge of collapse, in addition to other ailments have been left aside by the resources that have been allocated to combat the pandemic.

New models for predicting future waves or peaks of contagion are currently being developed, all thanks to the data that has been collected since this new virus emerged.¹⁵

COVID-19 DIAGNOSTIC STRATEGIES

Currently, millions of screening tests have already been applied for COVID-19, the most reliable has turned out to be the PCR test, with an approximate 90-98% effectiveness in detecting the presence of the virus, followed by the antigen test with an approximate of 82-90% detection efficiency.^{16,17} This type of test is followed by laboratory studies that are generally also used to detect the level of severity of the disease in patients, in addition to initially being used to diagnose whether a laboratory test cannot be accessed.¹⁸ However, in view of the high demand for tests, the scarcity and cost of such tests, the time they take, and the lack of medical personnel applying them due to the high risk of contagion when applying them, the development of other types of detection methods has been resorted to, among which statistical methods are applied to try to predict as accurately as possible, the outcome that patients would have for the COVID-19 disease. One of the most effective methods within this category has turned out to be mathematical modeling through logistic regression and artificial intelligence (AI), since they use different independent variables. Although there are variations between the models developed, most model-independent variables include the most frequent symptomatology in infected patients, exposure to the virus, place of residence, and previous medical conditions of the patient.¹⁹ These variables were analyzed separately and then in combinations, to know

what the predictor variables for the result of the test for COVID-19 would be.²⁰⁻²⁶

In other words, several previously known independent variables are used to predict the dependent variable, which has shown us a large percentage of success when predicting the result of the test, being able to even compare with the results obtained with the most used laboratory tests, but with less response time, exposure of both patients and staff, and lower cost.

STATIC PREDICTIONS MODELS FOR COVID-19

There are innovative models for the individualized prediction of the result of the COVID-19 test, developed with statistical methods and that do not require in their variables the result of laboratory analyzes (except for the COVID-19 diagnostic test used as a gold standard to compare the outcome) or physically review patients.⁴

The models that stand out from the others have been those of logistic regression and machine-learning as this is the most used method and that had reported the best results in the various works that were reviewed, and as mentioned above, one of its characteristics. The key is that they don't need laboratory studies to make their predictions.²⁰⁻²⁶

The methods have variations in their population type, sampling method, sample, data collection method, statistical analysis, variables used, and results obtained.²⁰⁻²⁶ A comparison analysis of the models was carried out, considering their main characteristics, such as the method of solution of the model, the dependent and independent variables, the population studied, the predictor variables, and the final results.

MODEL OVERVIEW AND ANALYSIS

The models analyzed showed similar structures in terms of research development (Table 1 and Table 2). The information was collected through questionnaires, interviews or databases, with the independent variables previously determined through research that showed which were the risk factors for positive COVID-19 test. Many similar prescribing variables were found, such as the main symptomatology presented by previous confirmed cases (sore throat, fever, cough, headache, changes in smell or taste, difficulty breathing), exposure to the virus by contact with people who were known to be infected, smokers, or recent trips (last two weeks) or co-morbid diseases.^{20,21,23,24,26} However, others were not as common, such as gender (women), race (African-Americans) and demographic data (whether or not they live in metropolitan areas or with large percentages of the population) and the level of physical activity they performed.^{22,25} Another key point that was considered in only one study is the psychological problems derived from the COVID-19 disease, which could also be determinants as predictive factors (anxiety, depression, insomnia).²⁷ This was determined for the population and subsequent sample, which in general only required that they have their COVID-19 test, either the PCR or the quick test.²⁰⁻²⁶ Having access to patients data before and after the test was useful so they were able to perform a validation phase.^{23,25}

After identifying the characteristics of their sample, the most frequent variables in these patients were analyzed, in order to determine which were the predictor variables for the models, as well as to make a

combination of them to make the predictions with a greater degree of accuracy.²⁸ Some studies used one or more predictive or statistical methods, the most common being logistic regression (Table 1), and AI (Table 2).²⁰⁻²⁶

MODEL COMPARISONS

The main variations found in the models, in addition to the type of analysis that was performed, were the independent variables selected for the predictions (Table 1 and Table 2). When a large number of independent variables were reported in various studies, and as they varied at the time they were taken into account for the analyses, it can be considered that there was more or less information about the disease in general, so the researchers decided each one by their own variables. In most models the independent variables were similar, but in other cases they were not. This could explain the difference in the performance of the models, which overall was between 70 and 87% of success effectiveness in positive COVID-19 test.²⁰⁻²⁶

Most of the models were made with samples taken in short periods of time, about one month, and the longest was 4 months.²⁰⁻²⁶ In most cases, the investigations were conducted in the United States of America, and all in the year 2020 (Table 1 and Table 2).

ADVANTAGES AND DISADVANTAGES OF MODELS

The models found and analyzed showed effective results (Table 1 and Table 2). However, the diagnostic value of the models is lower than the laboratory results, since there are variants that cannot be controlled, such as the large number of symptoms that the COVID-19 disease has presented, in addition to the fact that these vary from person to person depending on various characteristics that patients present.²⁹ External factors that contribute to the fact that the results are not more uniform, such as their demographic characteristics and the fact that a large percentage of the population travels due to their daily activities or work, may also affect the predictive value of the models.²⁰⁻²⁶ Another factor to consider is that, while research has determined that people have been exposed to the virus by having contact with an infected patient as a highly valuable predictor, many of these patients who may have been exposed do not know it because the other people have not been tested, or did not have the general symptoms.³⁰

Some research mentions the fact that one of its limitations is not to give registered follow-up to patients, since having subsequent information from them to be able to reinforce the prediction systems with this data.^{20-24,26}

However, and despite all the above, the models analyzed have shown optimal performance when determining the predictor variables, and derived from it, final results above 70% of positive COVID-19 test prediction, some even reaching more than 87% efficacy.²⁰⁻²⁶ With this performance, it would be considered that patients analyzed with this type of method could have high certainty of the results, because these percentages can be compared with those of the rapid test for COVID-19, this being one of the most used. Other great advantages to highlight about these methods are their cost, since, in times of pandemic, this type of tests increased their monetary value, increasing the number of people not taking the test.³¹ In addition, a shortage of tests has been reported due to

the large number of infections recorded, mainly during the waves of contagion, in addition to the lack of medical personnel to apply them due to the possible exposure to the virus, and the time it takes people to perform the test in a laboratory.³¹ For these reasons, individualized prediction models based on statistical models and patient symptoms could be more comfortable, cheaper and faster for people.

SYNTHESIS OF THE MAIN CHARACTERISTICS OF THE PUBLISHED MODELS

The results showed mostly models that used statistical or mathematical methods (Table 1), in addition to artificial intelligence (Table 2) to make predictions from certain variables considered and studied by the different authors, which yielded predictor variables for the final prediction model. Most of the research reviewed showed at least 70% effectiveness in predicting outcome, and in the best cases, almost 90% efficacy.²⁰⁻²⁶ Therefore, they could be a useful option for more people with symptoms or risk factors to be tested and, if necessary, seek appropriate medical attention.

Most of the models were developed in the US, through electronic surveys or databases of health centers that treated Covid-19 patients, which had the result of their Covid-19 test (PCR, nasopharyngeal or pharyngeal), and almost all considered adult patients, all with the main objective of predicting the result of the positive test for Covid-19, through statistical methods of logistic regression or applying artificial intelligence. The most suitable study designs, mostly cohort, prospective and retrospective, were used as models. The independent variables taken to predict the dependent (Covid-19 test result) were the main ones reported for patients confirmed by the virus, such as the most frequent symptomatology (sore throat, body, headache, changes in smell or taste), co-morbid diseases, exposure to the virus, workplaces, travel, age, gender, race, psychological problems, economic position, among others.²⁰⁻²⁶

Most of these models performed validation stages so that the models could improve or adjust to their first results.^{21,23-25}

The development of this research provides the population with new methods to try to stop the infections, and with it, the probability that the virus continues to mutate. In this way, the world could recover little by little from the ravages in all areas that the pandemic has been causing, not to mention that vaccines are still scarce and their annual application has been recommended, despite the fact that a large percentage of the world's population has not been able to receive even the first dose, so today more than ever, the prevention and control of infections is fundamental.²⁰⁻²⁶

OPPORTUNITIES, CONSTRAINTS AND PERSPECTIVE

As already mentioned, the fact that new methods of diagnosing COVID-19 can be developed, of more scope, lower cost and time, that do not expose the health personnel and above all, is faster, which would avoid early detection and that the patient continues to infect people. It turns out to be a solution to various problems caused by the pandemic, especially to try to curb the number of infections and prevent new variants of interest and new waves of contagion from emerging.²⁰⁻²⁶

Some of the limitations reported by the researchers are the variation in symptoms, the displacement of people, and the lack of knowledge of whether they were exposed to the virus or not, this also derived from the lack of tests to detect the virus.³² In addition, the information used is from 2020, so it is possible to deduce the lack of information that we have today, and which can be considered important for this type of model, such as the symptoms of each variant of interest, which has been shown to be very different from each other, and take into account the vaccination that has been applied as a possible confounding variant to reduce the symptoms, but not to prevent contagion itself.²⁰⁻²⁶

It is essential that new effective and reliable methods for the early detection of the virus continue to be developed, these methods could be of great help for the entire medical system, and in the future for patients, since it has been shown that the virus leaves sequelae in up to 50% of infected patients, depending on the time they were infected and if they had any dose of vaccination.³³

CONCLUSIONS

The models analyzed proved to be reliable and with less expenditure of resources; however, it is important to incorporate new information that in 2020 was still unknown, in order to have more accurate and reliable models, however, these new methods have great potential, in addition to being able to be applied as a first way to analyze patients, and in this way, if necessary, direct to laboratory tests to those who require it most, and thus be able to optimize medical, personnel and time resources.

Table 1. Comparison of current individualized test prediction models for Covid-19, part 1

Research name/ characteristics	Smell and taste symptom-based predictive model for Covid-19 diagnosis. ²⁰	Individualizing risk prediction for positive coronavirus disease 2019 testing. ²¹	Beyond predicting the number of infections: predicting who is likely to be covid negative or positive. ²²	Development of an individualized risk prediction model for Covid-19 using electronic health record data. ²³
-----------------------------------	---	---	--	---

Country-region	USA ^c	USA ^c	Iran	USA ^c
Objective	Covid-19 Positive Test Prediction	Covid-19 positive test prediction	Covid-19 Test Result Prediction	Covid-19 Test Result Prediction
Statistical method	Binary logistic regression	Logistic regression	Logistic regression	Logistic regression
Study design	Cohort	Prospective, validation phase: cohort	Not mentioned	Longitudinal cohort
Population	Covid-19 patients in California, USA, who have a PCR ^a test	Covid-19 patients in Florida, USA	Population in Iran	University of Alabama Patients Tested for Covid-19
Sampling method	Estimation of a proportion	Not mentioned	Not mentioned	Not mentioned
Sample size	145 participants with a positive test for Covid-19 and 157 with a negative test for Covid-19	11672 patients with general characteristics	521 adults with general characteristics	7262 patients admitted to the University of Alabama hospital
Method of obtaining the information	Anonymous surveys via the internet	Registry of patients analyzed for Covid-19, in Cleveland clinic	Primary study through surveys	Database of patients in the health system of the University of Alabama who have been tested for Covid-19
Dependent variable	PCR ^a test result	PCR ^a or nasal test result	PCR ^a test result	PCR ^a test result
Independent variable	Symptoms, demographic information, co-morbidities	Sex, age, symptoms, economic status, race	Sex, age, co-morbidities, active workers, demographics, exercise, depression, anxiety	Age, gender, race, health condition, smoker, addictions, weight
Confusor variant	Myalgia, gastrointestinal problems	Common Symptoms	Gender and age	Age, gender and race
Significance predictors	Changes in smell, unexplained body aches, fever, or chills	Female, African American, elderly, and having been exposed to the virus	Physical activity, workplace, co-morbid diseases	Substance abuse, smoking and co-morbidities
Model Solution Method	Stepwise	Multivariable	Ordination	Linear
Scan time	March-April 2020	April 2020	April 2020	January-June 2020
Validation stage	Not carried out, but considered necessary	Performed with 2295 patients	Not realized	Performed
Results	82% in Covid-19 test result discrimination, 75% correct predictions	The concordance index corrected by bootstrap in the development cohort was 0.863 (95% CI ^b , 0.852-0.874). The concordance index in the Florida validation cohort was 0.839 (95% CI ^b , 0.817-0.861)	Separate analysis of predictor variables, all showed more than 70% chance of positive Covid-19 test	Time-separated and variant analysis, general analysis with more than 70% prediction efficiency

^a PCR: polymerase chain reaction, ^b CI: confidence index, ^c USA: United States of America

Table2. Comparison of current individualized test prediction models for Covid-19, part 2

Name of the research/ characteristics	Machine learning-based prediction of Covid-19 diagnosis based on symptoms. ²⁴	Prediction of individual Covid-19 diagnosis using baseline demographics and lab data. ²⁵	Screening for Covid-19: patient factors predicting positive PCRA test. ²⁶
Country-region	Israel	USA ^c	USA ^c
Objective	Covid-9 Positive Test Prediction	Covid-9 Positive Test Prediction	Covid-19 Positive Test Prediction
Statistical method	Machine-Learning	Machine-Learning, baseline data	Logistic regression
Study design	Prospective	Cohort	Retrospective
Population	Patients in Israel tested for Covid-19	Patients from hospitals in the New York metropolitan area	Patients admitted to Rochester, Minnesota clinic
Sampling method	Not mentioned	Stratified sub-populations	Not mentioned
Sample size	51,831 individuals tested (with 4769 confirmed by Covid-19)	31739 adults without a health system	48 positive and 98 negative patients from the Rochester, Minnesota clinic
Method of obtaining the information	Public information reported by the Minister of Health and Israel	Clinic databases	Questionnaire applied to patients by a nurse
Dependent variable	PCR ^a test and nasopharyngeal test	PCR ^a test	PCR ^a test result
Independent variable	Cough, fever, contact with infected people, sex, age over 60, headache and breathing problems	Demographics, common co-morbidities and laboratory tests, calcium levels, temperature, age, blood tests, smokers, oxygen saturation	Fever, sneezing, respiratory problems, co-morbidities, travel and exposure to the virus
Confusor variant	Headache, shortness of breath and cough	Temperature and blood tests	Fever, chills
Significance predictors	Sex, age over 60 years, exposure to the virus and appearance of at least 5 clinical symptoms	Common co-morbidities and laboratory tests, calcium levels, temperature, age, blood tests, smokers, oxygen saturation	Exposure to the virus and travel to metropolitan areas
Model Solution Method	Decision Tree	Decision tree, random forest, XGBoost multi-tree and logistic regression	Multivariable
Scan time	March-April 2020	April-June 2020	March 2020
Validation stage	Performed with 1000 repetitions (ROC ^b)	Realized (ROC ^b)	Not realized
Results	87.30% sensitivity, 71.98% specificity, or 85.76% sensitivity and 79.18% specificity	Random forest: 79.10% accuracy, multi-tree XGBoost: 77.66% accuracy, logistic regression: 79.05% accuracy and single-tree XGBoost: 79.37% accuracy	Contact with confirmed cases increases the odds of positive test by 17 times (95% CI ^d 4.6–88.4), and recent trips increases the odds of positive test by 4.7 times (95% CI 1.9–12.7).

^a PCR: polymerase chain reaction, ^b ROC: operational characteristic of the receptor, ^c USA: United States of America, ^d CI: Confidence indicator

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest with any institution or commercial association of any kind or field. This research was carried out following the corresponding guidelines.

REFERENCES

- [1] Berekaa MM. Insights into the COVID-19 pandemic: Origin, pathogenesis, diagnosis, and therapeutic interventions. *Front. Biosci. (Elite Ed)*. 2021;13(1):117–139.
- [2] World Health Organization. Coronavirus disease (COVID-19) pandemic. *W. H. O*. 2021:1-2.
- [3] Uddin M, Mustafa F, Rizvi TA, Loney T, Suwaidi HA, Al-Marzouqi A, et al. SARS-CoV-2/COVID-19: Viral Genomics, Epidemiology, Vaccines, and Therapeutic Interventions. *Vir*. 2020;12(5):526.
- [3] Asselah T, Durantel D, Pasmant E, Lau G, Schinazi RF. COVID-19: Discovery, diagnostics, and drug development. *J. Hepatol*. 2021;74(1):168–184.

- [4] He X, Hong W, Pan X, Lu G, Wei X. SARS-CoV-2 Omicron variant: Characteristics and prevention. *MedCom*. 2021;2(4): 838–845.
- [5] Yong SJ. Long COVID or post-COVID-19 syndrome: putative pathophysiology, risk factors, and treatments. *J. Infect. Dis*. 2021;53(10):737–754.
- [6] Kabir MA, Ahmed R, Iqbal S, Chowdhury R, Paulmurugan R, Demirci U, et al. Diagnosis for COVID-19: current status and future prospects. *Expert. Rev. Mol. Diagn*. 2021;21(3):269–288.
- [7] Wong R. COVID-19 testing and diagnosis: A comparison of current approaches. *Malays. J. Pathol*. 2021;43(1):3–8.
- [8] Al-Najjar D, Al-Najjar H, Al-Rousan N. Evaluation of the prediction of CoVID-19 recovered and unrecovered cases using symptoms and patient's meta data based on support vector machine, neural network, CHAID and QUEST Models. *Eur. Rev. Med. Pharmacol. Sci*. 2021;25(17):5556–5560.
- [10] Hemmer CJ, Löbermann M, Reisinger EC. COVID-19: Epidemiologie und Mutationen: Ein Update [COVID-19: epidemiology and mutations: An update]. *Radiologe*. 2021;61(10):880–887.
- [11] Wang R, Hozumi Y, Yin C, Wey GW. Mutations on COVID-19 diagnostic targets. *Genom*. 2020;112(6):5204–5213.
- [12] Solís-Arce JS, Warren SS, Meriggi NF, Scacco A, McMurry N, Voors, et al. COVID-19 vaccine acceptance and hesitancy in low- and middle-income countries. *Nat. Med*. 2021;27(8):1385–1394.
- [13] Bose-O'Reilly S, Daanen H, Deering K, Gerrett N, Huynen M, Lee J, et al. COVID-19 and heat waves: New challenges for healthcare systems. *Environ. Res*. 2021;198:111153.
- [14] Centers for Disease Control, 24/7 DBC: We Save Lives. Classifications and definitions of SARS-CoV-2 variants. *C. D. C*. 2021:1-2.
- [15] Cherednik I. Modeling the Waves of Covid-19. *Act. Biotheor*. 2021;70(1):8.
- [16] Lai C, Lam W. Laboratory testing for the diagnosis of COVID-19. *Biochem. Biophys. Res. Commun*. 2021;538:226–230.
- [17] Mohamadian M, Chiti H, Shoghli A, Biglari S, Parsamanesh N, Esmaeilzadeh A. COVID-19: Virology, biology and novel laboratory diagnosis. *J. Gene. Med*. 2021;23(2):e3303.
- [18] Martínez-Chamorro E, Díez-Tascón A, Ibáñez-Sanz L, Ossaba-Vélez S, Borruel-Nacenta S. Radiologic diagnosis of patients with COVID-19. Radiological diagnosis of the patient with COVID-19. *Radiolog*. 2021;63(1):56–73.
- [19] Alizadehsani R, Alizadeh-Sani Z, Behjati M, Roshanzamir Z, Hussain S, Abedini N, et al. Risk factors prediction, clinical outcomes, and mortality in COVID-19 patients. *J. Med. Virol*. 2021;93(4):2307–2320.
- [20] Roland LT, Gurrola JG, Loftus PA, Cheung SW, Chang JL. Smell and taste symptom-based predictive model for COVID-19 diagnosis. *Int. Forum. Allergi. Rhinol*. 2020;10(7):832–838.
- [21] Jehi L, Ji X, Milinovich A, Erzurum S, Rubin BP, Gordon S, Young JB, Kattan MW. Individualizing Risk Prediction for Positive Coronavirus Disease 2019 Testing: Results From 11,672 Patients. *CHEST*. 2020;158(4):1364–1375.
- [22] Zhang SX, Sun S, Afshar-Jahanshahi A, Wang Y, Nazarian-Madavani A, Li J, et al. Beyond Predicting the Number of Infections: Predicting Who is Likely to Be COVID Negative or Positive. *Risk. Manag. Healthc. Policy*. 2020;13:2811–2818.
- [23] Mamidi T, Tran-Nguyen TK, Melvin RL, Worthey EA. Development of An Individualized Risk Prediction Model for COVID-19 Using Electronic Health Record Data. *Fron. Big. Data*. 2021;4:675882.
- [24] Zoabi Y, Deri-Rozov S, Shomron N. Machine learning-based prediction of COVID-19 diagnosis based on symptoms. *N. P. J. Digit. Med*. 2021;4(1):3.
- [25] Zhang J, Jun T, Frank J, Nirenberg S, Kovatch P, Huang KL. Prediction of individual COVID-19 diagnosis using baseline demographics and lab data. *Sci. Rep*. 2021;11(1):13913.
- [26] Challener DW, Challener GJ, Glow-Lee VJ, Fida M, Shah AS, O'Horo JC. Screening for COVID-19: Patient factors predicting positive PCR test. *Infect. Control. Hosp. Epidemiol*. 2020;41(8): 968–969.
- [27] Galindo-Vázquez O, Ramírez-Orozco M, Costas-Muñiz R, Mendoza-Contreras LA, Calderillo-Ruiz G, Meneses-García A. Symptoms of anxiety, depression and self-care behaviors during the COVID-19 pandemic in the general population. *Gac. Med. Mex*. 2021;156(4):298–305.
- [28] Bender R, Grouven U. Ordinal logistic regression in medical research. *Clin. Med. (Lond.)*. 1997;31(5):546–551.
- [29] Iser B, Sliva I, Raymundo VT, Poletto MB, Schuelter-Trevisol F, Bobinski F. Suspected COVID-19 case definition: a narrative review of the most frequent signs and symptoms among confirmed cases. *Epidemiol. Serv. Saude*. 2020;29(3):1-10
- [30] Johansson MA, Quandelacy TM, Kada S, Prasad P. V, Steele M, Brooks JT. SARS-CoV-2 Transmission from People Without COVID-19 Symptoms. *J. A. M. A. NETW. OPEN*. 2021;4(1):e2035057.
- [31] Vila-Muntadas M, Agustí-Sunyer I, Garcia-Navarro A. COVID-19 diagnostic tests: importance of the clinical context. *Med. Clin. (Barc)*. 2021;157(4):185–190.
- [32] Chadeau-Hyam M, Bodinier B, Elliott J, Whitaker MD, Tzoulaki I, Vermeulen R, et al. Risk factors for positive and negative COVID-19 tests: a cautious and in-depth analysis of UK biobank data. *Int. J. Epidemiol*. 2020;49(5):1454–1467.
- [33] Halpin S, O'Connor R, Sivan M. Long COVID and chronic COVID syndromes. *J. Med. Virol*. 2021;93(3):1242–1243.

Individualized prediction of SARS-CoV-2 infection in Mexico City

Municipality during the first six waves of the pandemic

Short title: SARS-CoV-2 individual test prediction

Mariel Victorino- Aguilar,¹ Abel Lerma,² José Humberto Badillo- Alonso,³ Víctor Manuel

Ramos- Lojero,⁴ Claudia Lerma^{5,6*}

¹ Master's Program in Biomedical Sciences, Institute of Health Sciences, Autonomous University of the State of Hidalgo, San Agustín Tlaxiaca 42160, Mexico; biomédica.ingmva@outlook.com, ORCID: 0000-0002-1338-3253

² Area of Psychology, Institute of Health Sciences, Autonomous University of the State of Hidalgo, San Agustín Tlaxiaca 42160, Mexico; abel_lerma@uaeh.edu.mx, ORCID: 0000-0001-7212-641X

³ Jalalpa el Grande Health Center, Mexico City Health Secretariat, 01296, Mexico City, Mexico; humbertobadilloalonso@gmail.com, ORCID: 0000-0001-5746-3944

⁴ Health Jurisdiction Alvaro Obregon, Mexico City Secretary of Health, 01470, Mexico City, Mexico; victorlojero1@gmail.com, ORCID: 0009-0000-2397-2118

⁵ Department of Electromechanical Instrumentation, Instituto Nacional de Cardiología Ignacio Chávez, México City 14080, Mexico; dr.claudialerma@gmail.com, ORCID: 0000-0002-4679-7751

⁶ Centro de Investigación en Ciencias de la Salud (CICSA), FCS, Universidad Anáhuac México Campus Norte, Huixquilucan Edo. de Mexico 52786, Mexico

* Corresponding author

Claudia Lerma, PhD

Instituto Nacional de Cardiología Ignacio Chávez

Juan Badiano 1, Col. Section 16, Mexico City, 14080, Mexico

Email: dr.claudialerma@gmail.com

Acknowledgments

M.V.A. was supported by CONACyT through a scholarship for master's studies.

Abstract

Purpose: After COVID-19, alternative methods to laboratory tests for the prediction of SARS-CoV-2 were developed. The objective of this investigation was to develop and validate a model for individualized prediction of SARS-CoV-2 infection. Our research question was: Can a reliable prediction model of COVID-19 test result be developed with patient data from health centers in Mexico City using a remote data collection method?

Methodology: We included data from patients of the health centers of the Alvaro Obregon Jurisdiction in Mexico City registered in the Epidemiological Surveillance System for Viral Respiratory Diseases (SISVER-SINAVE). Samples were obtained from 36,949 patients for the first phase (model development) and 2957 for the second phase (validation). The variables that were different between a positive test and a negative test were used to generate binary logistic regression models. For validation, the best performance model was tested in the wave 6 data.

Results: The models obtained overall accuracy of 73% (63-83%), sensitivity of 67% (18-86%), and specificity of 74% (59-91%). The wave 5 model had the best performance (accuracy of 83%). In wave 6 data validation, the wave 5 model equation had 100% sensitivity, 0% specificity, and 39% efficacy, while the general model equation had 66% sensitivity, 43% specificity, and 52% efficacy.

Conclusion: The individualized prediction models of a positive COVID-19 test based on SISVER-SINAVE data had good performance in the development phase and lower performance in the validation phase. These prediction models could be applied in early case identification strategies, especially in vulnerable populations.

Keywords: COVID-19; statistical modeling; logistic regression; SISVER; risk factors; symptoms.

Introduction

SARS-CoV-2 is a type of coronavirus that causes COVID-19 disease, which has several symptoms, and in most patients, it is not fatal. However, these symptoms vary from person to person. In Mexico during the first months of the pandemic about 24% of patients needed hospitalization [1,2]. Since the beginning of the COVID-19 pandemic, the world was overwhelmed by the number of infections in very short periods of time [2]. Supplies and health personnel were scarce, especially in the poorest countries, so the need arose to develop methods of detection of the virus, an alternative to existing ones, but at lower cost and accessible to a large part of the population, in order to reduce contagion.

Six waves of the COVID-19 pandemic

Over the last two years, the pandemic has developed through different peaks of infection with different types of SARS-CoV-2 virus variants, and until the beginning of 2023, 6 waves of contagion have been reported in the world [2]. These waves were related to 5 variants of interest, derived from the original strain: alpha, beta, gamma, delta, and omicron (with their sublineages). From these strains, alpha and beta are known to affect humans the most. In Mexico the waves corresponded to those reported worldwide by the WHO, reporting up to November 29, 2022, a total of 7.13 million infections and 330,000 deaths, with almost 2500 new infections confirmed daily [1-4].

Diagnostic tests for SARS-CoV-2 infection

At the beginning of the pandemic, in Mexico, the only diagnostic methods were SARS-CoV-2 PCR tests from the Diagnostic and Epidemiological Reference Institute (InDRE) with 98% efficacy, which were scarce compared to the number of daily infections [1]. In Mexico City, they applied only to patients who came to specialized centers designated to treat COVID-19, and who showed obvious signs of the disease with severe symptoms. Tests were then started in private laboratories but at a high cost of up to 3,500-5,000 Mexican pesos (MMX), and were only performed with a prescription. In early 2021, the Panamerican Health Organization made a first donation of 100,000 rapid tests for SARS-CoV-2 antigens to Mexico, which were first endorsed by InDRE and the Federal Commission for Protection against Health Risks and were 80-97% effective, and as vaccination progressed, began to buy and apply private laboratories and clinics, with a cost of 250-1000 MMX according to the Consumer Protection Federal Agency [5-7].

Since the beginning of the pandemic, testing was only available in large hospitals and COVID centers in cities, and rural populations had virtually no access to them, as well as to treatment for the disease. After more than two years, there are currently many more medical centers with access to testing and treatment at lower costs than two years ago. However, many people still cannot afford to pay for them or must travel long hours and wait long periods of time to get one from the public health system [1,2]. All these complications despite the fact that the disease is still very present in the population, adding hundreds of cases daily around the world, and has become a stationary disease according to recent studies [8].

New methods for COVID-19 detection

Since the emergence and spread of the virus in 2020, several proposals have been published to solve the problem of the lack of testing that led to many more infections. These new methods of individual prediction, mostly positive testing, are focused on not requiring laboratories, and to have access to more information from patients positive for COVID-19. Statistical models were developed by different methods, from mathematical analysis to artificial intelligence [9-16]. The new methods, mostly using similar variables in their prediction models, use known data from patients who have been diagnosed as positive for the virus by laboratory tests. Known data may include symptomatology, comorbid diseases, demographic characteristics, and recent mobility, although some models included data from other laboratory tests obtained from hospitalized patients [9-16].

New methods have been developed to diagnose COVID-19 applying artificial intelligence (AI) (up to 99% accuracy), "immunoassay" and "biosensor-based identification" (66-95% accuracy), mobile devices or smart rings, (up to 99% predicted), "plasma biosensors", and "electromechanical biosensors". Analysis of chest X-rays with artificial intelligence, (98% of hits), use of cough and breath sounds, and "COVID-19 symptom tracker" (98% sensitivity) [16].

SINAVE database application for prediction model in Mexico

In Mexico, diagnostic test data have been collected in the case of the National Epidemiological Surveillance System (SINAVE), which allows access to databases whose fields (i.e., variables) have been changing as more is known about virus variants and disease manifestation. The variables in the SINAVE databases include patient data, the symptoms of the virus, contact with a suspicious case, contact with animals, disease evolution, co-morbid diseases, as well as information from the sampling and the test result [3]. Although epidemiological models have been developed for the prediction of COVID-19 tests or to estimate the risk of hospitalization or mortality, no models have been reported for individualized prediction of SARS-CoV-2 infection for the Mexican population [9-16].

The objective of this work was to develop a model for individual test prediction for COVID-19 based on SINAVE data, for patients who presented for the PCR test at Alvaro Obregon Jurisdiction health centers in Mexico City during the 6 waves of contagion that occurred between 2020 and 2023.

Methods

Study design and participants

The study design is nonexperimental (observational), retrospective (Phase 1), prospective (Phase 2), and analytical (statistical models through binary logistic regression).

The study was conducted with data from patients who attended the health centers of the Alvaro Obregon Jurisdiction of Mexico City between April 2020 and July 2022 for Phase 1 and November 2022 to January 2023 for Phase 2. All patients were selected within 3 weeks at the peak of each of the waves according to data reported in SINAVE 3 (Table 1). We included patients who requested the CRP test due to suspicion of COVID-19 and who were older than 18 years. Patients with inconclusive results in their PCR test were excluded. The research protocol was approved by the Ethics and Research Committee of the Universidad Autónoma del Estado de Hidalgo (protocol number Comiteei.icsa 017/2022 /2022). All procedures adhered to the principles stated in the Declaration of Helsinki.

Procedure for obtaining study data and variables

Data were obtained from the SISVER/SINAVE database of the Ministry of Health of Mexico, which contains the information variables of those patients who came to the health centers of the Alvaro Obregon Jurisdiction for suspicion of infection with the COVID virus-19 [3]. These data were collected by the health professionals attending to the patients and were recorded in the system with an identifier number per patient.

The samples of each of the patients were taken in the COVID centers specially designated for this type of care, by means of pharyngeal and nasopharyngeal exudates; they were sent to InDRE and the results were subsequently recorded in the SINAVE database. The results were recorded as a field in the database named "Final result" and were classified as: "SARS-Cov-2", "Negative", "Rejected", "Unsuitable", "Non-subscribed", "Positive" and empty fields. The results classified as "SARS-Cov-2" and "Positive" were considered positive tests and recoded as 1, while the results classified as "Negative" were considered negative tests and recoded as 0 for further analysis. The results classified with other criteria ("Rejected", "Unsuitable", "Non-subscribed", or empty fields) were removed as they were not suitable for analysis [3].

The variables included in the regression analysis were selected from a literature review regarding the main characteristics presented by patients diagnosed as positive within each wave, as it was shown that some of these characteristics changed with respect to the SARS-Cov-2 virus subtypes, so they were included in a general way for the models since the database did not record information on the variant of the virus with which the positive patient was infected [3, 9-19].

The following variables were included: sex, age, main comorbidities for the Mexican population (diabetes, hypertension, and obesity [9]), smoker, contact with a suspicious case, and presence of symptoms (fever, cough, odynophagy, dyspnea, irritability, diarrhea, chest pain, chills, headache, myalgia, arthralgia, rhinorrhea, polypnea, vomit, abdominal pain, conjunctivitis, cyanosis, anosmia, dysgeusia). All variables except age and sex were classified as "YES" and "NO" and were recoded in 1 and 0 respectively. Cases with captured data such as "IGNORED" and empty fields were excluded [3].

The SISVER/SINAVE database was consulted to obtain all the cases registered for the health centers of Alvaro Obregon Jurisdiction during each peak of the first 6 waves of the pandemic (Table 1). The dates with the highest number of cases of each wave were identified. For waves 2, 3, 4, 5 and 6, an interval of one week before and one week after the week with more cases was selected (total sampling period of 3 weeks per wave). During wave 1, there was little availability of evidence and the criteria for defining a suspicious case

were stricter, resulting in a low rate of negative cases. Therefore, for the wave 1 sample, an interval of two weeks before and two weeks after the week with more cases was selected (total sampling period of 5 weeks).

Statistical analysis

Age had no normal distribution (Kolmogorov-Smirnov test, $p < 0.01$), so it is described as median (25th percentile - 75th percentile) and was compared between waves and between positive or negative results of each wave by U of Mann-Whitney test. The nominal variables are described as absolute value and percentage and were compared between waves and between test results (positive or negative) with Chi-square test.

For the prediction of the SARS-CoV-2 test result binary logistic regression analyses were performed, obtaining a regression model for each of the 6 waves and a more general model (for all waves). Only the variables that showed significance between the positive and negative groups of each wave were included. We used the conditional forward method of binary logistic regression with the goodness of Hosmer-Lemeshow adjustment. For the significant variables in each model, the odds ratio (OR) with a 95% confidence interval (CI) is reported. The statistical analysis was performed with SPSS version 21.0. A value of $p < 0.01$ was considered significant.

Validation phase

The general equation for calculating the value of $P(Y)$ which is the probability that Y took the value of 1 or the presence of the property, which in this case would be the probability that a case took the positive individual test value by COVID-19, in this case, data obtained from the SPSS program were taken for the models made, and subsequently applied to the databases of each of the waves to check the models. The equation is as follows:

$$\text{logit}[p(x)] = \log \left[\frac{p(x)}{1 - p(x)} \right] = a + b^1x^1 + b^2x^2 + b^3x^3 \dots \quad \text{Equation 1}$$

Simplifying Equation 1, it takes the following form:

$$p = \frac{\exp(a + b^1x^1 + b^2x^2 + b^3x^3 \dots)}{1 + \exp(a + b^1x^1 + b^2x^2 + b^3x^3 \dots)} \quad \text{Equation 2}$$

Where:

p: probability of the presence of the property

e: Euler number=2.71828 approx.

a: value of the constant

bn: B value of the variable

x^n : set of covariables

Results

Table 2 shows the characteristics of the study participants. The asterisk symbols (*) show that wave 1 had greater age, a higher percentage of women, and more proportion of smokers, hypertension, obesity, diabetes, and contact with a suspicious case compared to subsequent waves (2 to 5). Compared to wave 2 (symbol &), the age in subsequent waves 3 and 4 was lower and in wave 5 was greater. Wave 2 also had a higher proportion of smokers (compared to waves 3 and 5), a higher proportion of hypertension and diabetes (compared to waves 3 and 4), a lower proportion of obesity (compared to wave 4), and less contact with a suspicious cases (in the subsequent 3 waves 3, 4 and 5). Wave 2 had the lowest proportion of comorbidities (hypertension, obesity, diabetes) and the lowest proportion of contact with a suspicious case. Compared to wave 3 (symbol #), waves 4 and 5 had significant differences in age, number of smokers, proportion of patients with hypertension, obesity (only in wave 4), diabetes (only in wave 5), and contact with a suspicious case (only in wave 5). Wave 6 (symbol ≠) showed statistically significant differences with the other waves in all variables.

Table 3 shows the sociodemographic and morbidity data of the participants compared with the test result (positive or negative), the asterisks indicate statistically significant differences between positive and negative within the same wave ($p < 0.01$). In wave 1, the different variables between positive and negative

(indicated with *) are: age, sex, hypertension, and diabetes. In wave 2 the different variables between positive and negative are: age, smoker, hypertension, obesity, diabetes, and contact with a suspicious case. In wave 3 the positive ones were different from the negative ones in age, sex, obesity, diabetes, and contact with a suspicious case. In wave 4 the positive cases were different from the negative ones in: hypertension, obesity and contact with a suspicious case, while in wave 5 the variables with difference between positive and negative were: age, smoker, hypertension, obesity, diabetes, and contact with a suspicious case. Wave 6 showed differences between age, diabetes mellitus and contact with a suspicious case.

Table 4 compares the percentage of presence of symptoms in patients with positive vs negative test within each wave and in general. In wave 1, the percentage of symptoms among positive and negative patients was different in most variables: cough, rhinorrhea, dyspnea, polypnea, anosmia, dysgeusia, headache, myalgia, arthralgia, chest pain, fever and chills. In wave 2 and 3, all variables presented statistically significant differences between positive and negative, which were: cough, rhinorrhea, dyspnea, polypnea, cyanosis, odynophagy, anosmia, dysgeusia, diarrhea, abdominal pain, vomit, headache, myalgia, arthralgia, chest pain, fever, chills, irritability, and conjunctivitis. Wave 4 presented differences in the following variables: cough, rhinorrhea, odynophagy, diarrhea, abdominal pain, headache, myalgia, arthralgia, chest pain, fever, chills, and irritability. Wave 5 presented differences in: cough, rhinorrhea, dyspnea, polypnea, odynophagy, dysgeusia, diarrhea, abdominal pain, headache, myalgia, arthralgia, chest pain, fever, chills, irritability and conjunctivitis. The general prevalence showed differences in all variables: cough, rhinorrhea, dyspnea, polypnea, cyanosis, odynophagy, anosmia, dysgeusia, diarrhea, abdominal pain, vomit, headache, myalgia, arthralgia, chest pain, fever, chills, irritability, and conjunctivitis. In wave 6, all variables showed statistically significant differences except polypnea, cyanosis, and abdominal pain.

Table 5 presents the results of the logistic regression analysis with models to predict the outcome of the SARS-CoV-2 test (positive or negative) as a dependent variable. In each wave, a different model was calculated that included as independent variables those that had significant differences between positive and negative (Tables 3 and 4). For example, the initial wave 1 model included age, female sex, hypertension, diabetes, cough, rhinorrhea, dyspnea, polypnea, anosmia, dysgeusia, headache, myalgia, arthritis, chest pain, fever, and chills. The variables that were not initially included in the wave 1 model, indicated as "Not included", are: smoker, obesity, contact with a suspicious case, cyanosis, odynophagy, diarrhea, abdominal pain, vomit, irritability and conjunctivitis. When applying the logistic regression analysis in model 1 with the later conditional method, the variables that did not have a significant contribution to the dependent variable, indicated as "Eliminated", were eliminated in the age, hypertension, rhinorrhea, polypnea, anosmia, myalgia, arthralgia, chest pain and chills. The final wave 1 model had the following variables with a significant odds ratio ($p < 0.05$): female sex, diabetes, cough, dyspnea, dysgeusia, headache, and fever.

Applying the same logistic regression analysis strategy, the final models of subsequent waves had different predictor variables (Table 5). The final wave 2 model had as predictors to age, smoker, obesity, contact with a suspicious case, cough, rhinorrhea, dyspnea, polypnea, odynophagy, anosmia, dysgeusia, vomit, headache, arthralgia, chest pain, fever, chills, irritability, and conjunctivitis. The final wave 3 model had as predictors age, female sex, cough, rhinorrhea, cyanosis, odynophagy, anosmia, dysgeusia, diarrhea, abdominal pain, headache, myalgia, arthralgia, fever, chills, and irritability. The final wave 4 model had as final predictor variables cough, odynophagy, myalgia, fever, chills, and irritability. The final wave 5 model had age, cough, rhinorrhea, headache, fever, and chills as predictors.

With the same regression analysis strategy, a general model was calculated with the 5 wave data (Table 5). The final general model had the following variables: *female* sex, hypertension, diabetes, cough, rhinorrhea, dyspnea, polypnea, cyanosis, odynophagy, anosmia, dysgeusia, diarrhea, abdominal pain, headache, myalgia, arthralgia, chest pain, fever, chills and irritability. Wave 6 included age variables, contact with a suspicious case, cough, dyspnea, vomit, arthralgia and conjunctivitis, excluding or eliminating the other variables.

Table 6 shows the performance results of the models. The determination coefficients (R^2) of the models ranged from 0.049 to 0.537. Global predictions also varied according to the wave, with an average overall accuracy of 73% (range between 63% and 83.1%), with average sensitivity of 67% (range between 18 and 86%), average specificity of 74% (range between 66 and 91%), average positive predictive value of 62% (ranges between 34 and 74%) and average negative predictive value of 77% (range between 58 and 91%).

Phase 2: Verification of prediction models

For the verification phase of the model, the descriptive analyses of the wave 6 database were performed (Tables 2-4), with a total of 3332 patients, to whom the elimination criteria were applied, leaving a final total of 3957 patients.

Supplementary material Online resource 1 shows the values obtained by Equation 2 applied to the variables corresponding to each model in the databases. True negatives (VN), false negatives (FN), false positives (FP) and true negatives (VN) were calculated. Then again, a total count of cases was made since Excel does not make adjustments in the case of lost values as it does SPSS, to correctly calculate the percentages obtained, it was determined the full N, the N of classifieds, total negatives and total positives of each model. Subsequently, the percentages of the correct negative, correct positive, the correct overall percentage, sensitivity, specificity, positive predictive value and negative predictive value were calculated. Table 7 shows the performance obtained in each of the models developed to predict an individual COVID-19 test result. The wave 5 model was the best overall percentage obtained with 84%, and for wave 6 of the 2 models applied (wave 5 and general). When applying the equation to the models it was found that they generally increased their values positively for the test predictions.

The wave 5 model applied to wave 6 data obtained a 100% positive correct prediction percentage, 0% negative correct prediction, and an overall accuracy percentage of 39%, while the general wave model obtained a 66% positive correct prediction percentage, 43% in negative correct prediction, and an overall success percentage of 52%. The cut-off value established for the binary logistic regression equation at >0.5 was taken into account as a positive prediction.

Discussion

The main contribution of this work is to demonstrate the development of statistical models for individual prediction of SARS-CoV-2 virus infection from sociodemographic data and symptoms usually collected in the epidemiological surveillance system of Mexico. As a test population, the work focused on a sample by the census of all the cases registered in the Alvaro Obregon Jurisdiction, one of the most populated and with great socioeconomic heterogeneity, in one of the most populous cities in the world, Mexico City. This municipality has remained one of the most affected since the beginning of the pandemic and has even been the epicenter of cases of COVID-19 infection in Mexico. The prediction models showed very good performance so they are a potential new tool to deal with potential new waves of contagion, promoting an early diagnosis of risk and thus favoring measures that reduce infections.

Some similar models have already been developed in different parts of the world, mainly in the United States. A previous work was based on variables and methods similar to the present work in California [9], with a model based on symptoms of smell and taste, which also includes morbidities and the symptomatology most frequently reported until then with that variable in the year of 2020 and also applied logistic regression models with a correct prediction of 75%. Another previous work [10], also determined the probability of having the disease or not from surveys and risk factors. Another model developed in Alabama [14] used a risk calculator from an electronic recording of the most reported risk factors, such as the symptoms already mentioned, smokers, and psychological factors. On the other hand, in Minnesota they developed another model that uses the same variables plus patient mobility, applying multivariate logistic regression, showing the predictor variables with the highest probability of being positive patients.

Most prediction models for COVID-19 tests are applied using logistic, binary, or multivariable regression [9-16], whose results raise percentages in general of between 70 and 80% of correct prediction. However, there are also other models applied to test prediction for such a disease. Some apply artificial intelligence or AI, to predict the development of the pandemic or make models of individualized prediction. An example was developed in Israel, in which they used basic information and symptomatology of patients as dichotomous variables, plus contact with positive cases, achieving an accuracy reported as high [11]. Some of the latest models apply electronic devices such as smartwatches or even new materials to measure biosignals that will be analyzed [16].

In each of the waves, different variables were used to make the final prediction model, which was previously determined from the significant differences between the positive and negative cases of each of the waves. As already mentioned, at each of the peaks of each wave different variants of the SARS-CoV-2 were presented, so the variables in the models changed depending on these characteristics. In addition, after the first wave, vaccination schemes began in older people, and as the pandemic progressed, they continued to be vaccinated by age group to reach the youngest. This could influence changes in the behavior of the variables of the models as the pandemic progressed, being that in the fifth wave raised again the age, with the hypothesis that this sector of over 50 years was the first to be vaccinated and by this time, adults over 60 were

already receiving the vaccination scheme again [9,10]. Currently, COVID-19 testing has become more accessible to the population, both in the public and private sectors, which influenced an increase of negative tests in the databases [8]. In this study and in the literature, the determinant variables agree with those previously reported in other populations [1-3, 17, 19-25]. It should be noted that no models had been made with the 5 waves separately and in general, so there is no comparison of present work with previous reports.

It can be observed, for example, that certain variables were not included in the model from the significance and then others were excluded during the regression analysis. Age was one of the variables included in wave 2, 3, and 5 by vaccination schemes, but not in the 1, 4, and the general model where either there was no vaccination, or the first block became the one that had the longest with the scheme. Sex was the only variable predictor in wave 1 and the general model, being a smoker only in wave 2, while morbidities such as hypertension only in the general model, obesity in wave 2, diabetes in wave 1 and the general model, while contact with a suspicious case was a predictor only on wave 2.

Regarding symptomatology, cough was present in all waves as a predictor variable, dyspnea was included in waves 1, 2 and the general model, polypnea in wave 1 and general, cyanosis in wave 3 and general, odynophagy and anosmia in wave 2, 3, and general dysgeusia in all except 4 and 5, diarrhea and abdominal pain in wave 3 and general. Vomit only occurred in wave 2 and was the only one along with conjunctivitis that was removed from the general model. Headache occurred in all except wave 4, myalgia in waves 3, 4, and general, arthralgia in waves 2, 3, and general, chest pain in wave 2 and general. Fever was present in all waves, chills only did not occur in wave 1, irritability was in waves 2, 3, 4, and general, and conjunctivitis only in wave 2. As can be seen, the gastrointestinal symptoms agreed with the appearance of Delta. Fever and cough were in all waves, most variables were present in at least 3 of the waves and others were only excluded in one, such as headache and chills, while it can be concluded that the first waves were those that showed the largest number of variables included, along with the general model that included all except two.

The validation phase obtained relatively low percentages of prediction, especially in the wave 5 model, which although it obtained 100% sensitivity, the specificity fell to 0%, obtaining a 39% of overall accuracy. Although the general model in the verification had better results, it shows that the populations seem to be quite different from those of the previous waves, mentioning, for example, the change in the percentages of positive and negative patients, and that no account is taken of the vaccination variable that appeared from little until wave 4 onwards.

The results of this work show that it is possible to create new methods for predicting COVID-19 tests, adapting them to the characteristics of the study population, whereas infections will increase with new waves in the future and studies suggest that it will be a stationary disease. The virus continues to change, and today more than ever it is vital that people have access to alternative methods to laboratory methods, which are also faster. The epidemiological implications arising from this research work are the possible reduction of infections in the Mexican population, by providing new tools to take care of their health, faster and cheaper. This also opens the doors for further development and investigation of such methods applied to the Mexican population. As mentioned, each population has its specific characteristics and the variables considered for the prediction models change between population and within the same population over time. Among the clinical considerations of the results, we can see the possible relief of the medical services that continue to provide services to COVID-19 patients, and that despite vaccination schemes have helped the population to face the pandemic, can not ignore the fact that the virus of SARS-CoV-2 came to stay as an endemic disease [8].

Among the limitations of the study are the variables mentioned above, the fact that during the first wave, only patients who had severe symptoms were tested, so the test results have few negatives to include in the model. Another major limitation was that recent travel records were not included in the database, since this variable was considered as a potential predictor at the beginning of the study and according to the literature review. Also, a limitation of the study is the lack of information on the virus variant from which patients are infected, since the test result is only classified as positive or negative. Therefore, it is not possible to classify the variables corresponding to each variant of the SARS-CoV-2 virus. However, the variable not included that is probably the most important is the vaccination scheme. Since vaccination began, many variables changed at the same time, such as symptomatology, the number of patients who came to the health centers to be tested, and the percentages of positive and negative patients. Vaccination-related information was not taken into account in the present work, as it was included in the SINAVE databases only recently and no data were available for most waves. The models applied in the data available in the verification phase produced low percentages of sensitivity and overall prediction. Future studies are needed to identify sources of variation of data and prediction results between waves.

Conclusions

In the present work, individual test prediction models were developed in a sample of Mexico City that showed a good percentage of global correct prediction. Within the verification phase, the overall model obtained the best prediction results applied to wave 6 data, with 66% sensitivity and 52% overall accuracy. The models developed here could be very useful for the health care of the Mexican population, and because they are based on a simple methodology and data publicly available through the country's health system, these models are likely to be possible with more sophisticated methods such as artificial intelligence to raise sensitivity and specificity.

Ethical considerations

The data on the bases of the Epidemiological Surveillance System are of public consultation.³ There was strict adherence to the Federal Data Protection Law, the National Institute of Access to Information, and the Regulations of the General Health Law on Health Research in force in Mexico, for the consultation of the surveillance sheets of the health centers of the Alvaro Obregon Jurisdiction. The data used does not violate the data protection of the registered persons, since only the symptoms, demographics, and folio number were used, which were not and will not be exposed individually [26-29].

Declaration

Observational studies (STROBE)

Conflicts of interest

The authors declare that they have no conflict of interest.

Competing interests and funding

The authors declare that they have no conflict of competing interests and funding. No funding was received for the study.

Table 1. Characteristics of the data for each of the 6 waves of the COVID-19 pandemic.

	<i>Wave 1</i>	<i>Wave 2</i>	<i>Wave 3</i>	<i>Wave 4</i>	<i>Wave 5</i>	<i>Wave 6</i>
<i>Date range</i>	19 April-21 June 2020	4- 24 January 2021	19 July- 8 August 2021	3- 23 January 2022	20 June- 10 July 2022	27 November 2022- 14 January 2023
<i>Main variant</i>	Original strain	Alpha, Beta y Gama	Delta	Omicron	Omicron (sublineages)	Omicron (sublineages)
<i>Total patients</i>	2081	13535	15368	6159	1177	3332
<i>Elimination criteria</i>	71	1196	2072	822	167	375
<i>No. Final patients</i>	2010	12339	13296	5337	1010	2957

Table 2. Sociodemographic characteristics and morbidities of the study population. Data are shown as count and percentages or median (25th percentile - 75th percentile).

	<i>Wave 1</i>	<i>Wave 2</i>	<i>Wave 3</i>	<i>Wave 4</i>	<i>Wave 5</i>	<i>Wave 6</i>
<i>Variable</i>	(N = 2009)	(N = 12339)	(N = 13296)	(N = 5337)	(N = 1010)	(N=2957)
<i>Age (years)</i>	46 (35 – 57)	40 * (29 – 52)	35 * $\&$ (26 – 48)	38 * $\&$ # (29 – 51)	43 * $\&$ # $\$$ (31 – 58)	43 * $\&$ # $\$$ \neq (30 – 55)
<i>Sex</i>		*	*	*	* $\&$ $\$$	* $\&$ # $\$$
<i>Female</i>	1006 (50%)	6668 (54%)	7306 (55%)	2854 (53%)	590 (58%)	
<i>Male</i>	1003 (50%)	5671 (46%)	5990 (45%)	2483 (47%)	420 (42%)	1754 (59.3%) 1203 (40.7%)
<i>Smoker</i>	240 (12%)	1234 (10%)	1036 (8%) * $\&$	560 (11%) #	41 (4%) * $\&$ # $\$$	180 (6.1%) * $\&$ # $\$$ \neq
<i>Hypertension</i>	367 (18%)	1124 (9%) *	818 (6%) * $\&$	267 (5%) * $\&$ #	112 (11%) * $\&$ # $\$$	255 (8.6%) * $\&$ # $\$$ \neq
<i>Obesity</i>	355 (18%)	672 (5%) *	721 (5%) *	110 (2%) * $\&$ #	41 (4%) * $\$$	180 (6.1%) * $\&$ # $\$$ \neq
<i>Diabetes mellitus</i>	301 (15%)	941 (8%) *	606 (5%) * $\&$	212 (4%) * $\&$	73 (7%) * $\&$ # $\$$	176 (6.0%) * $\&$ # $\$$
<i>Contact with a suspicious case</i>	833 (62%)	7202 (59%)	5198 (39%) * $\&$	2125 (41%) * $\&$	236 (24%) * $\&$ # $\$$	615 (20.8%) * $\&$ # $\$$ \neq

* $p < 0.01$ vs wave 1, $\&$ $p < 0.01$ vs wave 2, # $p < 0.01$ vs wave 3, $\$$ $p < 0.01$ vs wave 4, \neq $p < 0.01$ vs wave 5.

Table 3. Sociodemographic characteristics and morbidities of the study population (positive and negative for each wave). Data are shown as count and percentages or median (25th percentile - 75th percentile).

<i>Variable</i>	<i>Wave 1</i>		<i>Wave 2</i>		<i>Wave 3</i>		<i>Wave 4</i>		<i>Wave 5</i>		<i>Wave 6</i>	
	Positive (N = 1119)	Negative (N = 874)	Positive (N = 6372)	Negative (N = 5363)	Positive (N = 5435)	Negative (N = 7658)	Positive (N = 1150)	Negative (N = 3229)	Positive (N = 298)	Negative (N = 616)	Positive (N = 961)	Negative (N = 1532)
<i>Age (years)</i>	48 (37 – 60)	43 * (34 - 55)	40 (30 – 52)	39 * (28 – 51)	34 (25 – 46)	36 * (26 – 49)	39 (29 – 51)	38 (29 – 50)	49 (33 – 66)	40 * (28.5 – 55)	45 (33-58)	40* (30-55)
<i>Sex</i>		*				*						
<i>Female</i>	45.0%	56.4%	54.2%	53.8%	52.5%	56.8%	53.0%	53.7%	57.4%	59.3%	59.9%	59.7%
<i>Male</i>	55.0%	43.6%	45.8%	46.2%	47.5%	43.2%	47.0%	46.3%	42.6%	40.7%	40.1%	40.3%
<i>Smoker</i>	11.9%	12.2%	11.2%	8.5% *	8.4%	7.4%	11.6%	11.0%	7.1%	2.3% *	5.8% %	6.3%
<i>Hypertension</i>	21.6%	14.4% *	10.3%	7.5% *	6.5%	5.8%	7.0%	4.9% *	20.3%	5.5% *	9.4%	8.7%
<i>Obesity</i>	19.0%	16.1%	7.1%	3.1% *	6.6%	4.6% *	3.6%	2.0% *	7.8%	2.3% *	5.8%	6.1%
<i>Diabetes mellitus</i>	20.0%	8.9% *	8.6%	6.3% *	5.3%	3.9% *	5.3%	3.9%	11.9%	3.9% *	5.8%	6.9% *
<i>Contact with a suspicious case</i>	58.0%	65.0%	54.7%	62.7% *	42.5%	37.1% *	44.3%	37.1% *	37.5%	18.5% *	25.3%	19.9% *

* p < 0.01 compared with positive of the same wave.

Table 4. Prevalence of symptoms (positive and negative for each wave and general).

Variable	Wave 1		Wave 2		Wave 3		Wave 4		Wave 5		General		Wave 6	
	Positive (N = 1119)	Negative (N = 874)	Positive (N = 6372)	Negative (N = 5363)	Positive (N = 5435)	Negative (N = 7658)	Positive (N = 1150)	Negative (N = 3229)	Positive (N = 298)	Negative (N = 616)	Positive (N = 14374)	Negative (N = 17740)	Positive (N = 961)	Negative (N = 1532)
<i>Cough</i>	80.8	61.8 *	51.4	22.4 *	62.5	81.1 *	19.3	21.3 *	48.7	14.8 *	58.3	22.4 *	60.3	53.9*
<i>Rhinorrhoea</i>	34.4	27.0 *	27.8	9.0 *	26.7	37.4 *	5.4	6.5 *	28.3	11.0 *	28.1	9.0 *	41.3	35.7*
<i>Dyspnea</i>	46.7	26.2 *	9.9	5.1 *	8.5	20.9 *	4.5	3.9	2.5	1.7 *	11.9	5.0 *	9.8	14.6*
<i>Polypnea</i>	21.4	12.1 *	3.2	1.6 *	3.1	2.7 *	0.5	1.1	0.1	0.1 *	4.3	1.6 *	3.2	4.9
<i>Cyanosis</i>	6.9	4.2	1.9	0.7 *	1.4	0.7 *	0.2	0.5	0.1	0.1	1.9	0.7 *	0.9	1.9
<i>Odynophagy</i>	46.8	41.4	28.6	11.2 *	29.0	37.2 *	7.6	6.9 *	37.6	11.4 *	30.1	10.7 *	41.2	39.2*
<i>Anosmia</i>	17.1	8.9 *	19.0	4.8 *	14.3	2.1 *	0.7	1.6	2.0	1.1	15.4	2.7 *	4.1	2.5*
<i>Dysgeusia</i>	17.9	8.1 *	16.8	4.6 *	12.5	2.8 *	0.5	1.6	1.7	0.9 *	13.7	2.6 *	3.4	2.5*
<i>Diarrhea</i>	23.6	20.9	9.6	4.6 *	11.3	10.4 *	4.7	6.7 *	5.7	2.8 *	11.1	6.0 *	10.9	9.9*
<i>Abdominal pain</i>	13.9	12.3	5.3	1.8 *	7.3	17.2 *	3.4	3.2 *	3.3	1.5 *	6.8	2.9 *	5.9	7.3
<i>Vomit</i>	9.9	7.1	3.3	1.4 *	4.3	5.4 *	1.9	1.5	2.0	1.1	4.1	1.7 *	3.0	6.3*
<i>Headache</i>	75.2	70.0 *	44.4	19.7 *	45.6	65.0 *	17.0	15.5 *	32.0	12.3 *	46.7	18.9 *	48.1	46.2*
<i>Myalgia</i>	59.7	43.8 *	25.3	8.2 *	26.2	48.1 *	10.7	4.7 *	22.5	6.4 *	28.5	8.2 *	34.1	31.3*
<i>Arthralgia</i>	57.7	41.3 *	21.4	6.2 *	20.8	30.0 *	5.9	3.6 *	13.9	3.6 *	23.4	6.3 *	27.5	23.6*
<i>Chest pain</i>	31.4	22.0 *	14.3	6.0 *	15.4	30.3 *	8.6	6.2 *	7.0	2.5 *	15.8	6.3 *	16.8	16.9*
<i>Fever</i>	71.8	45.2 *	33.4	13.0 *	42.2	52.7 *	9.3	10.6 *	18.3	4.2 *	38.9	11.8 *	35.3	31.3*
<i>Chills</i>	43.8	29.1 *	21.8	6.5 *	27.5	40.7 *	6.7	7.2 *	22.0	5.7 *	26.0	7.7 *	28.6	26.6*
<i>Irritability</i>	19.4	16.0	8.7	5.1 *	10.1	13.1 *	5.4	5.9 *	2.9	0.7 *	9.7	5.2 *	11.6	11.9*
<i>Conjunctivitis</i>	10.9	11.3	7.3	2.2 *	8.5	15.5 *	2.3	2.7	1.6	1.1 *	7.8	2.7 *	7.8	10.8*

* p < 0.01 compared with positive of its group.

Table 5. Binary logistic regression analysis to predict an individual SARS-COV-2 test for each wave and a general model (all waves). Data are presented as odds ratio (95% confidence interval).

<i>Variable</i>	<i>Wave 1</i>	<i>Wave 2</i>	<i>Wave 3</i>	<i>Wave 4</i>	<i>Wave 5</i>	<i>General</i>	<i>Wave 6</i>
<i>Age (years)</i>	Eliminated	1.010 (1.007 – 1.013)	1.005 (1.002 – 1.008)	Not included	1.022 (1.011 – 1.033)	Not included	1.014(1.008 – 1.019)
<i>Female</i>	0.695 (0.542 – 0.891)	Not included	0.749 (0.687 – 0.817)	Not included	Not included	0.818 (0.776 – 0.863)	Not included
<i>Smoker</i>	Not included	1.173 (1.017 – 1.353)	Not included	Not included	Eliminated	Eliminated	Not included
<i>Hypertension</i>	Eliminated	Eliminated	Not included	Eliminated	Eliminated	1.184 (1.066 – 1.316)	Not included
<i>Obesity</i>	Not included	1.475 (1.199 – 1.815)	Eliminated	Eliminated	Eliminated	Eliminated	Not included
<i>Diabetes mellitus</i>	1.761 (1.199 – 2.585)	Eliminated	Eliminated	Not included	Eliminated	1.312 (1.166 – 1.475)	Eliminated
<i>Contact with a suspicious case</i>	Not included	0.718 (0.660 – 0.782)	Eliminated	Eliminated	Eliminated	Eliminated	1.315(1.077 – 1.606)
<i>Cough</i>	1.725 (1.280 – 2.324)	2.045 (1.853 – 2.256)	3.155 (2.875 – 3.464)	2.970 (2.479 – 3.558)	4.558 (2.882 – 7.209)	2.455 (2.311 – 2.608)	1.480(1.229 – 1.782)
<i>Rhinorrhea</i>	Eliminated	2.210 (1.951 – 2.503)	2.057 (1.799 – 2.352)	Eliminated	3.658 (2.211 – 6.052)	1.888 (1.747 – 2.041)	Eliminated
<i>Dyspnea</i>	1.548 (1.161 – 2.063)	0.806 (0.669 – 0.972)	Eliminated	Not included	Eliminated	0.865 (0.767 – 0.976)	0.530(0.399 – 0.704)
<i>Polypnea</i>	Eliminated	0.565 (0.403 – 0.792)	Eliminated	Not included	Not included	0.727 (0.588 – 0.899)	Not included
<i>Cyanosis</i>	Not included	Eliminated	0.529 (0.316 – 0.886)	Not included	Not included	0.649 (0.473 – 0.889)	Not included
<i>Odynophagy</i>	Not included	1.498 (1.328 – 1.689)	2.267 (1.987 – 2.586)	2.231 (1.841 –	Eliminated	1.627 (1.508 –	Eliminated

				2.704)		1.756)	
<i>Anosmia</i>	Eliminated	2.528 (2.034 – 3.142)	3.805 (2.825 – 5.124)	Not included	Not included	2.964 (2.525 – 3.478)	Eliminated
<i>Dysgeusia</i>	2.259 (1.520 – 3.359)	1.413 (1.127 – 1.771)	1.709 (1.255 – 2.327)	Not included	Not included	1.663 (1.407 – 1.965)	Eliminated
<i>Diarrhea</i>	Not included	Eliminated	0.725 (0.616 – 0.854)	Eliminated	Not included	0.726 (0.650 – 0.811)	Eliminated
<i>Abdominal pain</i>	Not included	Eliminated	0.777 (0.626 – 0.963)	Eliminated	Eliminated	0.826 (0.714 – 0.957)	Not included
<i>Vomit</i>	Not included	0.582 (0.413 – 0.820)	Eliminated	Not included	Not included	Eliminated	0.412 (0.249 – 0.682)
<i>Headache</i>	0.717 (0.528 – 0.966)	1.590 (1.434 – 1.762)	1.700 (1.531 – 1.887)	Eliminado	1.964 (1.283 – 3.007)	1.531 (1.433 – 1.636)	Eliminated
<i>Myalgia</i>	Eliminated	Eliminated	1.865 (1.570 – 2.215)	1.292 (1.014 – 1.647)	Eliminated	1.311 (1.185 – 1.449)	Eliminated
<i>Arthralgia</i>	Eliminated	1.552 (1.334 – 1.806)	1.387 (1.144 – 1.682)	Eliminated	Eliminated	1.285 (1.151 – 1.434)	1.241(1.012 – 1.521)
<i>Chest pain</i>	Eliminated	1.271 (1.079 – 1.497)	Eliminated	Eliminated	Eliminated	1.144 (1.034 – 1.266)	Eliminated
<i>Fever</i>	2.060 (1.575 – 2.695)	1.864 (1.664 – 2.089)	2.978 (2.670 – 3.322)	1.631 (1.240 – 2.146)	3.654 (2.340 – 5.704)	2.393 (2.229 – 2.570)	Eliminated
<i>Chills</i>	Eliminated	1.485 (1.275 – 1.729)	1.668 (1.454 – 1.913)	1.518 (1.176 – 1.959)	1.870 (1.134 – 3.083)	1.492 (1.365 – 1.632)	Eliminated
<i>Irritability</i>	Not included	0.760 (0.628 – 0.920)	0.802 (0.676 – 0.950)	2.005 (1.089 – 3.691)	Eliminated	0.857 (0.761 – 0.965)	Eliminated
<i>Conjunctivitis</i>	Not included	1.309 (1.025 – 1.670)	Eliminated	Not included	Eliminated	Eliminated	0.705(0.513 – 0.969)

Not included= variable not included in regression, Deleted: Variable eliminated by regression.

Table 6. Summary of the performance of SARS-Cov-2 individual test prediction models.

<i>Variable</i>	<i>Wave 1</i>	<i>Wave 2</i>	<i>Wave 3</i>	<i>Wave 4</i>	<i>Wave 5</i>	<i>General</i>	<i>Wave 6</i>
<i>R²</i>	0.151	0.267	0.405	0.201	0.537	0.313	0.049
<i>Overall prediction (%)</i>	66.2%	71.1%	76.9%	76.6%	83.1%	73.2%	62.5%
<i>Sensitivity</i>	65.8%	67.4%	77.4%	79.6%	86.4%	73.0%	18.1%
<i>Specificity</i>	66.4%	74.4%	75.8%	59.3%	74.9%	73.2%	91.1%
<i>Positive predictive value</i>	73.7%	71.0%	63.7%	34.1%	71.0%	60.8%	56.9%
<i>Negative predictive value</i>	57.7%	71.1%	86.0%	91.7%	88.9%	82.8%	63.2%

Table 7. Summary of performance obtained in each of the models developed to predict individual COVID-19 test results.

	<i>Model</i>	<i>Negative correct porcentaje</i>	<i>Positive correct porcentaje</i>	<i>Global correct porcentaje</i>	<i>Sensitivity</i>	<i>Specificity</i>	<i>Positive predictive value</i>	<i>Negative predictive value</i>
<i>Development</i>	<i>Wave 1</i>	54%	78%	67%	78%	54%	69%	65%
	<i>Wave 2</i>	71%	71%	71%	71%	71%	74%	67%
	<i>Wave 3</i>	86%	64%	77%	64%	86%	76%	77%
	<i>Wave 4</i>	92%	34%	77%	34%	92%	59%	80%
	<i>Wave 5</i>	88%	68%	84%	87%	88%	75%	89%
	<i>General</i>	81%	62%	73%	62%	81%	73%	72%
<i>Validation wave 6</i>	<i>Wave 5</i>	0%	100%	39%	100%	0%	39%	0%
	<i>General</i>	43%	66%	52%	66%	43%	42%	67%
	<i>Wave 6</i>	92%	18%	64%	18%	92%	59%	64%

References

- [1]. Gobierno de México. ¿Qué es el coronavirus? COVID-19. CDMX: Gobierno de México; 2020. <https://coronavirus.gob.mx/> del documento subordinado. Accessed 10 Mar 2023.
- [2]. World Health Organization. Coronavirus disease (COVID-19). Coronavirus disease (COVID-19) PANDEMIC. World Health Organization. 2020. https://www.who.int/es/emergencias/diseases/novel-coronavirus-2019?gclid=EAIaIQobChMIqYKjrvnw_gIVcCutBh0a-AYkEAAAYASAAEgL5BvD_BwE del documento subordinado. Accessed 8 Feb 2023.
- [3]. Gobierno de México. Sistema Nacional de Vigilancia Epidemiológica, dirección de vigilancia epidemiológica. Ciudad de México: SINAVE; 2023. <https://datos.covid-19.conacyt.mx/> del documento subordinado. Accessed 20 Mar 2023.
- [4]. Palacios M, Santos E, Velázquez MA, León M. COVID-19, a worldwide public health emergency, *Rev Clin Esp.* 2021; 221(1):55–61; <https://doi.org/10.1016/j.rceng.2020.03.001>
- [5]. Salud Digna. ¿Síntomas de COVID? Realízate una prueba. México: Salud Digna; 2023. <https://www.salud-digna.org/informativa-covid-19/> del documento subordinado. Accessed 10 May 2023.
- [6]. Laboratorio Médico del CHOPO. COVID. México: CHOPO; 2023. <https://www.chopo.com.mx/hidalgo/pruebas-para-covid-19> del documento subordinado. Accessed 10 May 2023.
- [7]. Organización Panamericana de la Salud. De Información Y Las Tecnologías En El Primer Nivel De Atención. Organización Mundial de la Salud. 2020. <https://www.paho.org/es/temas/coronavirus/brote-enfermedad-por-coronavirus-covid-19> del documento subordinado. Accessed 5 Feb 2023.
- [8]. Fontal A, Bouma MJ, San José A, Lopez L, Pascual M, Rodó X. Climatic signatures in the different COVID-19 pandemic waves across both hemispheres. *Nature Comput Sci.* 2021; (1): 655–665; <https://doi.org/10.1038/s43588-021-00136-6>
- [9]. Roland LT, Gurrola JG, Loftus PA, Cheung SW, Chang JL. Smell and taste symptom-based predictive model for COVID-19 diagnosis, *Int Forum Allergy Rhinol.* 2020; 10(7):832–838; <https://doi.org/10.1002/alr.22602>
- [10]. Jehi L, Ji X, Milinovich A, Erzurum S, Rubin BP, Gordon S, et al. Individualizing Risk Prediction for Positive Coronavirus Disease 2019 Testing: Results From 11,672 Patients, *Chest.* 2020; 158(4):1364–1375; <https://doi.org/10.1016/j.chest.2020.05.580>
- [11]. Zhang SX, Sun S, Afshar JA, Wang Y, Nazarian MA, Li J, Mokhtari DM. Beyond Predicting the Number of Infections: Predicting Who is Likely to Be COVID Negative or Positive, *Risk Manag Healthc Policy.* 2020; 13, 2811–2818; <https://doi.org/10.2147/RMHP.S273755>
- [12]. Wynants L, Van CB, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal, *BMJ.* 2020; 369 m1328; <https://doi.org/10.1136/bmj.m1328>
- [13]. Carrillo VMF, Salinas EG, García PC, Gutiérrez RLM, Parra RL. Early estimation of the risk factors for hospitalization and mortality by COVID-19 in Mexico, *PLoS One.* 2020; 15(9), e0238905; <https://doi.org/10.1371/journal.pone.0238905>
- [14]. Mamidi T, Tran NK, Melvin RL, Worthey EA. Development of An Individualized Risk Prediction Model for COVID-19 Using Electronic Health Record Data, *Front Big Data.* 2021; 4, 675882; <https://doi.org/10.3389/fdata.2021.675882>
- [15]. Challener D, Challener G, Gow-Lee V, Fida M, Shah A, O'Horo J. Screening for COVID-19: Patient factors predicting positive PCR test, *Infect Control Hosp Epidemiol.* 2020; 41(8):968–969; <https://doi.org/10.1017/ice.2020.249>
- [16]. Alyafei K, Ahmed R, Abir F, Chowdhury H, Naji K. A comprehensive review of COVID-19 detection techniques: From laboratory systems to wearable devices. *Comput Biol Med.* 2022; 149, 106070; <https://doi.org/10.1016/j.combiomed.2022.106070>
- [17]. Sharma A, Ahmad Farouk I, Lal SK. COVID-19: A Review on the Novel Coronavirus Disease Evolution, Transmission, Detection, Control and Prevention, *Viruses.* 2021; 13(2), 202; <https://doi.org/10.3390/v13020202>
- [18]. Gao Z, Xu Y, Sun C, Wang X, Guo Y, Qiu S. A systematic review of asymptomatic infections with COVID-19, *J Microbiol Immunol Infect.* 2021; 54(1), 12–16; <https://doi.org/10.1016/j.jmii.2020.05.001>

- [19]. Kandeel M, Ibrahim A, Fayez M, Al-Nazawi M. From SARS and MERS CoVs to SARS-CoV-2: Moving toward more biased codon usage in viral structural and nonstructural genes, *J Med Virol.* 2020; 92(6):660–666; <https://doi.org/10.1002/jmv.25754>
- [20]. Chilamakuri R, Agarwal S. COVID-19: Characteristics and Therapeutics, *Cells.* 2021; 10(2), 206; <https://doi.org/10.3390/cells10020206>
- [21]. Centros para el control de enfermedades E. Salud Pública en los Estados Unidos y México | CDC. DBC 24/7: Salvamos vidas, EUA. 2020. <https://www.cdc.gov/spanish/> del documento subordinado. Accessed 25 Nov 2022.
- [22]. Iser BP, Sliva I, Raymundo V. Poletto M, Trevisol BF. Suspected COVID-19 case definition: a narrative review of the most frequent signs and symptoms among confirmed cases, *Epidemiol e Serv saude Rev do Sist Unico Saude do Bras.* 2020; 29(3):e2020233; <https://doi.org/10.5123/S1679-49742020000300018>
- [23]. Cascella M, Rajnik MA, Scott A, Dulebohn CR. Features, Evaluation, and Treatment of Coronavirus (COVID-19), *PubMed, StatPearls Publishing.* 2023.
- [24]. Khan M, Khan H, Khan S, Nawaz M. Epidemiological and clinical characteristics of coronavirus disease (COVID-19) cases at a screening clinic during the early outbreak period: a single-centre study, *Journal Med Microbiol.* 2020; 69(8), 1114–1123; <https://doi.org/10.1099/jmm.0.001231>
- [25]. Tandam M, Acharya Y, Pokharel S, Timilsina M. Discovering symptom patterns of COVID-19 patients using association rule mining. *Computers Biol Med.* 2021; 131, 104249; <https://doi.org/10.1016/j.combiomed.2021.104249>
- [26]. De Diputados C, Congreso De DH, Unión LA. REGLAMENTO DE LA LEY GENERAL DE SALUD EN MATERIA DE INVESTIGACION PARA LA SALUD. D Of la Fed. 2014. http://www.diputados.gob.mx/LeyesBiblio/regley/Reg_LGS_MIS.pdf del documento subordinado. Accessed 30 Aug 2021.
- [27]. Gobierno de la Ciudad de México. Decretos, normas y acuerdos oficiales. Publicaciones en la Gaceta Oficial de la Ciudad de México, derivadas de la contingencia sanitaria por COVID-19. Ciudad de México: Gobierno de México; 2020. <https://servicios.covid19.cdmx.gob.mx/decretos-normas-y-acuerdos-oficiales> del documento subordinado. Accessed 11 Sep 2021.
- [28]. Gobierno de México. Datos Abiertos de México - Información referente a casos COVID-19 en México. Secretaría de Salud. 2021. <https://datos.gob.mx/busca/dataset/informacion-referente-a-casos-covid-19-en-mexico> del documento subordinado. Accessed 24 Oct 2021.
- [29]. COVID-19 y el tratamiento de datos personales. <https://micrositios.inai.org.mx/covid-19/> del documento subordinado. Accessed 24 Oct 2021.