



**UNIVERSIDAD AUTÓNOMA  
DE AGUASCALIENTES  
CENTRO DE CIENCIAS BÁSICAS**

**DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN,  
ESTADÍSTICA, MATEMÁTICAS Y FÍSICA**

**TESIS**

**MODELO INTELIGENTE PARA CONOCER LA EVOLUCIÓN DE LA  
PANDEMIA DE COVID-19 EN MÉXICO**

**PRESENTA**

**Itzel Paola Cervera Argüelles**

**PARA OBTENER EL GRADO DE MAESTRÍA EN CIENCIAS CON  
OPCIONES A LA COMPUTACIÓN, MATEMÁTICAS APLICADAS**

**TUTORES**

**Dr. Hermilo Sánchez Cruz**

**Dr. Juan Humberto Sossa Azuela**

**ASESOR**

**Dr. Ángel Eduardo Muñoz Zavala**

**Aguascalientes, Ags, 9 de octubre de 2023**

## Autorizaciones

### CARTA DE VOTO APROBATORIO INDIVIDUAL

**M. EN C. JORGE MARTÍN ALFÉREZ CHÁVEZ**  
DECANO DEL CENTRO DE CIENCIAS BÁSICAS  
P R E S E N T E

Por medio del presente como **TUTOR** designado del estudiante **ITZEL PAOLA CERVERA ARGUELLES** con ID 174803 quien realizó la tesis titulada: **MODELO INTELIGENTE PARA CONOCER LA EVOLUCIÓN DE LA PANDEMIA DE COVID-19 EN MÉXICO**, un trabajo propio, innovador, relevante e inédito y con fundamento en el Artículo 175, Apartado II del Reglamento General de Docencia doy mi consentimiento de que la versión final del documento ha sido revisada y las correcciones se han incorporado apropiadamente, por lo que me permito emitir el **VOTO APROBATORIO**, para que ella pueda proceder a imprimirla así como continuar con el procedimiento administrativo para la obtención del grado.

Pongo lo anterior a su digna consideración y sin otro particular por el momento, me permito enviarle un cordial saludo.

**ATENTAMENTE**  
"Se Lumen Proferre"  
Aguascalientes, Ags., a 09 de octubre de 2023.



---

**Dr. Hermilo Sánchez Cruz**  
Tutor de tesis

c.c.p.- Interesado  
c.c.p.- Secretaría Técnica del Programa de Posgrado

Elaborado por: Depto. Apoyo al Posgrado.  
Revisado por: Depto. Control Escolar/Depto. Gestión de Calidad.  
Aprobado por: Depto. Control Escolar/ Depto. Apoyo al Posgrado.

Código: DO-SEE-FO-07  
Actualización: 01  
Emisión: 17/05/19

CARTA DE VOTO APROBATORIO  
INDIVIDUAL

**M. EN C. JORGE MARTÍN ALFÉREZ CHÁVEZ**  
DECANO DEL CENTRO DE CIENCIAS BÁSICAS  
P R E S E N T E

Por medio del presente como **TUTOR** designado del estudiante **ITZEL PAOLA CERVERA ARGUELLES** con ID 174803 quien realizó la tesis titulada: **MODELO INTELIGENTE PARA CONOCER LA EVOLUCIÓN DE LA PANDEMIA DE COVID-19 EN MÉXICO**, un trabajo propio, innovador, relevante e inédito y con fundamento en el Artículo 175, Apartado II del Reglamento General de Docencia doy mi consentimiento de que la versión final del documento ha sido revisada y las correcciones se han incorporado apropiadamente, por lo que me permito emitir el **VOTO APROBATORIO**, para que ella pueda proceder a imprimirla así como continuar con el procedimiento administrativo para la obtención del grado.

Pongo lo anterior a su digna consideración y sin otro particular por el momento, me permito enviarle un cordial saludo.

**A T E N T A M E N T E**  
"Se Lumen Proferre"  
Aguascalientes, Ags., a 09 de octubre de 2023.



---

**Dr. Juan Humberto Sossa Azuela**  
Tutor de tesis

c.c.p.- Interesado  
c.c.p.- Secretaría Técnica del Programa de Posgrado

CARTA DE VOTO APROBATORIO  
INDIVIDUAL

**M. EN C. JORGE MARTÍN ALFÉREZ CHÁVEZ**  
DECANO DEL CENTRO DE CIENCIAS BÁSICAS  
P R E S E N T E

Por medio del presente como **ASESOR** designado del estudiante **ITZEL PAOLA CERVERA ARGUELLES** con ID 174803 quien realizó la tesis titulada: **MODELO INTELIGENTE PARA CONOCER LA EVOLUCIÓN DE LA PANDEMIA DE COVID-19 EN MÉXICO**, un trabajo propio, innovador, relevante e inédito y con fundamento en el Artículo 175, Apartado II del Reglamento General de Docencia doy mi consentimiento de que la versión final del documento ha sido revisada y las correcciones se han incorporado apropiadamente, por lo que me permito emitir el **VOTO APROBATORIO**, para que ella pueda proceder a imprimirla así como continuar con el procedimiento administrativo para la obtención del grado.

Pongo lo anterior a su digna consideración y sin otro particular por el momento, me permito enviarle un cordial saludo.

**ATENTAMENTE**  
"Se Lumen Proferre"  
Aguascalientes, Ags., a 09 de octubre de 2023.



---

**Dr. Ángel Eduardo Muñoz Zavala**  
Asesor de tesis

c.c.p.- Interesado  
c.c.p.- Secretaría Técnica del Programa de Posgrado

Elaborado por: Depto. Apoyo al Posgrado.  
Revisado por: Depto. Control Escolar/Depto. Gestión de Calidad.  
Aprobado por: Depto. Control Escolar/ Depto. Apoyo al Posgrado.

Código: DO-SEE-FO-07  
Actualización: 01  
Emisión: 17/05/19



DICTAMEN DE LIBERACIÓN ACADÉMICA PARA INICIAR LOS TRAMITES DEL EXAMEN DE GRADO



Fecha de dictaminación dd/mm/aaaa: 03/11/2023

**NOMBRE:** ITZEL PAOLA CERVERA ARGUELLES **ID** 174803

**PROGRAMA:** MAESTRÍA EN CIENCIAS, CON OPCIONES A LA COMPUTACIÓN, MATEMÁTICAS APLICADAS **LGAC (del posgrado):** COMPUTACIÓN-INTELIGENCIA ARTIFICIAL

**TIPO DE TRABAJO:** ( X ) Tesis ( ) Trabajo Práctico

**TITULO:** MODELO INTELIGENTE PARA CONOCER LA EVOLUCIÓN DE LA PANDEMIA DE COVID-19 EN MÉXICO

**IMPACTO SOCIAL (señalar el impacto logrado):** Se elaboró un modelo que permite observar los factores de riesgo para que una persona fallezca por COVID-19 en México, igualmente se analizan estadísticas nacionales e internacionales para evaluar el impacto de la pandemia en la mortalidad nacional.

**INDICAR SI NO N.A. (NO APLICA) SEGÚN CORRESPONDA:**

INDICAR	SI	NO	N.A. (NO APLICA)	SEGÚN CORRESPONDA:
<i>Elementos para la revisión académica del trabajo de tesis o trabajo práctico:</i>				
SI				El trabajo es congruente con las LGAC del programa de posgrado
SI				La problemática fue abordada desde un enfoque multidisciplinario
SI				Existe coherencia, continuidad y orden lógico del tema central con cada apartado
SI				Los resultados del trabajo dan respuesta a las preguntas de investigación o a la problemática que aborda
SI				Los resultados presentados en el trabajo son de gran relevancia científica, tecnológica o profesional según el área
SI				El trabajo demuestra más de una aportación original al conocimiento de su área
SI				Las aportaciones responden a los problemas prioritarios del país
N.A.				Generó transferencia del conocimiento o tecnológica
SI				Cumple con la ética para la investigación (reporte de la herramienta anti plagio)
<i>El egresado cumple con lo siguiente:</i>				
SI				Cumple con lo señalado por el Reglamento General de Docencia
SI				Cumple con los requisitos señalados en el plan de estudios (créditos curriculares, optativos, actividades complementarias, estancia, predoctoral, etc.)
SI				Cuenta con los votos aprobatorios del comité tutorial, en caso de los posgrados profesionales si tiene solo tutor podrá liberar solo el tutor
N.A.				Cuenta con la carta de satisfacción del Usuario
SI				Coincide con el título y objetivo registrado
SI				Tiene congruencia con cuerpos académicos
SI				Tiene el CVU del Conacyt actualizado
N.A.				Tiene el artículo aceptado o publicado y cumple con los requisitos institucionales (en caso que proceda)
<i>En caso de Tesis por artículos científicos publicados</i>				
N.A.				Aceptación o Publicación de los artículos según el nivel del programa
N.A.				El estudiante es el primer autor
N.A.				El autor de correspondencia es el Tutor del Núcleo Académico Básico
N.A.				En los artículos se ven reflejados los objetivos de la tesis, ya que son producto de este trabajo de investigación.
N.A.				Los artículos integran los capítulos de la tesis y se presentan en el idioma en que fueron publicados
N.A.				La aceptación o publicación de los artículos en revistas indexadas de alto impacto

Con base a estos criterios, se autoriza se continúen con los trámites de titulación y programación del examen de grado: Sí  No

FIRMAS

**Elaboró:**  
 \* NOMBRE Y FIRMA DEL CONSEJERO SEGÚN LA LGAC DE ADSCRIPCION: Dr. Hermilo Sánchez Cruz  
 NOMBRE Y FIRMA DEL SECRETARIO TÉCNICO: Dr. Hermilo Sánchez Cruz  
 \* En caso de conflicto de intereses, firmará un revisor miembro del NAB de la LGAC correspondiente distinto al tutor o miembro del comité tutorial, señalado por el Decano  
**Revisó:**  
 NOMBRE Y FIRMA DEL SECRETARIO DE INVESTIGACIÓN Y POSGRADO: Dr. Juan Jáuregui Brincón  
**Autorizó:**  
 NOMBRE Y FIRMA DEL DECANO: M. en C. Jorge Martín Alférez Chávez

**Nota: procede el trámite para el Depto. de Apoyo al Posgrado**  
 En cumplimiento con el Art. 105C del Reglamento General de Docencia que a la letra señala entre las funciones del Consejo Académico: ... Cuidar la eficiencia terminal del programa de posgrado y el Art. 105F las funciones del Secretario Técnico, llevar el seguimiento de los alumnos.

## Agradecimientos

En primer lugar al Consejo Nacional de Humanidades, Ciencias y Tecnología (*Conahcyt*) y a la Universidad Autónoma de Aguascalientes ya que sin su valioso apoyo económico este trabajo no se hubiera podido realizar.

Me gustaría agradecer también a mis tutores el Dr. Hermilo Sánchez Cruz y el Dr. Juan Humberto Sossa Azuela, a mi asesor el Dr. Ángel Eduardo Muñoz Zavala cuyos comentarios fueron indispensables para poder resolver los problemas que se presentaron y poder entregar un trabajo que tuviera un impacto social.

Finalmente agradecer a mi familia cuyo apoyo me permitió concluir satisfactoriamente con mi educación.

## Dedicatorias

A mi familia

## Índice general

Resumen.....	7
Abstract.....	8
1. Introducción .....	9
2. Antecedentes del problema.....	12
3. Fundamentos teóricos.....	14
3.1 Antecedentes de salud pública y epidemiología.....	14
3.1.1 Incidencia.....	16
3.1.2 Prevalencia .....	16
3.1.3 Letalidad .....	17
3.1.4 Mortalidad general y específica.....	17
3.1.5 Riesgo relativo (RR).....	18
3.1.6 Riesgo absoluto ( $R_{Abs}$ ).....	19
3.1.7 Riesgo atribuible (RA).....	19
3.1.8 Inmunidad de rebaño .....	19
3.2 Pandemia de COVID-19 y campaña mundial de vacunación .....	21
3.3 Modelos de aprendizaje de máquina.....	26
3.3.1 Máquinas de soporte vectorial (SVM).....	28
3.3.2 Árboles de decisión.....	30
3.3.3 Bosques aleatorios.....	32
3.3.4 Preprocesamiento de los datos .....	32
3.3.5 Mediciones para la calidad del modelo.....	33
4 Metodología .....	36
4.1 Preprocesamiento de los datos .....	36
4.2 Estadísticas y gráficas.....	43
4.3 Selección del algoritmo para el modelo .....	44
4.4 Construcción del modelo.....	45
4.5 Revisión de información sobre campaña de vacunación en el mundo.....	47
5 Resultados y discusiones.....	48
5.1 Resultados .....	48
5.1.1 Estadísticas mundiales .....	48
5.1.2 Estadísticas nacionales.....	62

5.1.3	Modelo .....	72
5.1.4	Resultados de vacunación .....	80
5.1.5	Estadísticas mundiales .....	92
5.1.6	Modelo y estadísticas nacionales.....	94
5.1.7	Vacunación .....	96
6	Conclusiones y trabajo a futuro .....	100
6.1	Conclusiones .....	100
6.2	Trabajos a futuro.....	100
	Referencias.....	102
	Anexos.....	108



## Índice de Tablas

Tabla 1. Tabla de contingencia para el cálculo de RR [16] .....	18
Tabla 2. Variantes de SARS-COV-2 [25], [27].....	22
Tabla 3. Lista de vacunas disponibles para México [31]–[34].....	25
Tabla 4. Estructura de una matriz de confusión sin importar el tipo de problema [38] .....	34
Tabla 5. Variables consideradas para la creación del modelo de aprendizaje de máquina .....	38
Tabla 6. Variables consideradas para la comparación de países de interés .....	39
Tabla 7. Variables que fueron modificadas para realizar la graficación .....	40
Tabla 8. Países y regiones de interés consideradas para realizar las comparaciones .....	41
Tabla 9. Variables consideradas para el análisis de los resultados de la campaña nacional de vacunación .....	43
Tabla 10. Letalidad y mortalidad de los países de interés ordenado de acuerdo con la mortalidad .....	48
Tabla 11. Promedio por grupos de características de la población.....	52
Tabla 12. Promedio por grupo de las características específicas de la pandemia .....	54
Tabla 13. Casos registrados según el nivel de severidad de la enfermedad de COVID-19 .....	64
Tabla 14. Casos por año de las diferentes etapas de la COVID-19 .....	64
Tabla 15. Comparación entre fallecimientos con comorbilidades y sin comorbilidades .....	66
Tabla 16. Tabla de contingencia para cálculo de riesgo relativo y riesgo absoluto .....	66
Tabla 17. Casos y defunciones por comorbilidad .....	67
Tabla 18. Valores de riesgo para cada comorbilidad .....	68
Tabla 19. Tiempo de ejecución según la cantidad de datos usados.....	72
Tabla 20. Valor de las métricas según la cantidad de datos usados .....	73
Tabla 21. Resultado de las métricas según las variables utilizadas .....	74
Tabla 22. Resultados para las métricas del modelo .....	79
Tabla 23. Grados de severidad de un evento adverso .....	81
Tabla 24. Plataformas de vigilancia para reportes de vacunación en el mundo.....	81
Tabla 25. Eventos más reportados por vacuna en las plataformas de vigilancia .....	82
Tabla 26. Casos de miocarditis o pericarditis esperados y observados después de la segunda dosis de las vacunas de ARN mensajero considerando que los síntomas aparecen en los primeros 7 días [48].....	84
Tabla 27. Reportes de miocarditis en Israel [49].....	85
Tabla 28. Casos de miocarditis registrados por grupo de edad [50].....	86
Tabla 29. Características de los casos severos y no severos de miocarditis [50].....	86
Tabla 30. Características de los 8 pacientes con muerte súbita cardíaca [50].....	87
Tabla 31. Casos de miocarditis por vacuna y dosis recibida [50] .....	87
Tabla 32. Reportes de efectos adversos registrados en VigiAccess para las vacunas del esquema nacional de vacunación [53].....	89
Tabla 33. Dosis aplicadas en México por vacuna [54].....	90
Tabla 34. Características de las defunciones .....	95
Tabla 35. Criterios de exclusión para participar en el estudio de las vacunas .....	97

## Índice de Figuras

Fig. 1 Tasa de defunciones en 2021 publicada por INEGI [2].....	10
Fig. 2 Clasificación de las medidas usadas en epidemiología [14] .....	16
Fig. 3 Interpretación gráfica del riesgo relativo [16] .....	19
Fig. 4 Dinámica de contagio en una población [19] .....	20
Fig. 5 Proceso de desarrollo de una vacuna normal contra el proceso seguido para las vacunas de COVID-19 [69].....	24
Fig. 6 Clasificación del aprendizaje automático .....	27
Fig. 7 Caso de dos clases separables [38] .....	29
Fig. 8 Posibles hiperplanos de separación [38] .....	29
Fig. 9 Clasificación de las dos clases considerando los vectores de soporte [38] .....	29
Fig. 10 Partes de un gráfico de árbol de decisión [70].....	31
Fig. 11 Proceso de reducción de la matriz de datos original para eliminar las clases desiguales ...	45
Fig. 12 Mortalidad por país de interés.....	50
Fig. 13. Letalidad por país de interés.....	50
Fig. 14 Casos totales de COVID-19 por país.....	51
Fig. 15 Defunciones totales por COVID-19 por país.....	51
Fig. 16 Comparación de mediana de edad de los países de interés.....	53
Fig. 17 Comparación de prevalencia de diabetes para los países de interés .....	53
Fig. 18 Comparación de mortalidad por enfermedades cardiovasculares para los países de interés .....	54
Fig. 19 Pruebas de COVID-19 realizadas en los países de interés .....	55
Fig. 20 Promedio de pruebas realizadas para encontrar un caso positivo por país de interés .....	55
Fig. 21 Total de vacunas aplicadas en los países de interés .....	56
Fig. 22 Porcentaje de la población que ha recibido una vacuna en los países de interés .....	56
Fig. 23 Porcentaje de la población que ha completado su esquema de vacunación en los países de interés.....	57
Fig. 24 Total de refuerzos aplicados por cada 100,000 habitantes en los países de interés .....	57
Fig. 25 Índice de restricción máximo observado en cada país de interés .....	58
Fig. 26 Casos diarios en Perú, México y China .....	59
Fig. 27 Casos diarios en Colombia, México y España.....	60
Fig. 28 Defunciones diarias en Perú, México y China .....	60
Fig. 29 Defunciones diarias en Colombia, México y España.....	61
Fig. 30. Comparación entre casos nuevos e índice de restricciones en México .....	61
Fig. 31 Casos nuevos en México .....	62
Fig. 32 Defunciones diarias en México .....	63
Fig. 33 Defunciones y sobrevivientes por sexo en México .....	63
Fig. 34 Número de sobrevivientes y fallecidos por sexo y grupo de edad .....	65
Fig. 35 Comparación de porcentaje de fallecidos por comorbilidad .....	69
Fig. 36 Porcentaje de defunciones por comorbilidad sin considerar hipertensión, diabetes y obesidad .....	70
Fig. 37 Casos de neumonía y que fueron hospitalizados por comorbilidad .....	71
Fig. 38 Casos que visitaron UCI y requirieron intubación por comorbilidad .....	71
Fig. 39 Cambio en la exactitud según el número de hojas usadas .....	75
Fig. 40 Modelo completo desarrollado para las variables mencionadas .....	76
Fig. 41 Subárbol 1. Casos que sobreviven a la COVID-19.....	77
Fig. 42 Subárbol 2. Casos de fallecen por COVID-19 .....	78
Fig. 43 Subárbol 3. Casos que a pesar de no ser hospitalizado fallecen después de desarrollar neumonía .....	79

Fig. 44 Matriz de confusión del modelo ..... 79  
Fig. 45 Defunciones por COVID-19 de personas vacunadas en México por sexo ..... 91  
Fig. 46 Defunciones por COVID-19 de personas vacunadas en México por grupo de edad ..... 91  
Fig. 47 Defunciones por COVID-19 de personas vacunadas en México por tipo de vacuna que recibieron..... 92



## Acrónimos

AE: Evento adverso por sus siglas en inglés *Adverse Event*.

AR: Reacción adversa por sus siglas en inglés *Adverse Reaction*.

ARNm: Vacunas de ARN mensajero (*mRNA* en inglés)

COVID-19: Enfermedad por el nuevo virus SARS-COV-2 por sus siglas en inglés *Coronavirus Disease 2019*

IA: Inteligencia Artificial

MAAE: Eventos adversos médicamente atendidos por sus siglas en inglés *Medically Attended Adverse Event*.

ML: Aprendizaje máquina por sus siglas en inglés *Machine Learning*.

OMS: Organización Mundial de la Salud

OPS: Organización Panamericana de la Salud

OWID: Proyecto *Our World In Data* de donde se obtuvo la base de datos para las comparaciones con otros países.

SAE: Evento adverso severo por sus siglas en inglés *Severe Adverse Event*.

SVM: Algoritmo de máquina de soporte vectorial por sus siglas en inglés *Support Vector Machine*.

## Resumen

Esta investigación recopila información de las bases de datos de *Our World in Data* y del Gobierno de México respecto a diferentes variables que permiten conocer los efectos que tuvo la pandemia causada por el nuevo virus SARS-COV-2 en México y el mundo. Se revisan y comparan características de un grupo de países de interés, entre los que se encuentra México, para conocer si hubo algún factor de riesgo en común entre los países que presentaron las mortalidades más elevadas. Igualmente se estudian las estadísticas nacionales para conocer quiénes fueron los más afectados así como las características más comunes entre las personas que fallecieron a causa de la COVID-19.

Al considerar los resultados elevados de letalidad y mortalidad en México para esta pandemia se busca desarrollar un modelo de aprendizaje de máquina que permita predecir cuándo una persona está en riesgo de fallecer por COVID-19 para así identificar los factores de riesgo en la población.

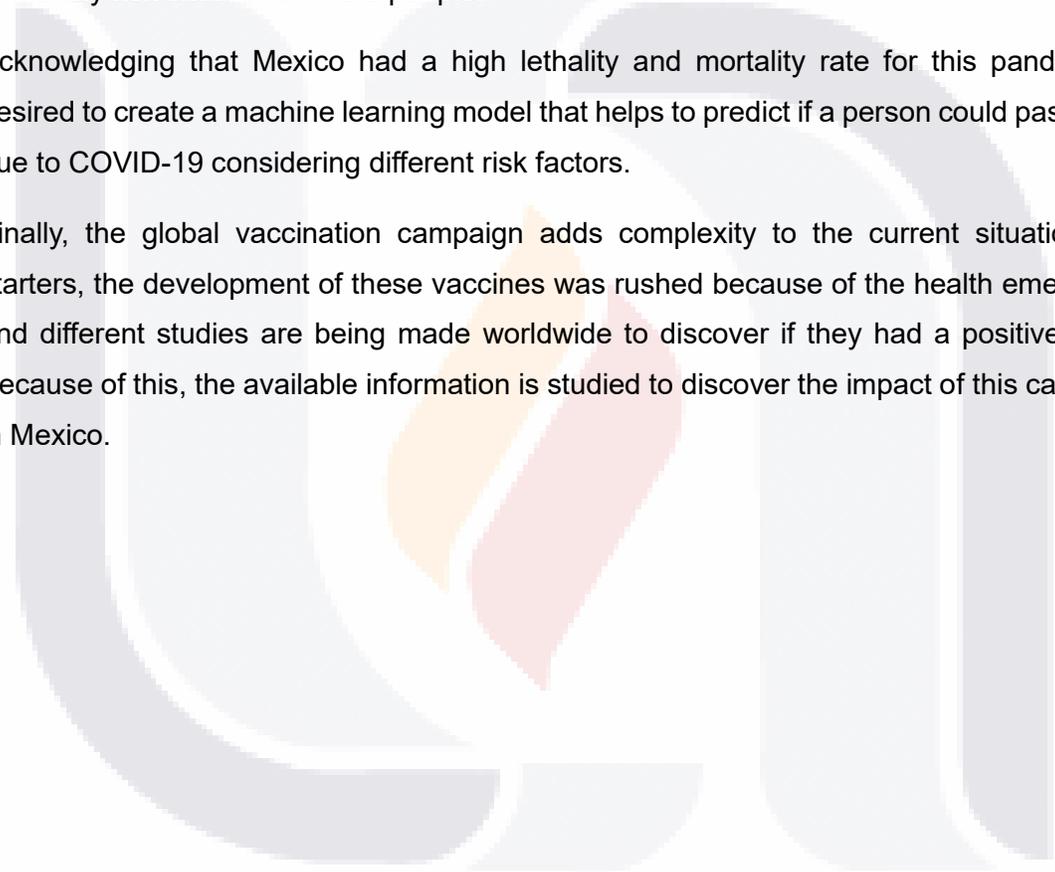
Debido a que un aspecto muy importante de esta pandemia fue el desarrollo acelerado de diferentes vacunas, se complementa la información anterior con una revisión de las características de la campaña masiva de vacunación que sucedió en el mundo con el fin de determinar si tuvo un impacto positivo o no en el mundo.

## Abstract

In this research, information from Our World in Data and Mexican Health Bureau databases is collected to know the consequences that the COVID-19 pandemic had in Mexico and the world. Distinct characteristics were compared between Mexico and a group of countries to discover if there is a common risk factor for a high mortality rate. The statistics in Mexico are studied to describe the characteristics of the population that succumbed to this illness and to identify risk situations for the people.

Acknowledging that Mexico had a high lethality and mortality rate for this pandemic is desired to create a machine learning model that helps to predict if a person could pass away due to COVID-19 considering different risk factors.

Finally, the global vaccination campaign adds complexity to the current situation. For starters, the development of these vaccines was rushed because of the health emergency, and different studies are being made worldwide to discover if they had a positive result. Because of this, the available information is studied to discover the impact of this campaign in Mexico.



# TESIS TESIS TESIS TESIS TESIS

## 1. Introducción

Las pandemias no son un problema nuevo para la humanidad, no obstante, la pandemia más reciente de COVID-19 causada por el nuevo virus SARS-COV-2 ha implicado retos muy grandes para la población mundial y cada país ha intentado hacerle frente de la mejor manera. Este virus fue detectado por primera vez en la ciudad de Wuhan, China a finales de noviembre del 2019 y en la mayoría de los casos se presenta como una enfermedad de vías respiratorias leve, sin embargo, sus características le han permitido esparcirse con facilidad alrededor del mundo causando 6.94 millones de decesos<sup>1</sup>.

En el caso específico de México se registraron 334,199 defunciones con lo cual se colocó en el 5° lugar a nivel mundial en defunciones por COVID-19 después de Estados Unidos (1.13 millones), Brasil (702,664), India (531,870) y Rusia (399,057). Para los cuatro países mencionados anteriormente estas cifras no resultan extremas ya que se encuentran entre los primeros lugares a nivel mundial de contagios; por ejemplo, Rusia se posiciona en el lugar 11 y reporta tres veces más contagios que México.

En el caso de México, este elevado número de defunciones junto con un número no tan elevado de contagios da como resultado valores altos para la letalidad y la mortalidad de esta enfermedad. Realizando nuevamente una comparación con otros países se puede conocer que si bien la mortalidad de esta enfermedad en México (259.2 defunciones por cada 100,000 habitantes) fue menor que las registradas en Reino Unido (313.87) e Italia (306.75) fue mayor que en España (243.7) y Francia (234.5).

Más específicamente, el INEGI reportó un impacto considerable en las defunciones del país a causa de esta pandemia ya que tan solo en el 2020 la tasa de defunciones por cada 10,000 habitantes aumentó de 59 a 86, Fig. 1. Esta situación estableció a la COVID-19 como la segunda causa de muerte más común en el país con 200,256 defunciones debajo de las enfermedades cardiovasculares (218,704 defunciones) y seguido por la diabetes mellitus (151,019 defunciones) [1].

En los reportes publicados por el INEGI también es posible conocer que los principales afectados por esta enfermedad fueron los hombres y las personas entre 35 y 64 años ya

---

<sup>1</sup> Datos al 31 de mayo del 2023 según la publicación de *OWID*.

que para estos grupos la enfermedad de COVID-19 fue la primera causa de muerte [1]. Para el 2021 la situación se mantiene igual para los grupos vulnerables mencionados antes, sin embargo, de manera general la situación empeora ligeramente (Fig. 1) ya que en ese año la COVID-19 fue la principal causa de muerte en el país con 238,772 defunciones [2].



Fuente: Estadísticas de Defunciones Registradas, 2012 a 2021

Fig. 1 Tasa de defunciones en 2021 publicada por INEGI [2]

Estas cifras elevadas podrían indicar factores que ponen en riesgo a la población mexicana respecto a este virus, por lo que surge la siguiente pregunta de investigación: ¿cómo saber con certeza si una persona se encuentra en riesgo de enfermar de gravedad o fallecer a causa de COVID-19?

Igualmente, esta pandemia trajo consigo el desarrollo acelerado de diferentes vacunas para evitar el desarrollo de la COVID-19 las cuales se pudieron aplicar mediante un permiso de uso de emergencia. La creación y distribución acelerada de estas vacunas de emergencia implicó un gran gasto económico para las compañías encargadas de fabricarlas y los gobiernos que las adquirieron, en el caso de México el gobierno aprobó 8 vacunas diferentes y se reporta que el 76.22% de la población ha recibido al menos una dosis<sup>2</sup> de un esquema de dos.

Por lo anterior, es natural preguntarse si los protocolos para la fabricación y seguridad de las vacunas se cumplieron y si estas implicaron algún beneficio para la población en

<sup>2</sup> Datos al 7 de octubre del 2022 según la publicación de OWID.

general, por lo que se plantea una segunda pregunta de investigación: ¿Es posible obtener un modelo que permita conocer si la vacunación ha tenido un impacto importante en México?

Considerando los puntos anteriores, esta investigación tiene como objetivo el siguiente punto:

- Encontrar un modelo inteligente para describir la evolución de la pandemia en México y brindar información más acertada para una mejor toma de decisiones.

Y particularmente, considerando que existen dos problemas principales, se desea:

- Utilizar técnicas de Aprendizaje de Máquina para encontrar un modelo que permita conocer la evolución de la pandemia de COVID-19 en México.
- Plantear un modelo diferente que incluya información que sea de utilidad para la toma de decisiones respecto a la pandemia.
- Identificar de forma certera los factores de riesgo para las personas que contraen COVID-19
- Identificar el impacto que tuvo el programa de vacunación sobre la pandemia.

Mediante los objetivos planteados anteriormente se desea demostrar que se cumplen los siguientes puntos:

1. Existe un modelo inteligente capaz de identificar los diferentes factores de riesgo que agravan la enfermedad de COVID-19.
2. La vacunación ha tenido un papel importante para la reducción de muertes por COVID-19.

En los siguientes apartados se describirán los pasos tomados para el desarrollo del modelo, además de la investigación realizada para poder llegar a una conclusión respecto a la campaña de vacunación. De igual forma, se realizan una serie de comparaciones entre países para poder analizar las características que tienen en común aquellos que presentaron valores altos de mortalidad.

## 2. Antecedentes del problema

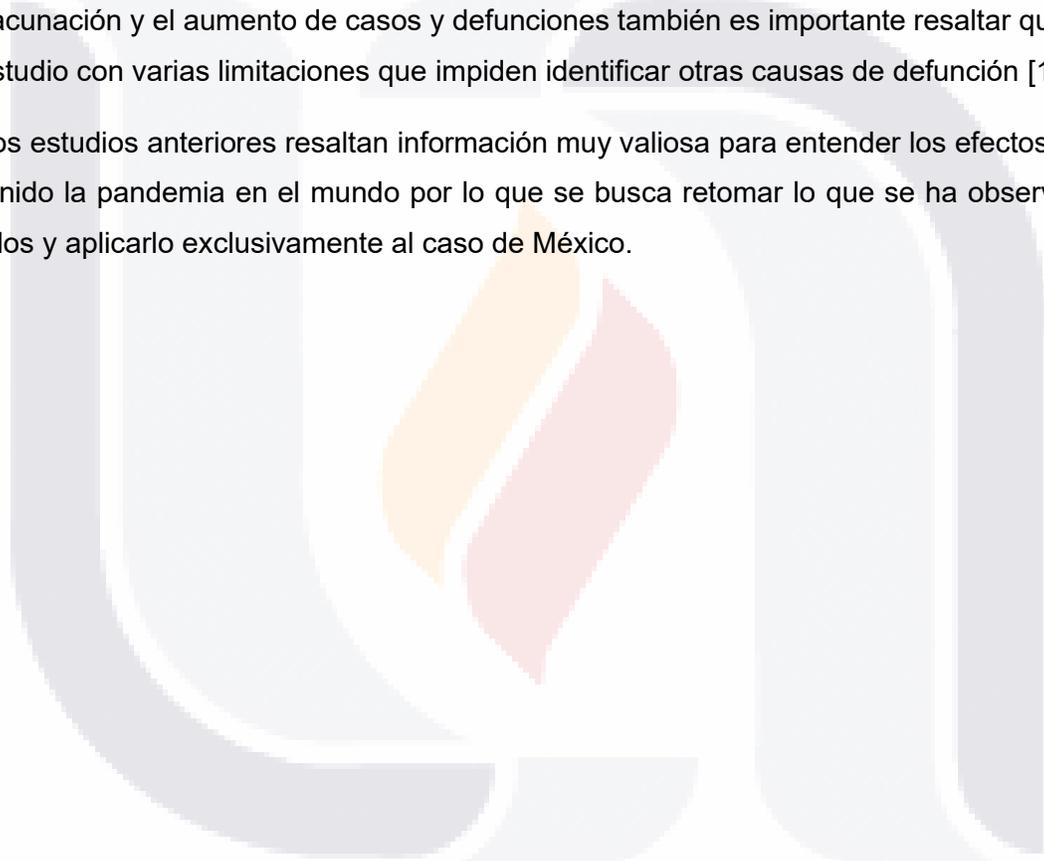
Tras la aparición de este nuevo virus los institutos se enfocaron en un principio a conocer las características de las poblaciones más afectadas, por lo que desde los primeros días de la pandemia fue posible conocer que la edad y la presencia de comorbilidades como hipertensión, diabetes, obesidad, inmunosupresión, enfermedad renal crónica, enfermedades cardiovasculares y pulmonares son factores de riesgo para que el estado de salud de una persona empeore después de contraer COVID-19 [3]–[5].

En el caso de México se realizaron estudios al inicio de la pandemia para conocer el impacto que las comorbilidades anteriores tienen sobre la población. Considerando los datos correspondientes al 2020, los dos estudios publicados hablan del impacto negativo de la obesidad y diabetes sobre el resultado esperado de la COVID-19 [6], [7], no obstante, contrastan respecto a las variables de edad y sexo ya que en uno de los estudios no resultan concluyentes [7] mientras que en el otro si se puede observar el impacto que tienen [6]. Estos dos estudios son muy relevantes ya que muestran como las comorbilidades preexistentes en la sociedad mexicana tienen un impacto en la mortalidad y letalidad que se presentan en el país.

Respecto a modelos para predecir el resultado de esta enfermedad en un principio el enfoque principal fue predecir el comportamiento que la pandemia presentaba en diferentes países [8]. Para México se observa que los primeros estudios se enfocaron en detectar las causas de riesgo para que las personas fallecieran debido a la COVID-19 sin embargo, no se planteaban modelos para realizar esta predicción [6], [7] sino que a través del análisis de riesgos determinaron que las comorbilidades tienen el impacto más grande en la letalidad por COVID-19 siendo la obesidad la principal responsable. En el 2022 la Universidad Veracruzana publicó un capítulo de libro en donde dan una introducción a temas de ciencia de datos tomando como caso práctico la información disponible sobre la pandemia en México [9]. Para desarrollar el caso práctico se consideraron los datos publicados por el Gobierno de México como Datos Abiertos, además de información obtenida del Sistema Nacional de Información Municipal (SNIM) para las características socioeconómicas de la región en la que se presentan los casos; para realizar los modelos se consideraron redes bayesianas y árboles de decisión.

Para la vacunación la información disponible se obtuvo principalmente de los artículos publicados por las farmacéuticas para demostrar la efectividad y seguridad de sus vacunas para obtener el permiso para uso de emergencia [10]–[12] además de múltiples estudios realizados alrededor del mundo para continuar analizando su seguridad y los efectos adversos relacionados con ellas. Para realizar un análisis más estadístico sobre el impacto que tuvo la vacunación en el mundo, en 2021 se publicó un artículo en donde se analiza la correlación entre la vacunación y el aumento de casos y defunciones por COVID-19; si bien este estudio tuvo como resultado que existe una relación positiva entre la campaña de vacunación y el aumento de casos y defunciones también es importante resaltar que es un estudio con varias limitaciones que impiden identificar otras causas de defunción [13].

Los estudios anteriores resaltan información muy valiosa para entender los efectos que ha tenido la pandemia en el mundo por lo que se busca retomar lo que se ha observado en ellos y aplicarlo exclusivamente al caso de México.



### 3. Fundamentos teóricos

Para analizar un problema como el que se plantea anteriormente, es fundamental integrar conocimientos de diferentes áreas, además de conocer la información disponible sobre el desarrollo que ha tenido. Por lo anterior, en esta sección se presentarán algunos conceptos importantes de las áreas de epidemiología, salud pública y computación inteligente fundamentales para los análisis realizados en los capítulos posteriores, de igual forma se incluye un breve resumen con las características más relevantes de la pandemia y la campaña de vacunación masiva que ocurrió a raíz de esta.

#### 3.1 Antecedentes de salud pública y epidemiología

La preservación de la salud es uno de los objetivos principales en todas las sociedades alrededor del mundo, por lo cual, es normal que aparezcan múltiples disciplinas que se encarguen de cumplir con este objetivo. Tal es el caso de la *Salud Pública* y de la *Epidemiología*. La primera es un área multidisciplinaria que busca concretar acciones en diversos ámbitos para obtener un beneficio en la salud de la población y la segunda se encarga de aplicar diversas herramientas para conocer los factores de riesgo, además de determinar las causas e impacto de las enfermedades que aparecen en una población [14].

En el caso de la Salud Pública, se abarcan una gran cantidad de acciones para garantizar el bienestar de los habitantes de una región, sin embargo, estas se pueden resumir en tres: 1) evaluación y vigilancia constante de los indicadores de salud para poder detectar oportunamente algún problema y establecer un orden de prioridad, 2) formular políticas que tengan como objetivo resolver los problemas identificados y 3) asegurar que la población sea capaz de acceder a un sistema de salud de calidad, el cual incluya los servicios de promoción y prevención de enfermedades [14].

Para desarrollar un sistema de vigilancia adecuado que pueda detectar correctamente los problemas de salud se deben considerar los siguientes componentes: determinantes de salud, demografía, epidemiología y sistemas de salud. En todos los casos es necesario recurrir a variables cuantitativas o cualitativas para obtener la información pertinente del estado de salud de la población y en el caso específico de una vigilancia epidemiológica se utilizan diversas medidas y variables para tomar decisiones que puedan influir positivamente en el mantenimiento del bienestar colectivo [14].

De acuerdo con Hernández [15], “la epidemiología es considerada la ciencia básica de la salud y ... estudia, sobre todo, las potenciales relaciones causales que se establecen entre la presencia de exposiciones determinadas y el desarrollo de enfermedades específicas, así como sus múltiples posibilidades de prevención”. Para cumplir con el objetivo de esta disciplina, se obtienen diversas medidas a través de las siguientes operaciones: proporciones, tasas y razones.

Por definición, las proporciones son medidas que establecen la frecuencia de un evento ( $F_E$ ) en relación con la población total que puede presentar dicho evento ( $P_R$ ), las tasas expresan el comportamiento de un evento en una población mediante la cantidad de ocurrencias de este ( $E_T$ ) y los periodos en los que la población se vio expuesta ( $P_R$ ); las razones ( $E1:E2$ ) se encargan de definir mediante un número la relación aritmética entre dos grupos o eventos ( $F_{E1}$  y  $F_{E2}$ ) de una misma población [15]. Las funciones correspondientes a cada concepto se encuentran listadas en las Ec.I, Ec.II, Ec.III [16] respectivamente y es conveniente enfatizar estas definiciones ya que, la investigación epidemiológica considera el tiempo, lugar y a los individuos al momento de obtener alguna medida [14].

$$Proporción = \frac{F_E}{P_R} \tag{Ec. I}$$

$$Tasa = \frac{E_T}{\sum_{t=1}^T P_R(t)} \times 10^n \tag{Ec. II}$$

$$E1 : E2 = \frac{F_{E2}}{F_{E1}} \tag{Ec. III}$$

Ahora bien, para agrupar las diversas medidas de la epidemiología se plantean tres clases: de frecuencia, para explicar la aparición de los eventos de salud a través del tiempo y comparar distintas poblaciones; de asociación, este tipo de medidas permiten determinar la fuerza con la que un evento de salud se relaciona con un factor de riesgo; y de impacto potencial, ayudan a identificar la fuerza con la que se relaciona la exposición y un evento de salud. Es necesario mencionar que las medidas consideran diferentes parámetros para obtener los resultados deseados [15]. En los siguientes apartados se explicarán las medidas más relevantes mientras que en la Fig. 2 se muestra la forma en la que estas medidas se agrupan.

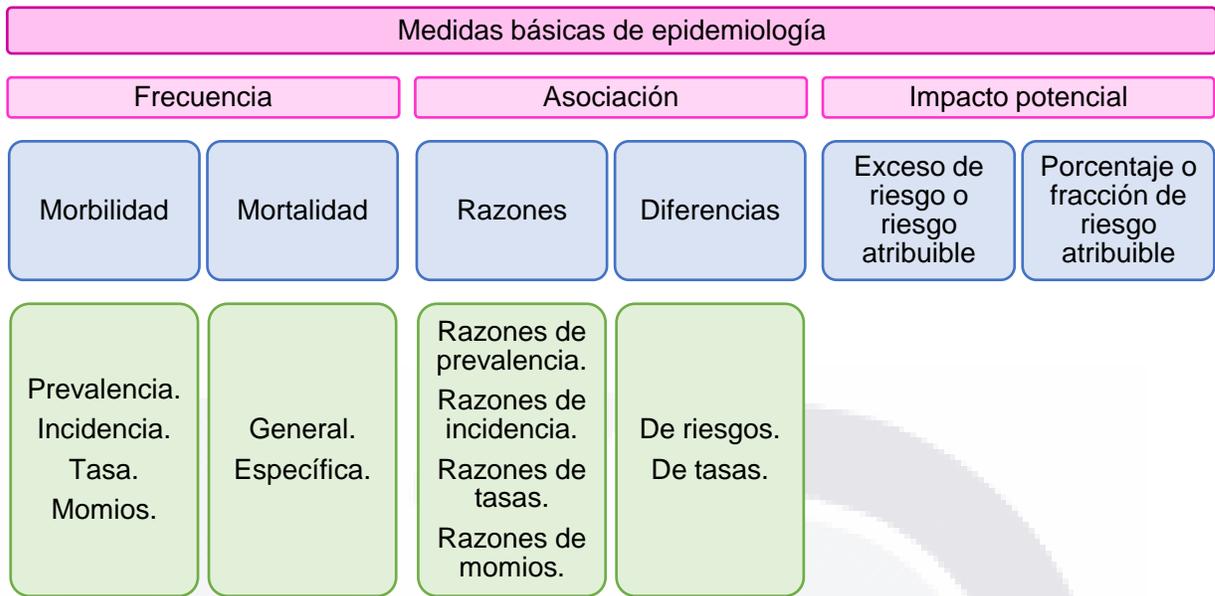


Fig. 2 Clasificación de las medidas usadas en epidemiología [14]

### 3.1.1 Incidencia

La tasa de incidencia ( $T.I$ ) es definida por Lemus [17] como “el número de casos nuevos de un evento relacionado con la salud ( $E_N$ ) en una población ( $P_R$ ) en un periodo determinado... Mide la aparición de la enfermedad, problema o evento de salud” y se modela mediante Ec. IV.

$$T.I = \frac{E_N}{P_R} \tag{Ec. IV}$$

Por lo anterior, se considera que la incidencia es la medida de frecuencia más significativa, además, esta tasa permite conocer el poder de contagio que tiene una enfermedad sobre una población [16].

### 3.1.2 Prevalencia

La prevalencia ( $P$ ), modelada mediante Ec. V, es una medida de frecuencia u ocurrencia que indica la existencia de un problema de salud. Por definición es “la proporción de casos de una cierta enfermedad ( $T_E$ ) en una población delimitada ( $P_T$ ), en un tiempo determinado” [17].

$$P = \frac{T_E}{P_T} \tag{Ec. V}$$

Esta medición resulta muy importante al momento de planificar la distribución de recursos, por lo cual es fundamental conocer que existen factores que pueden modificar su resultado;

por ejemplo, si mejora el tratamiento de pacientes con alguna enfermedad, pero no se erradica dicha enfermedad el resultado de la prevalencia va a aumentar [17].

### 3.1.3 Letalidad

Esta tasa, reportada como un porcentaje en Ec. VI [16], resulta de gran importancia ya que permite conocer la probabilidad de una persona de sobrevivir a una enfermedad. De manera más formal, es “la proporción de casos de una enfermedad que resultan mortales ( $D_E$ ) respecto del total de casos que contrajeron la enfermedad ( $P_E$ ) en un periodo especificado” [15].

$$L (\%) = \frac{D_E}{P_E} \times 100 \tag{Ec. VI}$$

Es fundamental resaltar que, de acuerdo con Lemus [17], “la letalidad está influenciada por los éxitos de las intervenciones médicas destinadas a curar las enfermedades... tiene una relación importante con las tasas de mortalidad de una enfermedad determinada”.

### 3.1.4 Mortalidad general y específica

Como concepto, la mortalidad explica “la magnitud con la que se presenta la muerte” dentro de una población durante un periodo de tiempo determinado; al ser un a medida estrictamente poblacional, la mortalidad puede ser comparada únicamente con mediciones que evalúen a grupos con las mismas características. La mortalidad puede estimarse de manera general para toda la población o para diferentes grupos poblacionales considerando sus edades, sexo o padecimiento en específico, por lo cual, se definen dos tipos [15].

*Mortalidad general ( $M_G$ )*. Este tipo de mortalidad, expresada matemáticamente en Ec. VII [16], habla del volumen de muertes ocurridas ( $D_T$ ) en una población ( $P_T$ ) sin importar la causa y sin hacer distinción por edad, sexo o cualquier otro motivo. Esta medición se obtiene de la siguiente forma [15]:

$$M_G = \frac{D_T}{P_T} (10^n) \tag{Ec. VII}$$

*Mortalidad específica ( $M_E$ )*. Esta tasa se obtiene al observar que existe algún evento de salud que está ocasionando un número mayor de muertes ( $D_E$ ), y para evitar alguna confusión se renombra la tasa de acuerdo con evento de salud [15]. Para poder estudiar el evento de interés se puede dividir a la población según su edad, sexo o padecimiento específico ( $P_R$ ), tal como se muestra en Ec. VIII. Para obtener esta medición se recurre a

una fórmula similar a la utilizada para obtener la mortalidad general, solamente realizando la siguiente modificación [16]:

$$M_E = \frac{D_E}{P_R} (10^n) \tag{Ec. VIII}$$

### 3.1.5 Riesgo relativo (RR)

Esta razón, clasificada como medida de asociación, permite conocer cuál es el riesgo de contraer una enfermedad dado que una persona se encuentra expuesta a algún factor de riesgo. Para esto, se compara la incidencia acumulada de un grupo expuesto contra un grupo que no está expuesto [16]. Para poder obtener el resultado de esta medición respecto a cualquier evento de salud es necesario recurrir a una tabla de 2 × 2 en donde se registran las comparaciones mencionadas con anterioridad y a la ecuación Ec. IX [15].

Tabla 1. Tabla de contingencia para el cálculo de RR [16]

		Exposición		
		Expuestos	No expuestos	
Casos	<i>A</i>	<i>b</i>	Total de casos ( <i>n<sub>i</sub></i> )	
No casos	<i>C</i>	<i>d</i>	Total de no casos ( <i>n<sub>o</sub></i> )	
	Total de expuestos ( <i>m<sub>i</sub></i> )	Total de no expuestos ( <i>m<sub>o</sub></i> )	Total de sujetos ( <i>n</i> )	

$$RR = \frac{a/n_i}{c/n_o} \tag{Ec. IX}$$

Ahora bien, para poder utilizar adecuadamente esta medición es necesario saber interpretarla. Para esto, se puede trazar una línea y ubicar tres puntos en ella: 0, 1 y α; el valor 0 representa el mínimo que se puede obtener, mientras que α puede ser cualquier valor mayor a 1; el resultado obtenido puede indicar un factor de protección o riesgo. Los resultados se clasifican de la siguiente manera:

- $RR = 1$  indica ausencia de asociación
- $RR > 1$  indica asociación positiva o factor de riesgo
- $RR < 1$  indica asociación negativa o factor de protección

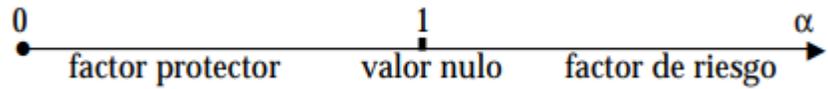


Fig. 3 Interpretación gráfica del riesgo relativo [16]

### 3.1.6 Riesgo absoluto (R<sub>Abs</sub>)

Por definición “a la incidencia de una enfermedad en una población se la denomina riesgo absoluto” y esta medición sirve para conocer “la magnitud del riesgo de sufrir una enfermedad en un grupo de personas con exposición a un factor de exposición” [18]. El resultado de esta medición es un porcentaje que, como se explica en la definición anterior, indica la presencia de la enfermedad en un grupo de personas que han sido expuestas o no, Ec. X y Ec. XI respectivamente; para obtener esta medición es necesario recurrir a la tabla de contingencia utilizada para obtener la medición del RR (Tabla 1) y a las siguientes ecuaciones [18]:

$$R_{AbsE} = \frac{a}{n_i}(100) \tag{Ec. X}$$

$$R_{AbsNE} = \frac{c}{n_o}(100) \tag{Ec. XI}$$

### 3.1.7 Riesgo atribuible (RA)

A esta medición también se le conoce como diferencia de incidencias, expresada como se muestra en Ec. XII, y se encarga de determinar el efecto que tiene un factor de riesgo sobre la frecuencia de aparición de una enfermedad. Esta medida es especialmente útil para tomar decisiones sobre la salud de la población y para poder obtener esta medición se requiere en gran medida del uso de la tabla de contingencia mostrada con anterioridad (Tabla 1) y de los valores de R<sub>Abs</sub> del grupo expuesto y no expuesto [17].

$$RA = R_{AbsE} - R_{AbsNE} \tag{Ec. XII}$$

### 3.1.8 Inmunidad de rebaño

La inmunidad de rebaño, también denominada inmunidad de grupo o colectiva, es definida por la OMS como la protección indirecta de una población contra un virus y se consigue cuando todas las personas (o una fracción considerable) han alcanzado la inmunidad, ya sea por haber recibido una vacuna o por haberse contagiado. Sanmartí menciona que esto

ocurre debido a que conforme se expande una enfermedad, el número de individuos que han adquirido una inmunización aumenta y la probabilidad de que una persona contagiada y una que no ha sido inmunizada se encuentren es más baja, por lo que el grupo de inmunizados bloquea la cadena de transmisión [19]. Esta dinámica se puede explicar mediante el siguiente gráfico:

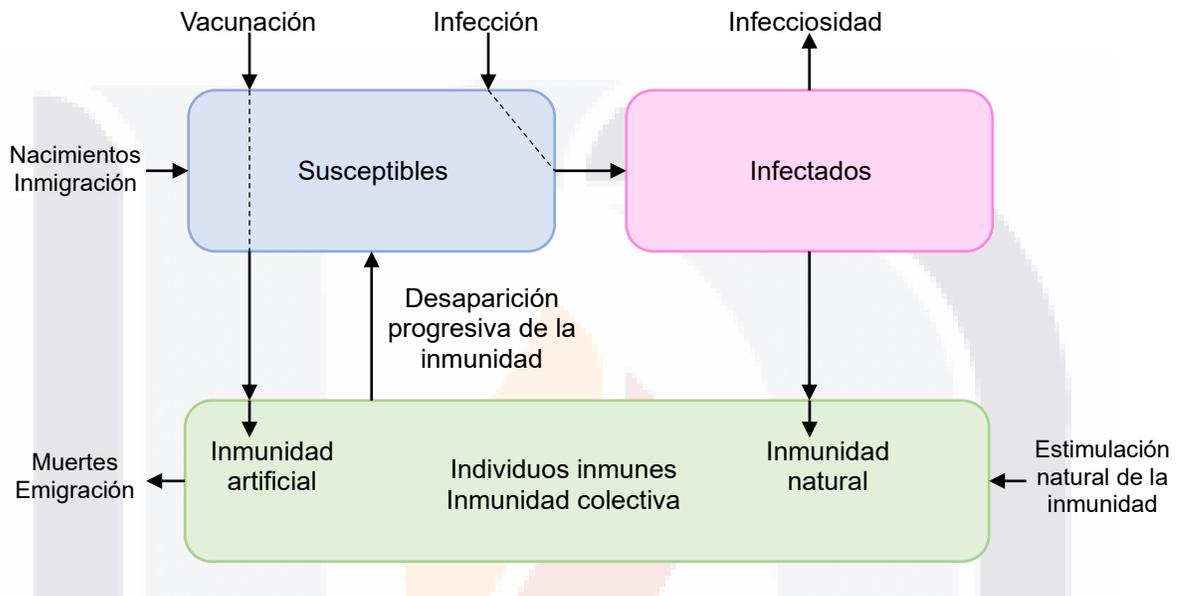


Fig. 4 Dinámica de contagio en una población [19]

La dinámica anterior es un ciclo que se ha observado para casi todas las infecciones y es fundamental para la planeación de campañas de vacunación que permitan conservar la salud. Es importante resaltar que, de acuerdo con Last, para determinar el número de individuos que deben ser inmunizados es necesario considerar los siguientes factores: el virus que se está analizando, sus características de transmisión, la forma en la que los individuos inmunes y susceptibles se encuentran distribuidos, entre otros [19].

La revisión de estos conceptos es fundamental para obtener algunos de los resultados de este estudio, además de ayudar a comprender algunas decisiones tomadas respecto a la campaña mundial de vacunación.

## 3.2 Pandemia de COVID-19 y campaña mundial de vacunación

La COVID-19 es una enfermedad infecciosa que afecta a las vías respiratorias y es causada por el nuevo virus SARS-COV-2. Este virus perteneciente a la familia de los coronavirus es altamente contagioso y, aunque la mayoría de las personas infectadas experimentarían una enfermedad respiratoria de leve a moderada, un grupo considerable de personas desarrollará una enfermedad grave que requerirá atención médica especializada y, en muchos casos aún con tratamiento médico causará la muerte [20].

Esta enfermedad se considera nueva dado que el primer caso identificado se presentó el 17 de noviembre del 2019 en Wuhan, China [21], mientras que el virus fue identificado por primera vez el 7 de enero del 2020 [22]. Al ser un virus altamente contagioso solo fueron necesarios dos meses para que, el 11 de marzo del 2020 [23], la enfermedad COVID-19 fuera catalogada como emergencia sanitaria y pandemia por la OMS<sup>3</sup>. A nivel nacional, el primer caso de COVID-19 fue detectado el 27 de febrero del 2020 y para el 30 de abril del mismo año se contabilizaban 19,224 casos en todo el país [22].

Como una posible estrategia para evitar que la situación empeorara y siguiendo el ejemplo de muchos otros países en el mundo, el 23 de marzo del 2020 el Gobierno de México declaró el inicio de la *Jornada Nacional de Sana Distancia* durante la cual se suspendieron completamente todas las actividades no esenciales y se invitó a la población a permanecer en confinamiento voluntario. Este periodo de aislamiento tenía previsto terminar un mes después, no obstante, se prolongó hasta el 30 de mayo del mismo año [24] y a partir de ese momento las actividades se fueron retomando muy lentamente hasta que se levantó la emergencia sanitaria por COVID-19 el 9 de mayo del 2023. Durante este tiempo las medidas de protección recomendadas fueron las siguientes: uso de cubrebocas para mayores de dos años, conservar una distancia de al menos 1.5 metros entre personas, higiene de manos, mantener los espacios ventilados, aislarse en caso de presentar síntomas o estar enfermo y vacunarse.

Las medidas anteriores tenían como objetivo disminuir los contagios ya que, además de las defunciones, uno de los problemas que se presentan durante la transmisión de un virus es la posibilidad de que la secuencia genética de este se vea alterada. A estas modificaciones se les denomina variantes y en algunos casos es necesario supervisarlas ya que el virus

---

<sup>3</sup> El 5 de mayo de 2023 la OMS declara el fin de la emergencia sanitaria por COVID-19.

puede volverse más contagioso, mortal o más agresivo con el sistema de los seres humanos. Para poder llevar un registro adecuado de estas alteraciones en la estructura del virus, la OMS se encarga de clasificar estas variantes en diferentes grupos: VOC, VOI y VUM<sup>4</sup> [25].

Las variantes VOC se refieren a las *Variantes Preocupantes (Variant of Concern)* y son aquellas que han demostrado un aumento en la transmisibilidad, virulencia, un cambio perjudicial en la epidemiología de la enfermedad, una disminución en la efectividad de las medidas de salud. Para el caso de las VOI (*Variants of Interest*) se habla de las *Variantes de Interés* y se incluyen en esta categoría las variantes que muestran un cambio en la secuencia genética del virus que podría afectar a las características de transmisibilidad, virulencia o que pudiera dificultar su detección o tratamiento. Finalmente, las VUM (*Variants Under Monitoring*) son las *Variantes Bajo Vigilancia* y son las variantes que presentan cambios en su genoma que pudieran representar alguna ventaja al momento de darse el contagio o el tratamiento, no obstante, aún no existen suficientes estudios para demostrar el impacto que pueda tener esta variante [25], [26].

En la Tabla 2, se muestran las variantes que la OMS consideró como preocupantes. Las variantes Alfa, Beta y Gamma dejaron esta clasificación en marzo del 2022 y la variante Ómicron dejó de serlo en marzo del 2023, además a esta fecha la OMS no reporta ninguna variante como preocupante, sin embargo, existe una (*variante XBB.1.5*) que aparece como variante de interés [25].

Tabla 2. Variantes de SARS-COV-2 [25], [27]

Linaje	Nombre	Origen	Fecha de primer caso
B.1.1.7	Alfa	Reino Unido	03-09-2020
B.1.351	Beta	Sudáfrica	01-09-2020
P.1 (B.1.1.28.1)	Gamma	Brasil	22-09-2020
B.1.617.2	Delta	India	01-03-2021
B.1.1.529	Ómicron	Sudáfrica y Botsuana	12-10-2021

<sup>4</sup> Además de la OMS, el Centro para el Control y la Prevención de Enfermedades (CDC) también realiza un seguimiento de las variantes del virus SARS-COV-2. En este caso se mantienen las clasificaciones de VOI y VOC pero se agregan las clases VBM (Variante Bajo Monitoreo) y VOHC (Variante de Gran Consecuencia). En esta referencia se puede observar una tabla que muestra una clasificación similar a la que presentaba la OMS [68].

TESIS TESIS TESIS TESIS TESIS

Como una posible solución a la pandemia de COVID-19 varios laboratorios en todo el mundo comenzaron a trabajar en el desarrollo de una vacuna que pudiera inmunizar a las personas.

Según Sanmartí, las vacunas son suspensiones de microorganismos vivos, atenuados o inactivados, que se administran a un individuo sano para que pueda desarrollar inmunidad contra alguna enfermedad infecciosa y, de acuerdo con Escobar, las vacunas representan el recurso más importante para la prevención de enfermedades. Históricamente las vacunas han estado presentes desde 1796 y desde entonces han demostrado ser altamente útiles en el control y disminución de las enfermedades para las cuales se han desarrollado, por lo cual, en los siglos siguientes se trabajó en mejorar el proceso necesario para obtener una vacuna nueva [19], [28].

La OMS describe que para poder desarrollar una vacuna es necesario primeramente identificar los antígenos necesarios para obtener la respuesta deseada, después se realiza una etapa de pruebas preclínicas utilizando animales como sujetos para demostrar la seguridad y respuesta inmunitaria de la vacuna. Al superar los protocolos anteriores, se comienza a probar la vacuna en seres humanos siguiendo tres fases: la primera fase consiste en aplicarla a un grupo reducido de pacientes jóvenes y sanos para confirmar los resultados obtenidos en la prueba preclínica y además determinar la dosis adecuada. La segunda fase considera un grupo mayor de sujetos los cuales deben presentar las características de las personas a las que se prevé inmunizar, además se incluye a un grupo de control compuesto con sujetos que no están dentro del grupo de interés para poder realizar una comparación más acertada del funcionamiento de la vacuna. La última fase contempla realizar las mismas pruebas a un grupo mucho más grande de sujetos y además comparar los resultados obtenidos con el de un grupo de control que no ha recibido la vacuna; es conveniente mencionar que tanto en esta fase como en la fase dos no se conoce a qué grupo pertenecen los sujetos que participan en el estudio, a esto se le denomina estudio doble ciego. Con los resultados obtenidos después de la tercera fase se realizan una serie de evaluaciones de seguridad y eficacia y en caso de ser acreditadas estas evaluaciones se aprueba la vacuna para su comercialización y depende ahora de cada país aprobarla de acuerdo con sus regulaciones [29].

En el caso de la vacunación de COVID-19 el protocolo anterior fue apresurado de tal manera que fue posible obtener diversas vacunas en menos de un año, mientras que por

lo regular este proceso puede tardar hasta 10 años. Esto se logró al acelerar el protocolo de tal forma que las fases 1, 2 y 3 fueron superpuestas y además se comenzó con la fabricación de estas antes de terminar la tercera fase (Fig. 5). Por lo anterior, las vacunas desarrolladas recibieron un permiso de emergencia el cual facilitó la fabricación y disponibilidad de estas a nivel mundial [29]. Es importante resaltar que a causa de este permiso el nombre adecuado es *vacuna de emergencia* y es la primera vez que se otorga un permiso así<sup>5</sup>.

Por lo general, el desarrollo de una vacuna es muy tardado debido a los **recursos limitados** y al **tiempo** que se necesita para preparar y obtener el financiamiento para realizar la siguiente fase.



En la búsqueda de una vacuna o anticuerpo contra la COVID-19, **no estamos tomando atajos ni omitiendo pasos**. Más bien estamos **superponiendo** algunos de los pasos en lugar de realizarlos uno detrás de otro.



Con el financiamiento del gobierno de los EE. UU., estas vacunas contra la COVID-19 se empezaron a fabricar en grandes cantidades **mientras** se realizaban los ensayos en fase 3. Esta fabricación fue un **riesgo** porque aún no sabíamos si las vacunas eran eficaces. Y este riesgo se asumió debido a la crisis de salud pública. Si los estudios muestran que las vacunas no son eficaces, estas se destruirán. En cada etapa del proceso de desarrollo de las vacunas, **nuestra preocupación principal es la seguridad del público**.

Fig. 5 Proceso de desarrollo de una vacuna normal contra el proceso seguido para las vacunas de COVID-19 [69]

<sup>5</sup> Al 31 de agosto del 2022 el New York Times reportaba que las vacunas Pfizer y Moderna ya cuentan con una autorización por parte de la FDA, mientras que Johnson & Johnson tiene una autorización limitada. Las vacunas SinoVac y Sinopharm han sido aprobadas en China; Sputnik V fue aprobada en Rusia, Novavax en Canadá y Corea del Sur; AstraZeneca y Covaxin han sido aprobadas en India. [34]

Para obtener dicho permiso de emergencia, fue necesario que los laboratorios demostraran que las vacunas presentaban una eficacia adecuada. En epidemiología se consideran dos conceptos para evaluar el desempeño de una vacuna: *eficacia* y *efectividad*. Ambos conceptos sirven para evaluar la capacidad protectora de una vacuna, sin embargo, la eficacia siempre se evalúa en ensayos clínicos aleatorios y la efectividad se obtiene después de comercializar la vacuna y realizar estudios observacionales del desempeño de la vacuna con la población general. Es importante considerar que un buen desempeño en la eficacia de los ensayos clínicos no implica que una vacuna tendrá un buen impacto en la población general [19].

Como resultado del proceso anterior, los laboratorios que pudieron comprobar más de un 50% de eficacia ante la ONU fueron autorizados para comercializar sus vacunas [30]. Actualmente, para la prevención de la COVID-19 se consideraron 3 tipos vacunas: virus inactivado, vector viral no replicante y ARN mensajero. En el caso de México están disponibles las siguientes vacunas:

Tabla 3. Lista de vacunas disponibles para México [31]–[34]

Farmacéutica	Fármaco	Nombre comercial	Plataforma de diseño	Eficacia
Pfizer-BioNTech	BNT162b2	Pfizer	ARN mensajero	95%
AstraZeneca	AZD1222	AstraZeneca	Vector viral no replicante	74% <sup>6</sup>
CanSino Biologics Inc.	Ad5-nCov	CanSino	Vector viral no replicante	58% <sup>7</sup>
Centro Nacional Gamaleya	Gam-COVID-Vac	Sputnik V	Vector viral no replicante	92% <sup>8</sup>
Sinovac Research and Development Co	CoronaVac	Sinovac	Virus inactivado	51% <sup>9</sup>
Bharat Biotech International Limited	BBV152 Covaxin	Covaxin	Virus inactivado	78%

<sup>6</sup> Eficacia contra COVID sintomático, 100% de eficacia contra COVID severo

<sup>7</sup> Eficacia contra COVID sintomático, 92% de eficacia contra COVID severo

<sup>8</sup> Contra variable alfa, para ómicron su eficacia disminuye a 75%

<sup>9</sup> Resultado de la prueba de eficacia en Brasil, en Turquía registró un 84% de eficacia

Janssen-Cilag	Ad26.COV2.S	Johnson & Johnson	Vector viral no replicante	72% <sup>10</sup>
Moderna	mRNA-1273	Moderna	ARN mensajero	94% <sup>11</sup>

Es necesario mencionar que, a pesar de ser aprobadas las vacunas continúan bajo una constante supervisión para poder llevar un registro de todos los efectos adversos que estas puedan causar. Las plataformas más relevantes para dar seguimiento a las vacunas son: *VigiAccess*, *EudraVigilance* y *VAERS*. La primera de estas plataformas corresponde a un sistema de vigilancia mundial dirigido por la OMS el cual contempla a más de 180 países y agrupa los reportes registrados según los efectos secundarios. La segunda concentra los reportes de eventos adversos de todas las vacunas aplicadas en la Unión Europea y algunas regiones de interés comercial. Finalmente, *VAERS* se encarga de recopilar los reportes de las diferentes campañas de vacunación en Estados Unidos y sus territorios.

### 3.3 Modelos de aprendizaje de máquina

Desde la revolución industrial la tecnología ha evolucionado a pasos agigantados lo cual ha favorecido enormemente el desarrollo de nuevas áreas del conocimiento. Una de estas es la *Inteligencia Artificial*.

Definir la inteligencia artificial ha representado un gran debate en todo el mundo, una de las primeras definiciones presentada por Alan Turing dice que “una máquina (computadora, robot o ente sintético) es inteligente si exhibe un comportamiento a nivel humano en una tarea o contexto dado” [35]; por otro lado, en 1956 John McCarthy establece el objetivo de la IA como “encontrar la forma de hacer que las máquinas utilicen el lenguaje, formen abstracciones y conceptos, resuelvan tipos de problemas ahora reservados a los humanos y que se mejoren a sí mismas” [36]. A pesar de que las definiciones mencionadas se refieren a diferentes características de la inteligencia humana, Turing habla de una capacidad más conductual mientras que McCarthy se refiere a una capacidad más cognitiva, ambas tienen en común que buscan imitar la inteligencia humana.

Dentro de las múltiples estrategias propuestas para resolver problemas usando IA surge en la década de los 80 el Aprendizaje de Máquina o Aprendizaje Automático (AM). Formalmente, el AM se puede definir como la adquisición de una representación útil y

<sup>10</sup> En pruebas realizadas en Estados Unidos, en Brasil tuvo un 68% de eficacia y en Sudáfrica un 64%

<sup>11</sup> Eficacia contra la enfermedad sintomática, para casos severos la eficacia aumenta a 98%

entendible de un conjunto de datos [37]; dicho de otra forma, es un enfoque de programación que ayuda a que las máquinas puedan aprender de forma similar a los humanos mediante el uso de la información disponible y una de sus características más importantes es que se busca desarrollar algoritmos que puedan ser aplicados para resolver diferentes problemas [38].

Existe una gran variedad de algoritmos que se clasifican, como se muestra en la Fig. 6, dependiendo de dos factores: 1) cómo se realiza el aprendizaje y 2) la aplicación que tiene el algoritmo. Para la clasificación por tipo de aprendizaje se consideran tres tipos:

- *Aprendizaje supervisado*: En este tipo se busca crear un modelo que pueda establecer la relación entre las entradas  $x \in X$  y las salidas  $y \in Y$ . Este tipo de aprendizaje es el más común y permite resolver problemas de clasificación y regresión [36].
- *Aprendizaje no supervisado*: A diferencia del tipo de aprendizaje anterior aquí se busca extraer información de los datos utilizando únicamente las entradas, esto es muy útil en situaciones en donde no existen datos de salida para el problema que se desea resolver. Algunos ejemplos de este tipo de aprendizaje son el agrupamiento de datos y reducción de la dimensionalidad [36].
- *Aprendizaje reforzado*: Es un aprendizaje que busca interactuar con su entorno y que a diferencia del primer tipo, implementa un sistema de recompensa-castigo para la toma de decisiones [36].

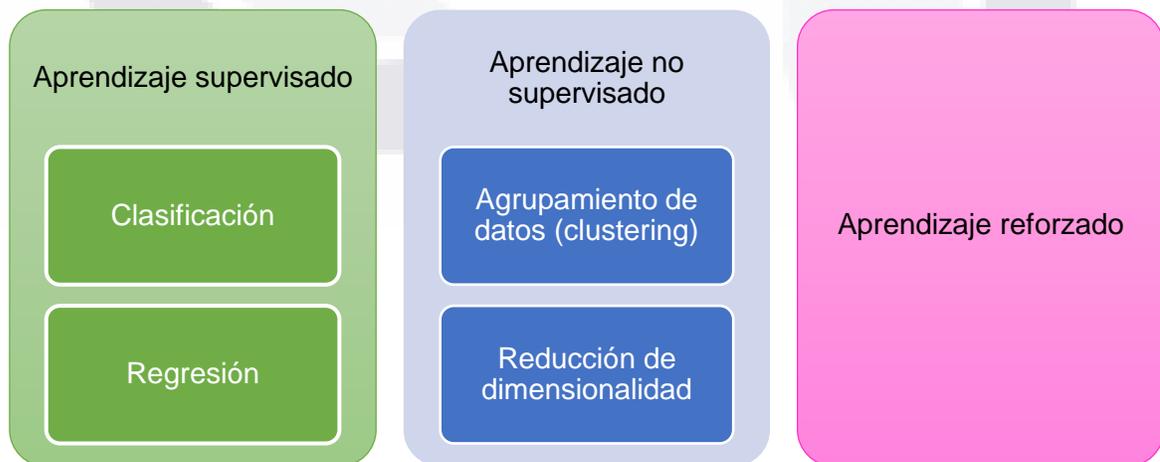


Fig. 6 Clasificación del aprendizaje automático

Ahora bien, para la clasificación por aplicación se consideran los dos siguientes tipos: predicción y clasificación. De forma general, un problema de predicción busca determinar el valor exacto de salida para una entrada que se tenga: por el contrario, un problema de clasificación busca determinar a la clase a la que pertenece un patrón de entrada [36].

### 3.3.1 Máquinas de soporte vectorial (SVM)

Este conjunto de métodos aplicables tanto a clasificación como regresión fueron introducidos en 1995 y debido a que tienen una base teórica bastante sólida han sido implementados en diversas áreas [39] principalmente para resolver problemas de clasificación [40].

En pocas palabras, para que estos métodos puedan resolver un problema es necesario mapear los puntos de entrada y después, con ayuda de dos vectores generados a través de las entradas, encontrar el hiperplano capaz de separar las clases deseadas considerando el margen más amplio entre ellas [40].

Formalmente las máquinas de soporte vectorial se basan en los conceptos de la teoría de aprendizaje estadístico, esto significa que, a lo que se llama aprendizaje es realmente la estimación de una función a través de un conjunto de datos de entrada, además, esta función debe cumplir con la condición de minimizar una determinada función de riesgo. Esta función de riesgo depende de la complejidad de las funciones que se utilizan así como de los datos de entrada [39].

Para el caso específico de la clasificación utilizando estos métodos se tienen dos enfoques: duro (*hard*) y suave (*soft*). Ambos enfoques siguen los mismos principios teóricos y mantienen las mismas funciones, no obstante, el segundo de ellos se considera más relajado dado que toma en cuenta los valores aislados o mal clasificados y, por tanto, este enfoque también resulta más realista [38]. En el caso duro, las ecuaciones que permiten obtener la función de costo son [38], [39]:

$$wx + b = 0 \quad w \in R^N, b \in R \quad \text{Ec. XIII}$$

$$wx + b \geq 1 \rightarrow x \in \text{clase1} \quad y = 1 \quad \text{Ec. XIV}$$

$$wx + b \leq -1 \rightarrow x \in \text{clase2} \quad y = -1 \quad \text{Ec. XV}$$

$$y \cdot (wx + b) \geq 1 \tag{Ec. XVI}$$

$$wx_2 + b = 1 \tag{Ec. XVII}$$

$$wx_1 + b = -1 \tag{Ec. XVIII}$$

Ahora bien, para entender de donde se obtienen las ecuaciones es necesario recurrir a las siguientes ilustraciones. En la Fig. 7 se observan dos clases que pueden ser fácilmente separadas por diferentes hiperplanos de la forma dada en Ec. XIII, como se muestra en la Fig. 8, no obstante, solo uno de estos puede ser considerado ideal.

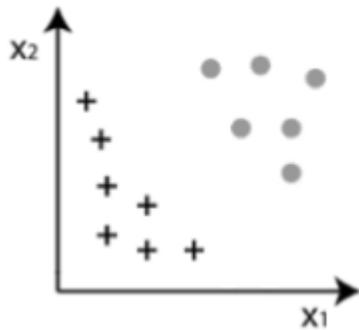


Fig. 7 Caso de dos clases separables [38]

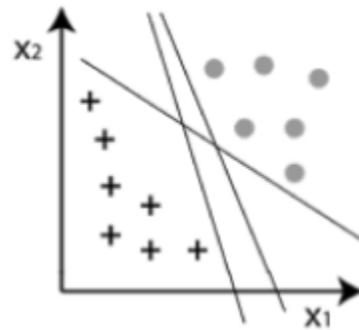


Fig. 8 Posibles hiperplanos de separación [38]

Para poder determinar cuál de todos los posibles hiperplanos es el adecuado se recurre a los vectores de soporte. Este concepto se refiere a las muestras en la frontera de los grupos que permiten trazar dos hiperplanos, uno positivo y uno negativo que cumplen con las Ec. XVII y XVIII respectivamente. En la Fig. 9 se muestran en el caso bidimensional.

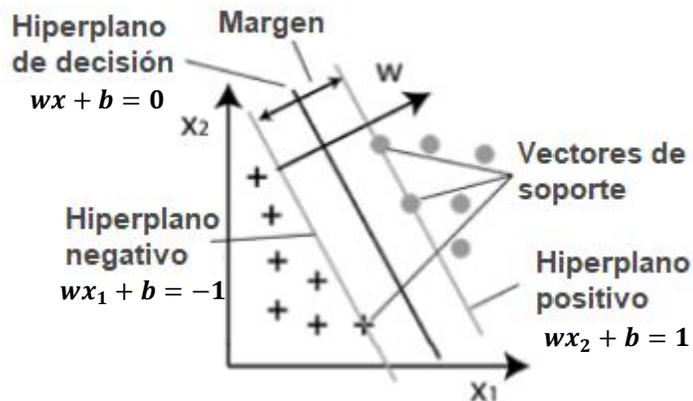


Fig. 9 Clasificación de las dos clases considerando los vectores de soporte [38]

El identificar tanto los vectores de soporte como los hiperplanos positivo y negativo es fundamental ya que el hiperplano de decisión ideal es aquel que maximiza el margen entre los hiperplanos [38]. Dado que el margen se puede definir como la distancia entre dos puntos se puede obtener la función a maximizar a través de las Ec. XVII y XVIII de la siguiente forma [38]:

$$(wx_2 + b) - (wx_1 + b) = 1 - (-1) \tag{Ec. XIX}$$

$$w(x_2 - x_1) = 2 \tag{Ec. XX}$$

$$\text{Margen} = x_2 - x_1 = \frac{2}{\|w\|} \tag{Ec. XXI}$$

$$\max\left(\frac{2}{\|w\|}\right) \approx \min\left(\frac{1}{2}\|w\|^2\right) \tag{Ec. XXII}$$

Finalmente, la ecuación de costo necesaria para poder realizar el aprendizaje se obtiene a partir de Ec. XVI y XXII [38]:

$$\text{Loss} = \min\left(\frac{1}{2}\|w\|^2\right), \quad \text{subject to } y \cdot (wx + b) \geq 1 \tag{Ec. XXIII}$$

Por sus características teóricas, los métodos de SVM demuestran un desempeño excelente en problemas pequeños, lo cual ha sido aplicado ampliamente en el campo médico en donde muchas veces obtener datos resulta complicado y caro [41].

### 3.3.2 Árboles de decisión

Los modelos de árboles de decisión, al igual que las máquinas de soporte vectorial, son métodos utilizados principalmente para la clasificación de datos en el aprendizaje automático y tienen la característica de ser de caja blanca por lo que explicar su comportamiento y resultado es muy sencillo. La principal característica de estos modelos es que son fundamentales para el desarrollo de los modelos de *Bosques Aleatorios* [38].

Los métodos de árboles de decisión se forman a partir de la idea de dividir los datos recursivamente siguiendo la forma de un árbol jerarquizado en donde, partiendo de un nodo raíz, se espera llegar a una hoja que pueda generalizar algún caso específico del problema [42]. Por lo anterior, este es un modelo que entrega un resultado de forma gráfica y que permite llegar a una decisión final a través de las hojas de este [43].

Las partes que conforman el gráfico se muestran en la Fig. 10. En todos los casos se observa que ninguna rama llega al nodo raíz mientras que a los demás nodos llega exactamente una rama [42].

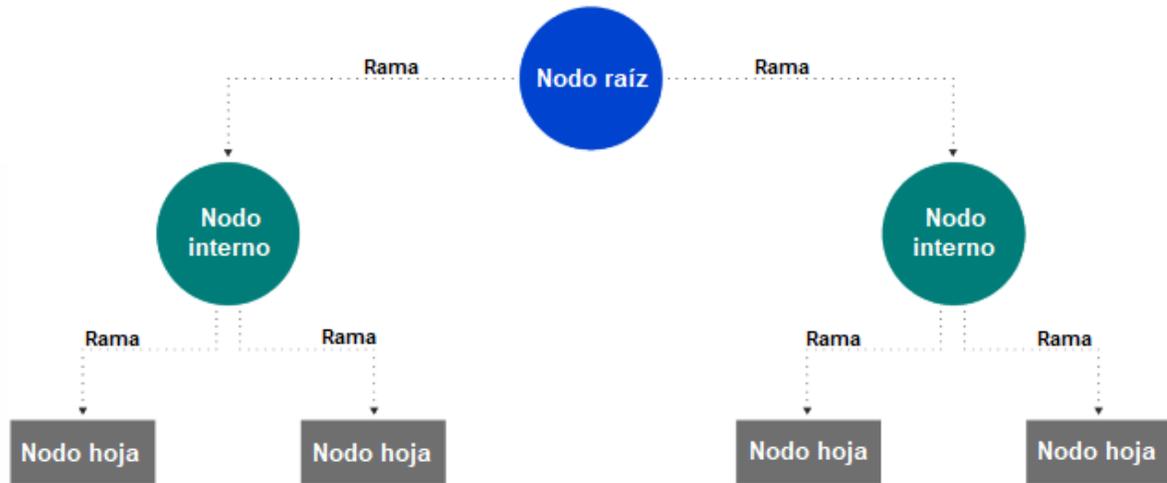


Fig. 10 Partes de un gráfico de árbol de decisión [70]

Para realizar las particiones de los datos, se selecciona primero una variable la cual se ramifica de acuerdo con el valor de los atributos seleccionados o, en casos numéricos, según los rangos [42]. A estos resultados siguen más ramificaciones hasta llegar a una hoja. En todos los casos la entrada de información está dada por una sentencia descrita por un conjunto de atributos y tiene una salida binaria del tipo verdadero o falso [43].

Para poder definir la variable que va a ser particionada se recurre al concepto de entropía. Este concepto describe la tendencia al orden en los datos y tiene un valor entre 0 y 1, donde el primero indica orden total y el segundo desorden total. Formalmente, se define como el valor esperado de la información y se obtiene mediante la siguiente ecuación [43]:

$$H = - \sum_{i=1}^n p(X_i) \log_2 p(X_i) \tag{Ec. XXIV}$$

Donde  $\log_2 p(X_i)$  indica el valor de la información y  $p(X_i)$  es la probabilidad de escoger una clase. Mediante la Ec. XXIV se busca que la partición realizada pueda crear un grupo más ordenado de datos que el grupo anterior [43].

Todo lo anterior convierte al algoritmo de árboles de decisión en un recurso muy útil, no obstante, tiene la desventaja de ser inestables y sensibles a la rotación de la información

[38], además de tener un alto riesgo de caer en el sobreajuste de la información. Algunas posibles soluciones para mejorar el desempeño de los árboles de decisión son: trabajar con conjuntos de datos equilibrados y recurrir a las diferentes técnicas disponibles para evitar el sobreajuste [43].

### **Poda del árbol de decisión**

Como se mencionó antes, estos métodos tienen la desventaja de caer con facilidad en un sobreajuste y dar como resultado reglas que solamente se ajustan a un número muy reducido de muestras. Para evitar este problema se busca tener un menor número de hojas que contengan a grupos de datos más grandes para poder tener una mejor generalización; para esto, se utilizan diferentes métodos denominados poda [38].

Para esto, algunas de las estrategias implementadas en Python son las siguientes: definir la profundidad máxima del árbol, establecer un mínimo de datos para poder particionar un nodo o para permitir que se cree una hoja y definir el número máximo de hojas [38].

### **3.3.3 Bosques aleatorios**

Al igual que los métodos anteriores, los Bosques Aleatorios son métodos de Aprendizaje de Máquina utilizados principalmente para la clasificación de datos, no obstante, a este se le denomina método de ensamble debido a que implementa el funcionamiento del método de árbol de decisión para funcionar [38].

El objetivo de implementar árboles de decisión en este método es obtener una mayor diversidad de resultados y encontrar la mejor solución posible, esto se logra alimentando a diferentes árboles con subconjuntos de los datos diferentes y aleatorios [38].

### **3.3.4 Preprocesamiento de los datos**

Al igual que todos los métodos de aprendizaje de máquina que hay, los tres anteriores necesitan que los datos de entrada cumplan con una serie de características para evitar que se produzcan errores al obtener los resultados.

Cuando los datos utilizados para trabajar con estos métodos no son los adecuados el aprendizaje sencillamente no se realiza o entrega resultados que no representan la realidad. Estos problemas se pueden presentar por diversas situaciones como son: tener datos incorrectos o incompletos, una cantidad excesiva o insuficiente de ellos y un número desigual de datos en las clases a las que pertenecen; incluso aunque no se presenten los

casos anteriores es fundamental asegurar que la información se de calidad y correcta para resolver el problema [38].

Algunas de estas situaciones se pueden resolver fácilmente al realizar unas pequeñas modificaciones en la información antes de que entren al modelo de aprendizaje como eliminar datos y variables inútiles, convertir los datos a valores numéricos de preferencia entre 0 y 1 [44], además de otras técnicas más complejas para situaciones específicas como las que se necesitan para solucionar el problema de la desigualdad de clases.

La desigualdad de clases es una dificultad frecuente al enfrentarse a problemas de clasificación del mundo real, especialmente en el área médica. Es fundamental evitar esta situación ya que afecta enormemente a los algoritmos de aprendizaje debido a que su aprendizaje se basa en un valor de error global y en algunos el clasificar a todos los datos como una sola clase no suele generar un error alto. Para poder tratar con el problema de tener clases desiguales se puede recurrir a dos técnicas: aumentación (*oversampling*) o reducción (*under sampling*) de datos [45].

De los dos casos anteriores la técnica de reducción de datos es la más sencilla ya que, consiste en conservar aleatoriamente solo una parte de los datos de la clase mayoritaria. Por el otro lado, para implementar la aumentación de datos es necesario agregar datos a la clase más reducida duplicando los patrones observados en dicha clase o los valores positivos en el límite de clasificación [45]. Seleccionar la técnica correcta dependerá de las características de la información y del problema que se desea resolver.

### **3.3.5 Mediciones para la calidad del modelo**

Una vez desarrollado el modelo de aprendizaje de máquina es necesario realizar una evaluación de sus resultados que nos permita conocer la calidad del aprendizaje que ha tenido el modelo. Para esto se recurre al uso de los conceptos de exactitud, precisión, exhaustividad, valor F1 y de un gráfico denominado matriz de confusión.

Para poder determinar la calidad del aprendizaje hay que evaluar los conceptos anteriores con diferentes conjuntos de datos para asegurar que el modelo es capaz de trabajar con datos diferentes a los que fueron utilizados para su creación. Para esto, los datos originales se pueden dividir en tres grupos: entrenamiento, validación y prueba [38].

El grupo de datos de entrenamiento es el más grande de los tres, por lo general abarca entre el 75% y el 80% de los datos originales. Este conjunto de datos tiene como objetivo

entrenar al modelo. Por otro lado, los datos pertenecientes al grupo de prueba corresponden al 25% o 20% no utilizados para el entrenamiento y son fundamentales para la evaluación de la calidad del modelo. Finalmente, los datos de validación son muestras utilizadas para realizar ajustes finos a los parámetros del modelo [38].

Con los datos de prueba definidos estos son introducidos al modelo de aprendizaje y a través de sus resultados es posible crear una matriz que indique la cantidad de veces que el algoritmo entregó un dato acertado. Esta matriz, denominada formalmente como matriz de confusión, se puede crear tanto para los problemas de clasificación y regresión; en el caso de los problemas de clasificación un resultado acertado significa que el algoritmo fue capaz de asignarle la clase correcta al dato analizado [38].

En la Tabla 4 se muestra un ejemplo de una matriz de confusión para un problema de clasificación. Los valores contenidos en esta matriz son de gran importancia ya que, a través de ellos se realizan los cálculos para conocer la calidad del aprendizaje [38].

Tabla 4. Estructura de una matriz de confusión sin importar el tipo de problema [38]

		Realidad	
		Clase positiva	Clase negativa
Predicción	Clase positiva	Verdadero positivo ( <i>TP</i> )	Falso negativo ( <i>FN</i> )
	Clase negativa	Falso positivo ( <i>FP</i> )	Verdadero negativo ( <i>TN</i> )

La primera medida de calidad del aprendizaje es la exactitud (*accuracy* en inglés, *ac*) y para poder utilizarse las clases deben estar equilibradas. Esta medida permite conocer el porcentaje de casos que ha clasificado correctamente el modelo y su valor se obtiene de la siguiente forma [38]:

$$ac = \frac{TP + TN}{TP + TN + FP + FN} \tag{Ec. XXV}$$

La precisión (*precision* en inglés, *pr*) y la exhaustividad (*recall* en inglés, *re*) son proporciones que indican la cantidad de muestras clasificadas correctamente, enfocándose en las predicciones realizadas (Ec. XXVI) y las muestras positivas disponibles respectivamente (Ec. XXVII). Es fundamental considerar que, a pesar de que estos

conceptos son muy parecidos tienen propósitos diferentes, es deseable que el valor de precisión sea elevado cuando se realizan muchas predicciones mientras que, un valor alto de exhaustividad es necesario si hay pocas muestras positivas [38].

$$pr = \frac{TP}{TP + FP} \quad \text{Ec. XXVI}$$

$$re = \frac{TP}{TP + FN} \quad \text{Ec. XXVII}$$

Finalmente, el valor F1 (en inglés *F1 score*) combina los dos conceptos anteriores y permite comparar prácticamente el desempeño de ambos mediante un solo valor. Si los valores de precisión y exhaustividad son altos el valor F1 también lo será, sin embargo, si uno de ellos es bajo esto se podrá observar en el resultado de este valor, esta relación se muestra en la Ec. XXVIII [38].

$$F1 = 2 \cdot \frac{pr \cdot re}{pr + re} \quad \text{Ec. XXVIII}$$

## 4 Metodología

Para cumplir con los objetivos establecidos del proyecto fue necesario recolectar los datos disponibles y procesarlos para resolver las preguntas de investigación planteadas al inicio. En este capítulo se explicará detalladamente el proceso seguido para obtener las diferentes gráficas utilizadas para analizar el comportamiento que presentó la pandemia en México y algunos países de interés; igualmente se incluyen los pasos seguidos para la comparación de tres algoritmos de aprendizaje de máquina para seleccionar el más adecuado para la construcción de un modelo que permita predecir si una persona va a fallecer o no por COVID-19 de acuerdo con sus factores de riesgo. Finalmente se muestra el proceso de investigación para obtener información sobre la campaña mundial de vacunación para llegar a una conclusión acerca del impacto que esta tuvo.

### 4.1 Preprocesamiento de los datos

Para resolver la pregunta de investigación respecto al comportamiento de la pandemia se recolectó la información necesaria de dos bases de datos principales: *Datos Abiertos y Coronavirus Pandemic (COVID-19)*. La primera corresponde a la base de datos publicada por primera vez el 14 de abril del 2020 por la Dirección General de Epidemiología del Gobierno de México, la cual es actualizada diariamente y muestra la información específica de cada caso, sea positivo o no, asociado a la pandemia. Al 22 de noviembre del 2022 esta base de datos acumuló un total de 18,717,783 registros para los cuales hay información de 40 variables. Para poder trabajar con esta base de datos fue necesario recurrir a los documentos de *201128 Descriptores\_* y *201128 Catálogos*, los cuales se encargan de explicar el significado de cada variable y cómo interpretar cada uno de los registros almacenados respectivamente.

Por otro lado, la segunda base de datos se publicó por parte del proyecto *Our World in Data (OWD)* perteneciente a la organización *Global Change Data Lab (GCDL)* y se muestran 67 variables generales para más de 200 países. Estos datos fueron recolectados de diversas fuentes siendo el Instituto Johns Hopkins y los reportes oficiales de cada gobierno las principales. Al 1 de diciembre del 2022 se tenían 238,967 registros de los cuales 1,065 son exclusivos de México. Es conveniente resaltar tres puntos: 1) cada registro corresponde a un día específico en alguno de los países para los que se recolectan datos, 2) para algunos países no hay información disponible de todas las variables que se recopilan y 3) algunas

variables se han dejado de actualizar. A diferencia de la base de datos anterior, la información contenida aquí aparece de forma clara por lo que no es necesario un documento que indique cómo se debe interpretar, aun así, en la plataforma de *GITHUB* se muestra la descripción detallada de las variables utilizadas.

Para el caso de la campaña de vacunación se observó que a nivel nacional no había información pública por lo que fue necesario realizar una solicitud de acceso a la información, además de analizar los datos publicados en diversos artículos y en los sistemas de vigilancia *VAERS*, *EudraVigilance* y *VigiAccess*. Como resultado de la solicitud con folio 330026923002342 realizada en la Plataforma de Transparencia para que la Secretaría de Salud<sup>12</sup> compartiera los datos disponibles se obtuvo una base de datos considerablemente más pequeña que las anteriores, 5 variables y 20,750 registros, que contiene la información referente a las defunciones por COVID-19 de personas que recibieron alguna vacuna en México.

Dado que las bases de datos principales son diferentes fue necesario identificar la aplicación que se le puede dar a cada una y, definir la información relevante que pueden entregar. La base de datos publicada por el gobierno de México se utilizó para la creación y entrenamiento de un modelo que ayude a determinar si un individuo enfermo de COVID-19 va a fallecer o no, mientras que, la base de *OWID* brindó información para comparar características generales entre algunos países de interés. Para todas las bases de datos utilizadas fue indispensable aplicar filtros para eliminar variables no relevantes y registros que tuvieran información incorrecta o incompleta, además se ajustaron los valores registrados para asegurar que estos se encontraran en el rango deseado. Finalmente, en el caso de la base de *OWID* se crearon algunas variables.

De forma específica, para la base de Datos Abiertos se conservaron 18 variables de las 40 que se tenía inicialmente y 7,003,759 registros de los 18,717,783 originales. Para el caso de las otras 23 variables es importante resaltar que dos de ellas no se eliminaron inmediatamente ya que fueron de utilidad para realizar la reducción del número de registros. Las variables utilizadas como filtro que después se eliminaron fueron: 1) *CLASIFICACION\_FINAL*, permitió conservar únicamente los registros de los casos que resultaron positivos a COVID-19 y 2) *NACIONALIDAD*, mediante esta variable se aseguró que los registros pertenezcan a personas de nacionalidad mexicana. Este proceso de

---

<sup>12</sup> Los documentos entregados se incluyen en la sección de anexos

filtrado tuvo un impacto muy significativo en la cantidad de registros ya que, solamente filtrando la variable de *CLASIFICACION\_FINAL* los datos se redujeron aproximadamente al 38%<sup>13</sup>.

Los siguientes filtros utilizados se realizaron mediante las variables que se conservaron para la realización del modelo. Para el primer filtro se utilizó la variable de *EDAD* y se fijó el límite en 105 años debido a que se observó se conserva una buena cantidad de datos a comparación de otros límites. El segundo filtro se elaboró utilizando las variables de comorbilidades, para esto, se eliminaron los registros en los cuales se ignora si la persona a la que pertenece el registro padece alguna de las comorbilidades registradas. Por último, se quitaron los registros que en las variables de *NEUMONIA*, *INTUBADO*, *UCI* y *TIPO\_PACIENTE* aparecieran como *no especificado*.

Para este punto se conservaron poco más de 7 millones de datos y solamente fue necesario modificar los valores registrados para poder trabajar con ellos adecuadamente. Para esto las variables correspondientes a comorbilidades, *UCI*, *INTUBADO*, *NEUMONIA*, *TIPO\_PACIENTE*, *SEXO* y *FECHA\_DEF* se convirtieron a valores binarios de 0 y 1 donde representan una negación y una afirmación respectivamente; por ejemplo, si un registro corresponde a una persona que sobrevivió se observa un 1 en la variable *FECHA\_DEF*, de lo contrario aparece un 0. En el caso específico de la variable *SEXO*, el 1 indica que el registro pertenece a una persona de sexo masculino, mientras que el 0 identifica los registros de las personas de sexo femenino. Tanto las variables que se conservaron como las que se utilizaron para filtro se muestran en la Tabla 5.

Tabla 5. Variables consideradas para la creación del modelo de aprendizaje de máquina

Datos Abiertos	
Variable filtro	Descripción
CLASIFICACION_FINAL	El diagnóstico que recibió la persona
NACIONALIDAD	Nacionalidad de la persona
Variable	Descripción
SEXO	Sexo de la persona
TIPO_PACIENTE	Si requirió hospitalización
FECHA_DEF	Si la persona sobrevivió
INTUBADO	Si la persona requirió intubación
NEUMONIA	Si enfermó de neumonía
EDAD	Edad de la persona

<sup>13</sup> La cantidad de registros original era de 18,717,783 y después del primer filtro se redujo hasta 7,125,176

EMBARAZO	Si la persona estaba embarazada
DIABETES	Si la persona tiene diabetes
EPOC	Si la persona tiene EPOC
ASMA	Si la persona tiene asma
INMUSUPR	Si la persona tiene inmunosupresión
HIPERTENSION	Si la persona tiene hipertensión
OTRA_COM	Si la persona tiene alguna otra comorbilidad
CARDIOVASCULAR	Si la persona tiene una enfermedad cardiovascular
OBESIDAD	Si la persona tiene obesidad
RENAL_CRONICA	Si la persona tiene enfermedad renal crónica
TABAQUISMO	Si la persona es fumadora
UCI	Si requirió ingresar a la unidad de cuidados intensivos

En el caso de la base de datos de *OWID* la información exclusiva de México se obtuvo utilizando la variable *location* como filtro, se eliminaron 49 variables y se crearon otras dos: *mortalidad* y *letalidad*. Las variables que se conservaron al final de los cambios anteriores aparecen en la Tabla 6 y estas permitieron crear gráficas para realizar comparaciones entre varios países de interés.

Tabla 6. Variables consideradas para la comparación de países de interés

OUR WORLD IN DATA (OWID)	
Variable (originales)	Descripción
date	Fecha del registro
location	Nombre del país en inglés
iso	Código ISO del país
total_cases	Casos acumulados a la fecha
new_cases	Casos nuevos reportados
total_deaths	Muertes acumuladas a la fecha
new_deaths	Nuevas muertes reportadas
total_tests	Total de pruebas aplicadas
tests_per_case	Número de pruebas para encontrar un positivo
total_vaccinations	Total de vacunas aplicadas
people_vaccinated	Personas que han recibido al menos una dosis
people_fully_vaccinated	Personas que completaron el esquema original
total_boosters	Total de refuerzos aplicados
stringency_index	Índice de restricción
population	Población total
median_age	Media de edad de la población
cardiovasc_death_rate	Índice de mortalidad por e. cardiovascular

diabetes_prevalence	Prevalencia de diabetes
<b>VARIABLES (NUEVAS)</b>	<b>DESCRIPCIÓN</b>
Mortalidad	Valor de mortalidad específica, se interpreta como las defunciones causadas por COVID-19 por 100,000 habitantes
Letalidad	Porcentaje de letalidad. Indica el porcentaje de personas contagiadas que sobreviven a la enfermedad.

Al igual que con la base de datos publicada por el gobierno de México fue necesario modificar los registros de algunas variables para poder trabajar adecuadamente con ellos. Las modificaciones que sufrió cada variable se muestran en la Tabla 7.

Tabla 7. Variables que fueron modificadas para realizar la graficación

<b>OUR WORLD IN DATA (OWID)</b>	
<b>VARIABLE</b>	<b>MODIFICACIÓN</b>
total_cases	Se conservó el último registro
total_deaths	Se conservó el último registro
total_tests	Se conservó el último registro
tests_per_case	Se obtuvo el promedio de todos los registros
total_vaccinations	Se modificó este valor para que representara el total de vacunas aplicadas por cada 100,000 habitantes
people_vaccinated	Se modificó para que representara a las personas que han recibido al menos una dosis por cada 100,000 habitantes
people_fully_vaccinated	Se modificó para que representara a las personas que completaron el esquema original por cada 100,000 habitantes
stringency_index	Se obtuvo el promedio y el valor máximo
population	Se conservó el último registro
median_age	Se conservó el último registro
cardiovasc_death_rate	Se conservó el último registro
diabetes_prevalence	Se conservó el último registro

Para obtener las variables nuevas se aplicaron las ecuaciones correspondientes a la mortalidad específica (Ec. VIII) y letalidad (Ec. VI) mencionadas con anterioridad. En el caso de la mortalidad específica fue necesario sustituir los valores contenidos en las variables *total\_deaths* y *total\_population* en  $D_E$  y  $P_R$  respectivamente, además de definir 5 como

múltiplo de 10 para tener un resultado que se ajuste a 100,000 habitantes. Luego de realizar la sustitución anterior la ecuación, el resultado para México se ve de la siguiente forma:

$$M_E = \frac{330,495}{127,504,120} (10^5) = 259.2 \quad \text{Ec. XXIX}$$

Para obtener el porcentaje de letalidad la sustitución se realizó considerando los valores de las variables *total\_deaths* y *total\_cases* en  $D_E$  y  $P_E$  respectivamente. El resultado obtenido en el caso de México se muestra en Ec. XXX.

$$L(\%) = \frac{330,495}{7,125,176} \times 100 = 4.64 \quad \text{Ec. XXX}$$

Después de obtener las variables anteriores para todos los países incluidos en la base de datos de *OWID* fue posible definir una lista de países de interés para comparar con México, de igual forma, en la base de datos se incluye información por región, la cual también fue útil para realizar algunas comparaciones. Los países y regiones seleccionados se listan en la Tabla 8.

Tabla 8. Países y regiones de interés consideradas para realizar las comparaciones

Países y regiones de interés	
Código ISO	Regiones
OWID_AFR	África
OWID_ASI	Asia
OWID_EUR	Europa
OWID_NAM	América del Norte
OWID_SAM	América del Sur
OWID_OCE	Oceanía
OWID_WRL	Mundial
Código ISO	Países
COD	República Democrática del Congo
EGY	Egipto
ETH	Etiopía
KEN	Kenia
MAR	Marruecos
NGA	Nigeria
SDN	Sudán
SOM	Somalia
TZA	Tanzania
YEM	Yemen

ZAF	Sudáfrica
CHN	China
IND	India
ISR	Israel
JPN	Japón
KOR	Corea del Sur
SYR <sup>14</sup>	Siria
BGR	Bulgaria
BIH	Bosnia y Herzegovina
DEU	Alemania
ESP	España
FRA	Francia
GBR	Reino Unido
GRC	Grecia
HUN	Hungría
ISL	Islandia
ITA	Italia
SWE	Suecia
CAN	Canadá
MEX	México
USA	Estados Unidos de América
ARG	Argentina
BRA	Brasil
CHL	Chile
COL	Colombia
ECU	Ecuador
PER	Perú
PRY	Paraguay
SLV	El Salvador
AUS	Australia
NZL	Nueva Zelanda

Las comparaciones se realizaron por tres etapas: la primera se realizó de forma general y permitió determinar los países de interés al comparar los valores de la letalidad y mortalidad, en la segunda se compararon algunas variables que muestran el desarrollo que tuvo la pandemia en los países seleccionados y en la última etapa se compararon las variables de

---

<sup>14</sup> En la base de datos de OWID se registró este código de manera incorrecta como SIR.

*population, median\_age, cardiovasc\_death\_rate, diabetes\_prevalence* lo que permitió conocer las diferencias presentes en las poblaciones de estos países.

Por último, los cambios realizados a la base de datos de personas vacunadas que fallecieron fueron más sencillos que los cambios de las bases de datos anteriores. Únicamente se definió el límite de edad en 105 años, el cual eliminó 4 registros y de las 5 variables originales se eliminaron *EVOLUCI* y *VACUNA\_COV* debido a que todos los registros contenían la misma información: la evolución fue negativa ya que todos fallecieron y todos completaron el esquema inicial de vacunación. Las variables que se conservaron, así como su descripción se muestran en la Tabla 9.

Tabla 9. Variables consideradas para el análisis de los resultados de la campaña nacional de vacunación

Personas vacunadas que fallecieron	
Variable	Descripción
SEXO	Si la persona era de sexo masculino o femenino
EDAD	La edad de la persona
TIPO_VAC_COV	Tipo de vacuna que recibió la persona

## 4.2 Estadísticas y gráficas

Para la construcción de las gráficas de comparación y estadística se utilizaron las tres bases de datos anteriores, además del software estadístico *R*. Para el caso de las gráficas de comparación se utilizó la base de datos de *OWID* y se consideraron las siguientes variables: *mortalidad, letalidad, total\_cases, total\_deaths, total\_test, tests\_per\_case, total\_vaccinations, people\_vaccinated, people\_fully\_vaccinated, stringency\_index, population, median\_age, cardiovasc\_death\_rate, diabetes\_prevalence*.

Para poder realizar las comparaciones se construyeron dos tipos diferentes de gráficos: de dispersión y barras. Las gráficas de dispersión se construyeron para las variables de *mortalidad* y *letalidad* considerando a todos los países del mundo, además de las regiones de la Tabla 8. Para las gráficas de barras se consideraron todas las variables mencionadas anteriormente y se elaboraron únicamente para los países de interés.

Para las gráficas estadísticas se consideraron únicamente datos nacionales provenientes de variables de la base de datos de gobierno. En este caso las variables utilizadas fueron las que se incluyen en la Tabla 5 y los gráficos construidos fueron de barras (histograma para la edad).

Finalmente, para conocer las estadísticas de defunciones por COVID-19 de personas vacunadas se utilizó la base de datos proporcionada por la Secretaría de Salud después de la solicitud realizada y se construyeron gráficas de barras para conocer las estadísticas de defunciones por sexo, grupo de edad y tipo de vacuna que recibieron las personas.

### 4.3 Selección del algoritmo para el modelo

Para desarrollar el modelo que predijera las características de las personas que fallecen por COVID-19 se realizó una comparación entre tres posibles algoritmos para seleccionar el que presentara el mejor desempeño. Los posibles algoritmos evaluados fueron: máquinas de soporte vectorial (*SVM*), árboles de decisión y bosques aleatorios. Después de realizar pruebas para determinar cuál presentaba mejores resultados, tanto para resolver las preguntas de investigación como mejor desempeño computacional, se seleccionó uno y se trabajó en el desarrollo del modelo final. Cabe resaltar que para los tres casos se implementaron los algoritmos de aprendizaje supervisado para clasificación de la biblioteca *scikit-learn* para *Python*.

Para realizar la comparación entre los algoritmos se trabajó exclusivamente con los datos nacionales publicados por el gobierno de México (*Datos Abiertos*) hasta el 22 de noviembre del 2022. Es fundamental mencionar que la base de datos presenta el problema de tener clases altamente desiguales (6,681,331 sobrevivientes vs 322,428 fallecidos) por lo que fue necesario aplicar alguna técnica que permitiera solucionar este problema para que los algoritmos pudieran realizar el aprendizaje. Debido al gran tamaño de la base de datos se optó por una reducción del tamaño de la clase *SOBREVIVIENTE*.

Esta reducción se realizó aplicando un filtro sobre la variable *FECHA\_DEF* al eliminar aleatoriamente registros que tuvieran un valor de 1 hasta que las clases tuvieran el mismo tamaño. Esta reducción de la información dio como resultado una matriz reducida de tamaño  $644,856 \times 18$ , como se muestra en la Fig. 11.

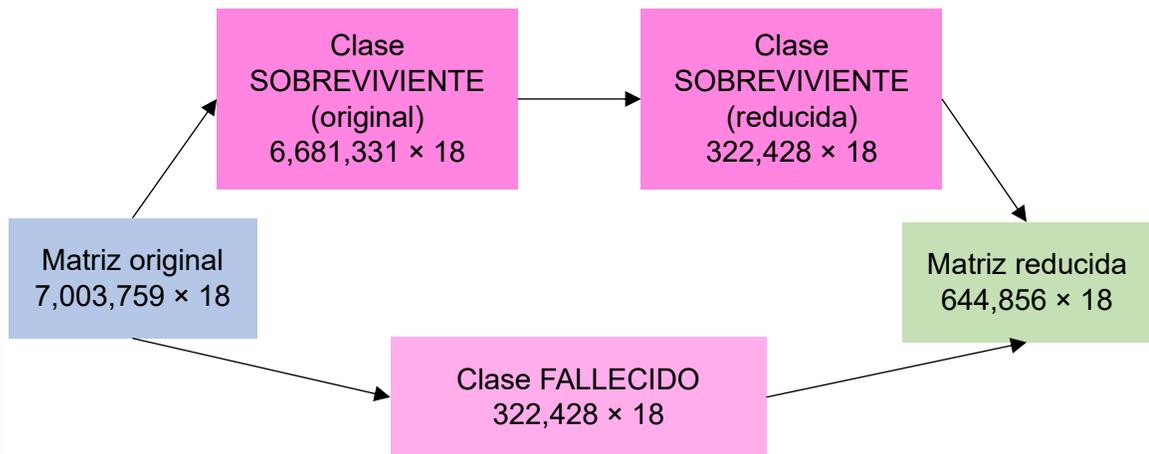


Fig. 11 Proceso de reducción de la matriz de datos original para eliminar las clases desiguales

Una vez solucionado el problema de la desigualdad de clases se procedió a realizar la comparación entre los algoritmos. Para comparar los algoritmos se desarrolló un modelo considerando únicamente dos variables de la matriz reducida y diferentes porcentajes de datos; las variables seleccionadas fueron *SEXO* y *EDAD*, la variable *FECHA\_DEF* se estableció como objetivo y para los datos se consideró el 1%, 2%, 10%, 25%, 50% y 100%. Esta segunda reducción en el tamaño y variación del porcentaje de información de la matriz se hizo con el objetivo de que los tres algoritmos terminaran de ejecutarse para poder comparar sus resultados de aprendizaje y observar cómo evolucionaba su desempeño computacional.

Después de comparar el desempeño de los algoritmos se determinó que la mejor opción era trabajar con árboles de decisión. Esto debido a que ese algoritmo presentó el mejor tiempo de ejecución de los tres y la diferencia en los resultados de aprendizaje entre los algoritmos no era significativa, además, brinda un apoyo visual muy útil para entender y difundir los resultados que entrega el modelo.

#### 4.4 Construcción del modelo

Al igual que para la comparación de los algoritmos para la creación del modelo se implementó el uso de *scikit-learn* y se mantuvo la clase objetivo, no obstante, se agregaron más variables hasta llegar al mejor resultado. Esta biblioteca se implementó dado que tiene la característica de poder incluir las técnicas de poda necesarias para evitar el sobreajuste del modelo y, mediante el paquete *graphviz* se puede tener la imagen correspondiente al

árbol. Estas características resultan muy útiles para poder tener un modelo comprensible y que realmente se ajuste a la realidad.

Al revisar la documentación correspondiente a la biblioteca de *scikit-learn* se pueden observar las características que puede tener el modelo construido. En este caso, las características más relevantes son: *max\_depth*, *min\_samples\_split*, *min\_samples\_leaf*, *criterion*, *splitter*.

Las características de *max\_depth*, *min\_samples\_split* y *min\_samples\_leaf* son muy relevantes ya que permiten realizar la poda del árbol. El primer parámetro es un valor entero que indica la profundidad que va a tener el árbol, por lo que es fundamental que sea un valor adecuado; en caso de no incluirlo el árbol crece hasta que las hojas contienen únicamente muestras de una clase, se cumpla el mínimo de muestras necesarias para realizar una división o se tenga el mínimo de muestras para una hoja. Las últimas dos condiciones se controlan mediante *min\_samples\_split* y *min\_samples\_leaf* respectivamente [46].

Para seleccionar el valor adecuado de profundidad para el árbol se realizó un ejercicio en donde se generaron 20 modelos diferentes en donde se ignoraron los parámetros de *min\_samples\_split* y *min\_samples\_leaf* y se modificó *max\_depth* para que tomara valores entre 5 y 100. Después, se graficó el valor de exactitud para cada uno de estos modelos y se evaluó cuál era el mejor valor para este parámetro.

Seleccionar el valor mínimo de muestras para cada partición fue más sencillo ya que debe ser el doble del valor asignado al mínimo de muestras en una hoja. Para el mínimo de datos en una hoja se decidió que fuera aproximadamente 1% de los datos, por lo que los valores asignados son 6,450 y 12,900 para las hojas y nodos.

El parámetro de *criterion* se utiliza para seleccionar el criterio que se va a utilizar para evaluar la calidad que tiene la partición hecha por el árbol de decisión. En este caso, se puede elegir entre tres posibles criterios: *gini*, *entropy* y *log\_less* [46].

Finalmente, el parámetro de *splitter* indica la estrategia que se utilizará para realizar la partición de los datos y se puede elegir entre *best* y *random*. El primero de estos valores asegura que la mejor partición sea seleccionada mientras que el segundo escoge la mejor partición aleatoria [46].

Para los dos últimos parámetros los valores seleccionados fueron: *gini* y *best*.

Después de seleccionar los parámetros anteriores se evaluó el comportamiento de diversas combinaciones de variables para el aprendizaje y finalmente se consideraron las variables de *EDAD*, *SEXO*, *TIPO\_PACIENTE*, *NEUMONIA*, *UCI*, *INTUBADO* y las variables de comorbilidades.

#### **4.5 Revisión de información sobre campaña de vacunación en el mundo**

Para conocer los efectos que tuvo la campaña de vacunación mundial fue necesario realizar una revisión bibliográfica de los artículos disponibles tanto de los ensayos clínicos que se utilizaron para aprobar de emergencia las vacunas como de los efectos adversos reportados por las mismas.

Igualmente para tener más información respecto a los eventos adversos causados por las vacunas fue necesario visitar diferentes plataformas de vigilancia del mundo. Estas plataformas tienen como objetivo recopilar los reportes acerca de cualquier efecto adverso que se presenta después de una vacuna en el mundo (*VigiAccess*), Estados Unidos (*VAERS*) y el Espacio Económico Europeo (*EudraVigilance*).

De acuerdo con su página oficial, *VigiAccess* es una plataforma mundial creada en 2015 y dirigida por la OMS que se encarga de difundir los eventos adversos causados por medicamentos o vacunas que se registran en la base de datos *VigiBase*. Estos datos sobre eventos adversos comenzaron a recolectarse en 1968 y son reportados por los departamentos de salud de 170 países involucrados.

Ahora bien, *VAERS* (*Vaccine Adverse Event Reporting System*) y *EudraVigilance* (*European Union Drug Regulating Authorities Pharmacovigilance*) también se encargan de recopilar los eventos adversos causados por medicamentos y vacunas, no obstante estos solo recopilan información de Estados Unidos y el Espacio Económico Europeo respectivamente.

Esta información es relevante ya que permite comparar los efectos adversos que presentan las vacunas de emergencia con los que tienen las vacunas que ya han sido autorizadas con anterioridad.

## 5 Resultados y discusiones

La presente sección tiene como objetivo incluir toda la información obtenida del capítulo anterior y después contrastarla con diferentes artículos para poder resolver las preguntas de investigación. Para organizar los resultados adecuadamente las gráficas y tablas obtenidas, así como su respectivo análisis, se encuentran divididas en estadísticas mundiales, nacionales, resultados del modelo e información sobre la vacunación.

### 5.1 Resultados

#### 5.1.1 Estadísticas mundiales

Los primeros resultados obtenidos fueron respecto al porcentaje de letalidad que presentó el COVID-19 en el mundo. De forma mundial se observó un valor de 1.0316%, sin embargo, en el caso de México se obtuvo un resultado mucho mayor (4.64%) que lo posicionó en el séptimo lugar a nivel mundial. Los primeros lugares de letalidad fueron para Yemen, Sudán y Siria, mientras que los resultados más bajos fueron para Nueva Zelanda, Islandia y Corea del Sur. El segundo concepto estudiado fue el de mortalidad y a nivel mundial se observó un valor de 83.19 defunciones por 100,000 habitantes; de la lista de países de interés 22 de ellos presentaron un resultado superior al valor mundial. Para esta medición el resultado de México mejoró; en este caso países como Italia, Reino Unido o Perú tuvieron un resultado mayor, sin embargo, otros países como España, Francia y Alemania presentaron una mortalidad menor a la de México.

En la Tabla 10 se muestran los tres grupos en los que se clasificaron los resultados de los países de interés para la mortalidad, el grupo rojo corresponde a la mitad superior de los países que presentaron una mortalidad mayor que la mundial; en el grupo amarillo se conforma de la mitad inferior que está por encima del valor mundial; de color verde se agrupan los países con un valor inferior a 83.19 defunciones por 100,000 habitantes. En las figuras 12 y 13 se muestra la distribución que presentan los datos de mortalidad y letalidad de los países de interés, en ambos casos la línea roja muestra el valor mundial siendo de 83.19 por 100,000 habitantes y de 1.03% respectivamente.

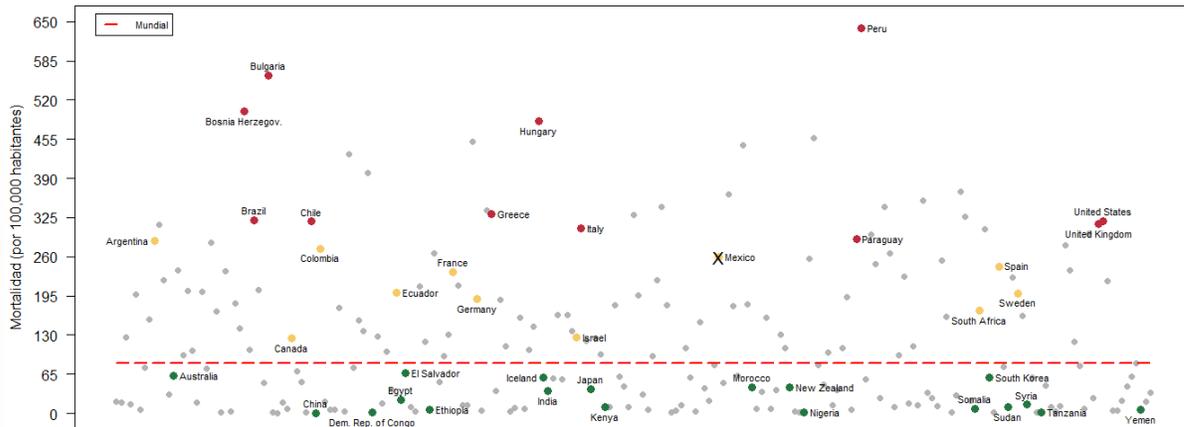
Tabla 10. Letalidad y mortalidad de los países de interés ordenado de acuerdo con la mortalidad

	ISO	Letalidad	Mortalidad
Perú	PER	5.11	638.52
Bulgaria	BGR	2.96	560.89

Bosnia y Herzegovina	BIH	4.04	501.06
Hungría	HUN	2.23	484.45
Grecia	GRC	0.63	330.37
Brasil	BRA	1.96	320.31
Estados Unidos	USA	1.09	319.41
Chile	CHL	1.27	318.42
Reino Unido	GBR	0.88	313.87
Italia	ITA	0.75	306.75
<b>América del sur</b>	<b>OWID_SAM</b>	<b>2.06</b>	<b>305.68</b>
Paraguay	PRY	2.73	289.36
Argentina	ARG	1.34	285.7
Colombia	COL	2.25	273.51
<b>Europa</b>	<b>OWID_EUR</b>	<b>0.82</b>	<b>265.96</b>
México	MEX	4.64	259.2
<b>América del norte</b>	<b>OWID_NAM</b>	<b>1.31</b>	<b>256.12</b>
España	ESP	0.85	243.7
Francia	FRA	0.42	234.5
Ecuador	ECU	3.55	199.66
Suecia	SWE	0.8	199.39
Alemania	DEU	0.43	189.45
Sudáfrica	ZAF	2.53	171.02
Israel	ISR	0.25	125.55
Canadá	CAN	1.08	124.83
<b>Mundial</b>	<b>OWID_WRL</b>	<b>1.03</b>	<b>83.19</b>
El Salvador	SLV	2.1	66.76
Australia	AUS	0.15	61.64
Corea del Sur	KOR	0.11	58.99
Islandia	ISL	0.11	58.73
<b>Oceanía</b>	<b>OWID_OCE</b>	<b>0.16</b>	<b>47.00</b>
Marruecos	MAR	1.28	43.47
Nueva Zelanda	NZL	0.11	42.66
Japón	JPN	0.2	40.06
India	IND	1.19	37.44
<b>Asia</b>	<b>OWID_ASI</b>	<b>0.77</b>	<b>31.83</b>
Egipto	EGY	4.81	22.34
<b>África</b>	<b>OWID_AFR</b>	<b>2.07</b>	<b>18.03</b>
Siria	SYR	5.51	14.3
Sudán	SDN	7.84	10.65
Kenia	KEN	1.66	10.52
Somalia	SOM	4.99	7.73
Yemen	YEM	18.07	6.41
Etiopía	ETH	1.53	6.14
República Democrática del Congo	COD	1.54	1.47

Nigeria	NGA	1.18	1.44
Tanzania	TZA	2.08	1.29
China	CHN	0.35	0.39

Mortalidad por país



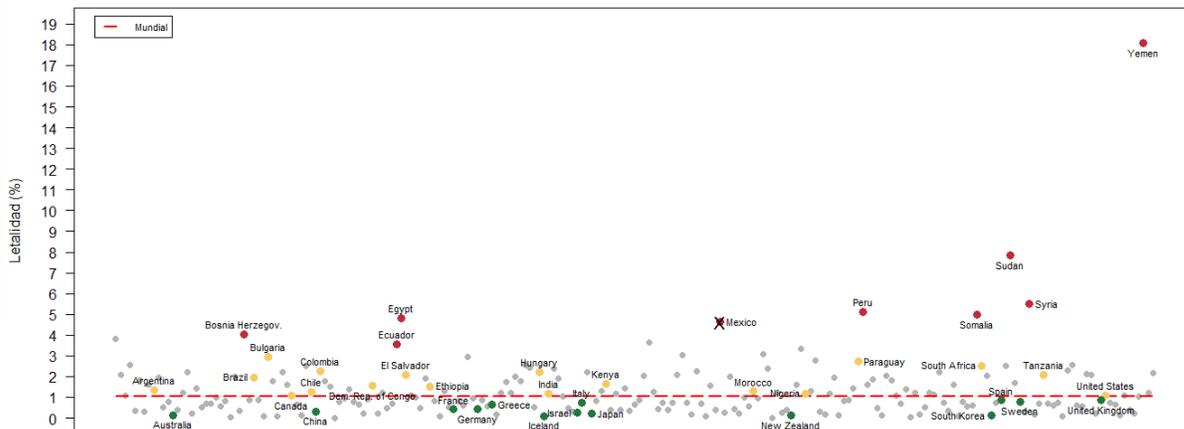
Países

\* Datos al 30 de noviembre del 2022

\*\* Mortalidad mundial: 83.19

Fig. 12 Mortalidad por país de interés

Letalidad por país (%)



Países

\* Datos al 30 de noviembre del 2022

\*\* Letalidad mundial: 1.02%

Fig. 13. Letalidad por país de interés

En las siguientes ilustraciones (Fig. 14 y Fig. 15) se muestran los valores totales de casos y defunciones que tuvo cada país ordenado de acuerdo con su resultado de mortalidad.

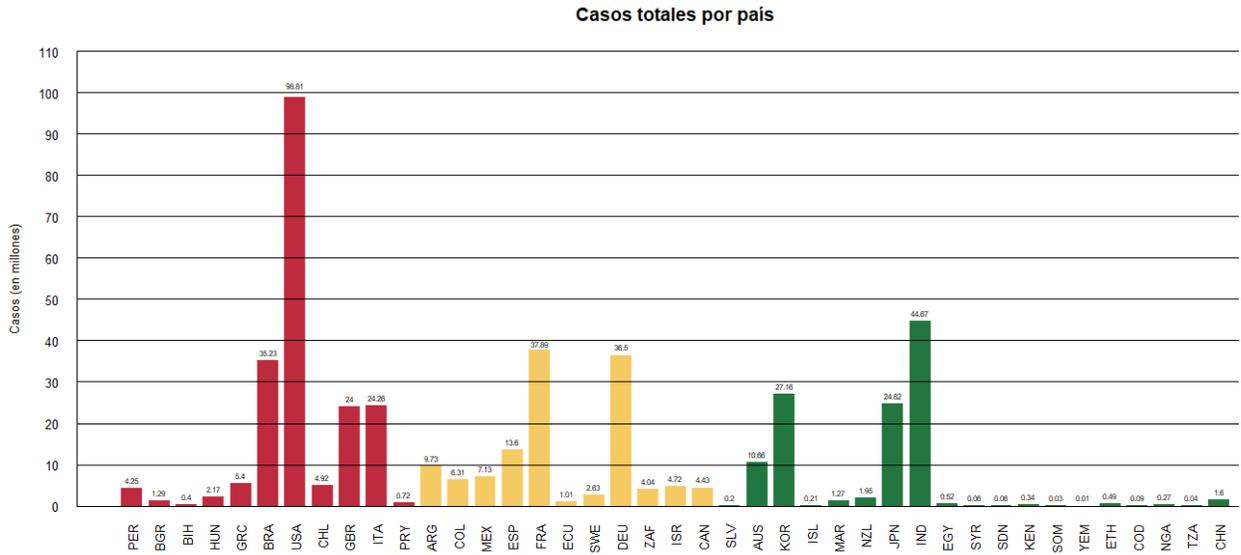


Fig. 14 Casos totales de COVID-19 por país

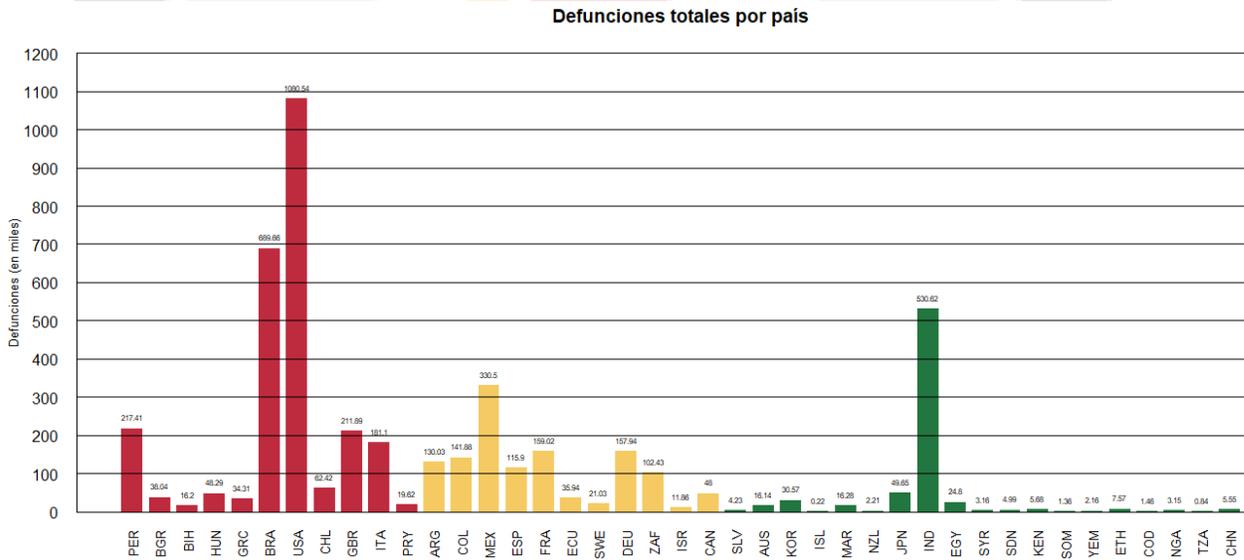


Fig. 15 Defunciones totales por COVID-19 por país

A través de la Tabla 10 y las ilustraciones anteriores se observó que los números altos en mortalidad se concentran en Europa y América mientras que, en los demás continentes se suele presentar una mortalidad inferior a las 80 defunciones por 100,000 habitantes. Este resultado podría indicar que hay alguna característica en común entre los países de estos continentes que pone en riesgo a sus habitantes, por lo que el siguiente paso es analizar las características de las personas que habitan en los países de interés. Para esto, se comparan las medianas de edad, los valores de prevalencia de diabetes y mortalidad para enfermedades cardiovasculares. Los resultados promedio se muestran en la siguiente Tabla:

*Tabla 11. Promedio por grupos de características de la población*

<b>Grupo</b>	<b>Mediana de edad</b>	<b>Prevalencia de diabetes</b>	<b>Mortalidad por enfermedades cardiovasculares</b>
Rojo	38.85	7.15	198.69
Amarillo	35.99	6.93	134.85
Verde	27.64	7.56	261.21

Los valores exactos para cada país se muestran en las Fig. 16, 17 y 18. A través de la Tabla 11, se observa que en promedio los países con un alto nivel de mortalidad por COVID-19 tienen las edades más elevadas además de una alta prevalencia de diabetes y mortalidad por enfermedades cardiovasculares. Por otro lado, a pesar de que los países con una mortalidad baja tienen los valores de mortalidad por enfermedades cardíacas y prevalencia de diabetes más altos de los tres grupos también tienen a la población más joven. Es importante resaltar el caso de Japón que, como se muestra en la Fig. 16, a pesar de tener la mediana de edad más elevada del grupo de interés tiene estadísticas muy bajas para las comorbilidades mencionadas, lo cual puede contribuir para que su mortalidad respecto al COVID-19 sea baja. En el caso específico de México, su población es joven (29.3 años de mediana) y tiene una mortalidad por enfermedades cardiovasculares baja, sin embargo, ocupa el tercer lugar en prevalencia de diabetes, Fig. 17.

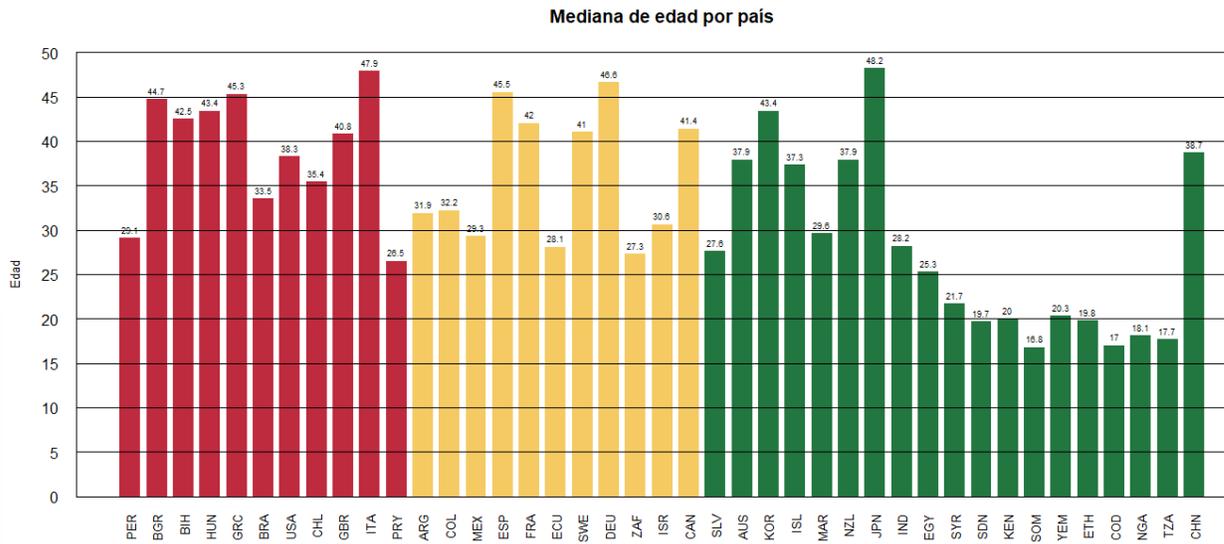


Fig. 16 Comparación de mediana de edad de los países de interés

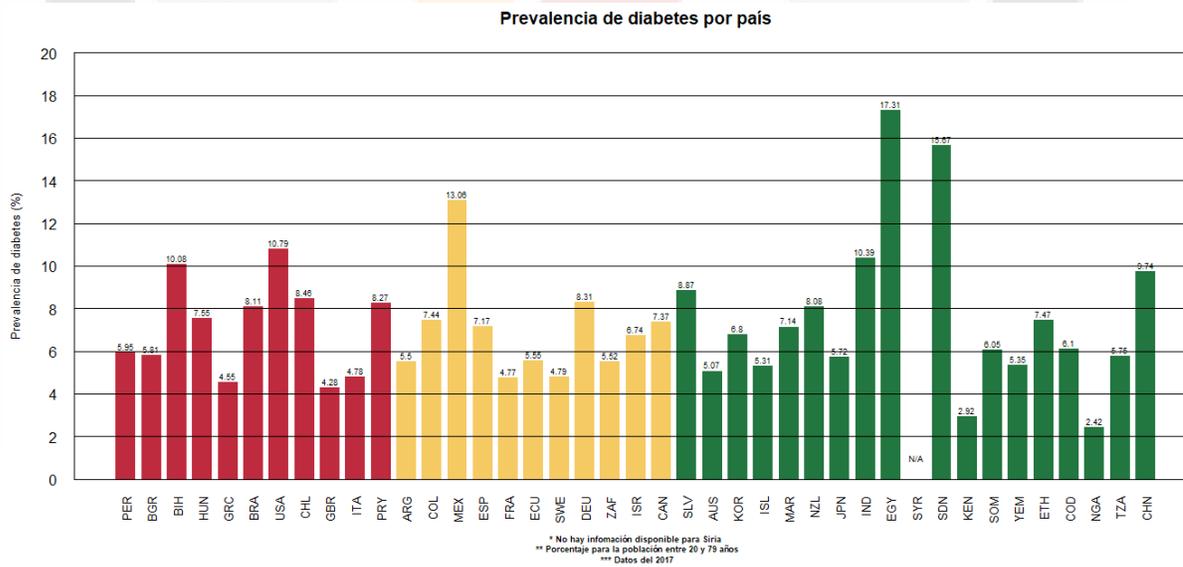


Fig. 17 Comparación de prevalencia de diabetes para los países de interés

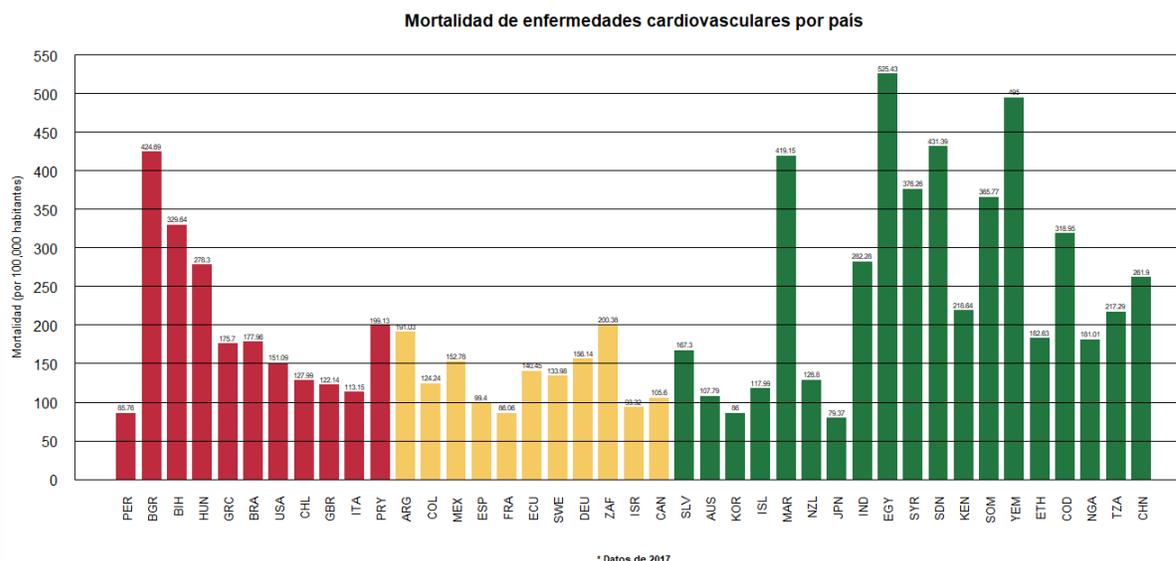


Fig. 18 Comparación de mortalidad por enfermedades cardiovasculares para los países de interés

Las siguientes comparaciones hablan de cómo cada país reaccionó durante la pandemia. Para esto se comparan el número de pruebas realizadas por cada país, el nivel promedio de restricción y el proceso de vacunación que se ha seguido desde que estas fueron aprobadas. En la Tabla 12 se muestran los promedios de cada grupo, mientras que en las figuras 19, 20, 21, 22, 23, 24 y 25 se muestran los datos completos.

Tabla 12. Promedio por grupo de las características específicas de la pandemia

Grupo	Total de pruebas realizadas por 100,000 habitantes	Promedio de pruebas realizadas para detectar un caso positivo	Total de vacunas aplicadas por 100,000 habitantes	Porcentaje de la población que ha recibido una dosis de la vacuna	Porcentaje de la población que ha recibido dos dosis de la vacuna	Total de refuerzos aplicados por 100,000 habitantes	Máximo valor del índice de restricción
Rojo	261.27	36.07	190.49	70.24	65.73	55.41	86.41
Amarillo	169.76	25.41	200.08	77.44	71.8	56.57	86.66
Verde	96.09	33,497.29 <sup>15</sup>	121.66	52.15	47.93	32.82	81.17

<sup>15</sup> El promedio de este grupo sin considerar a China que tuvo un resultado extremadamente alto es de 258.77

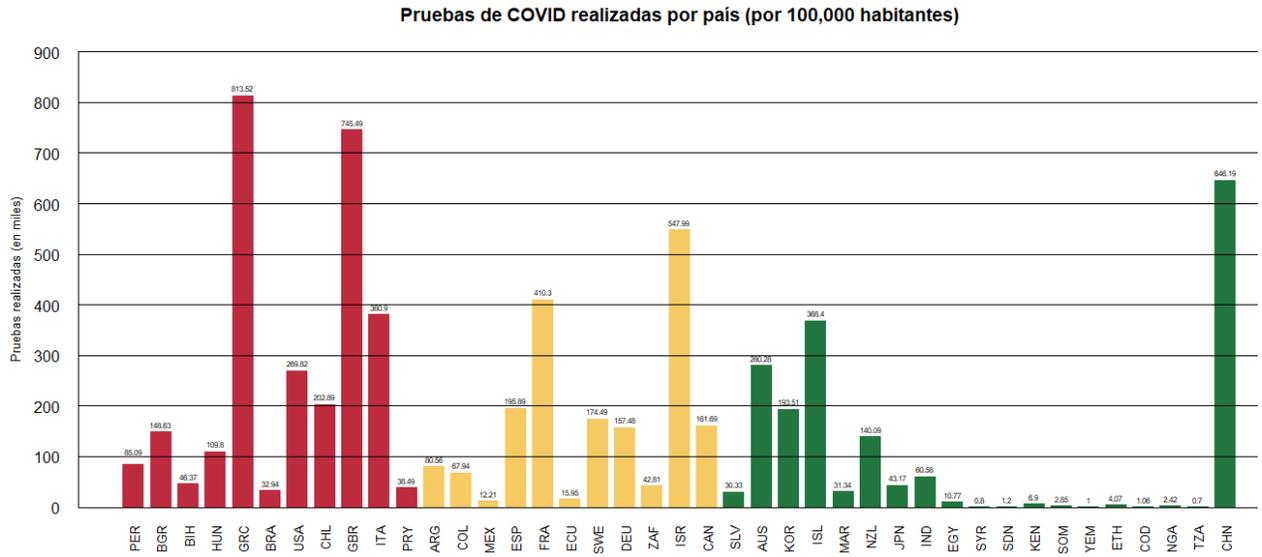


Fig. 19 Pruebas de COVID-19 realizadas en los países de interés

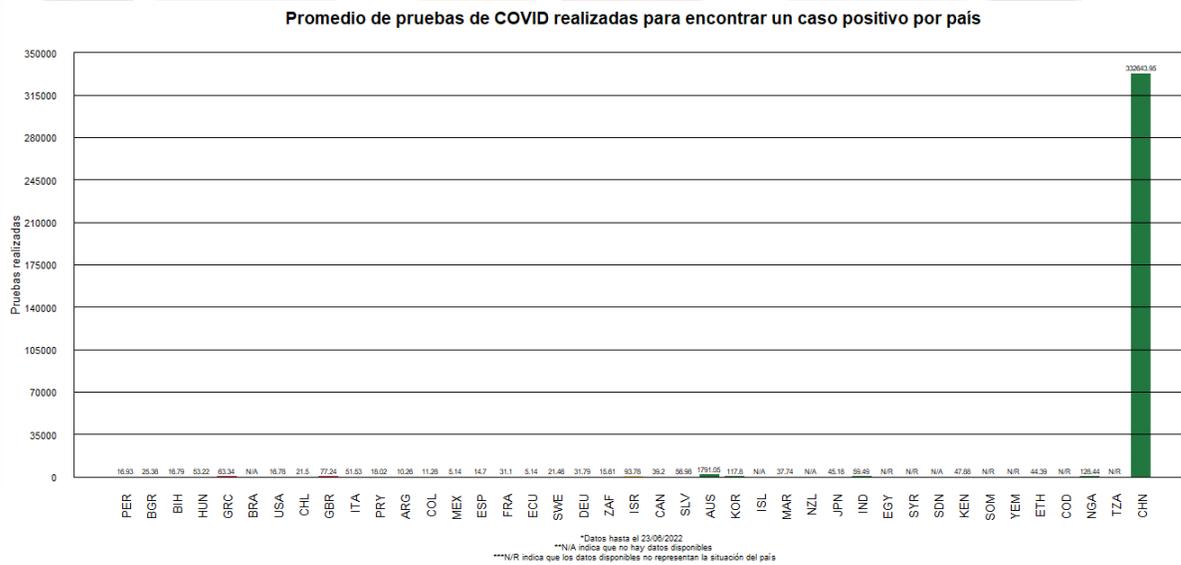


Fig. 20 Promedio de pruebas realizadas para encontrar un caso positivo por país de interés

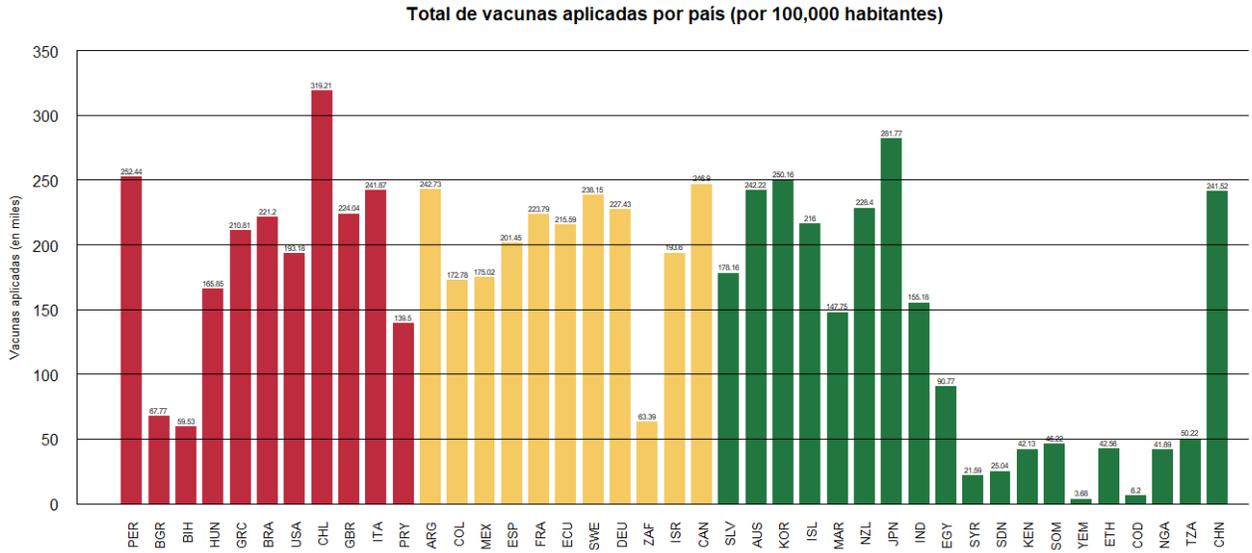


Fig. 21 Total de vacunas aplicadas en los países de interés

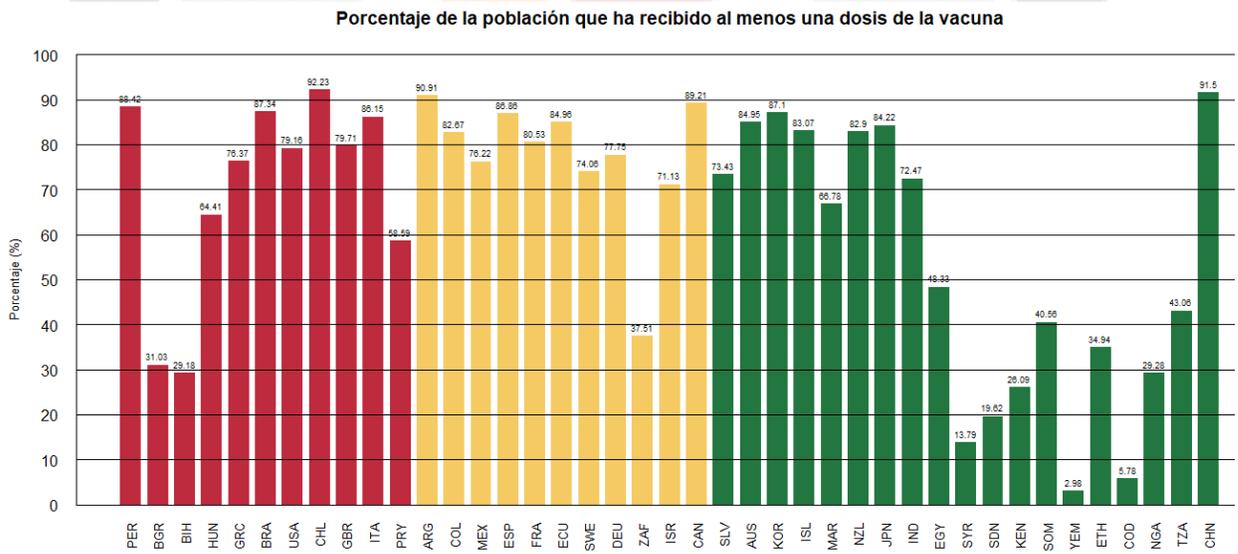
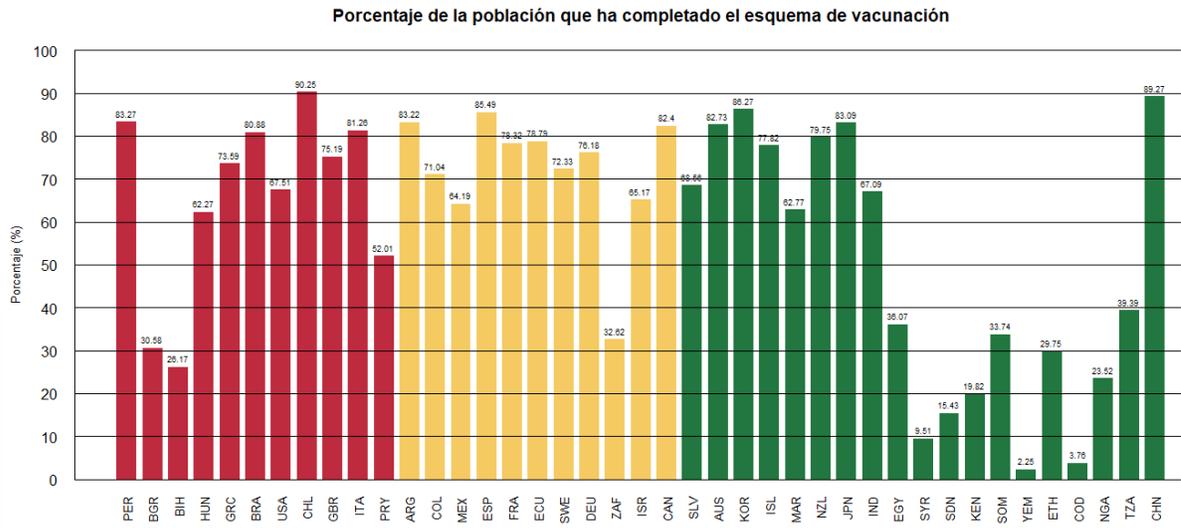


Fig. 22 Porcentaje de la población que ha recibido una vacuna en los países de interés



\* En este caso esquema completo hace referencia a la aplicación de las dos dosis iniciales

Fig. 23 Porcentaje de la población que ha completado su esquema de vacunación en los países de interés

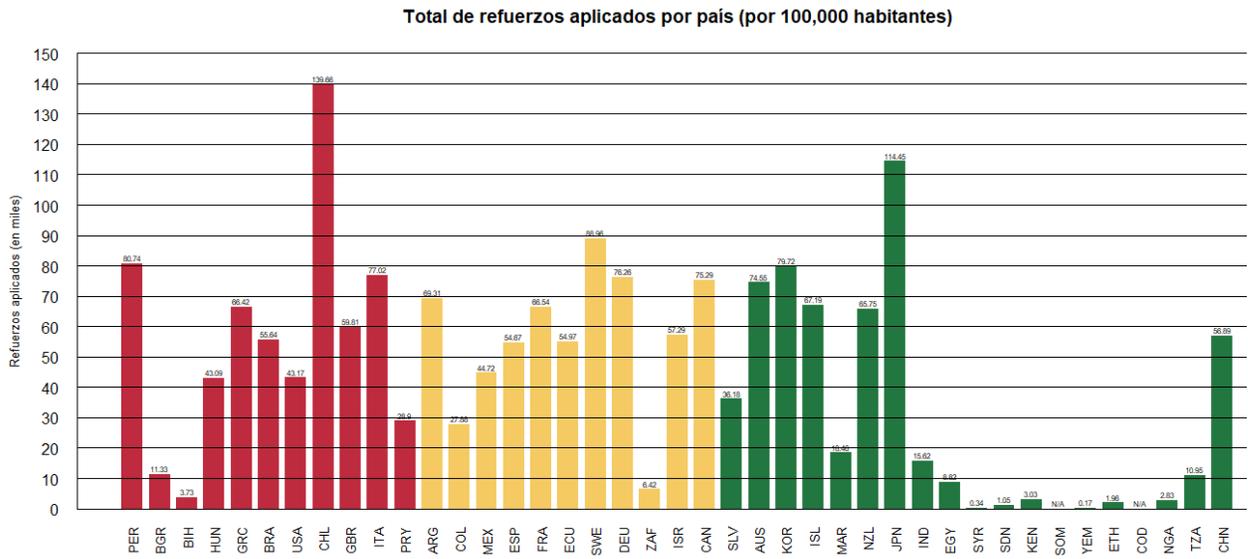
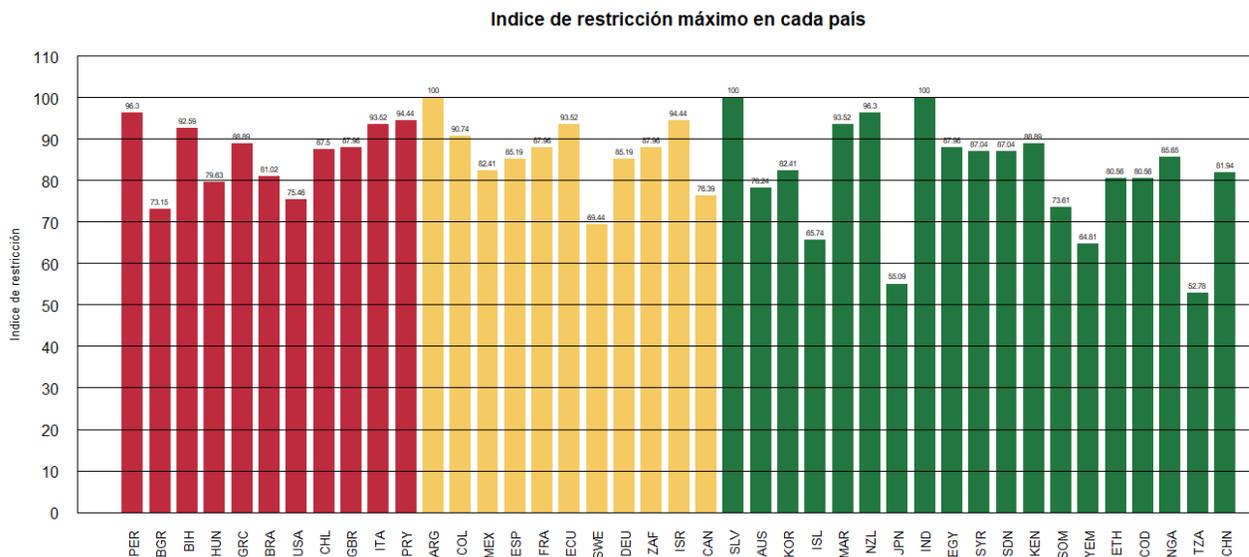


Fig. 24 Total de refuerzos aplicados por cada 100,000 habitantes en los países de interés



*Fig. 25 Índice de restricción máximo observado en cada país de interés*

Con las gráficas anteriores se observó que en general los países europeos fueron los que más pruebas de COVID-19 realizaron, por el contrario, los países africanos y algunos países latinos realizaron muy pocas pruebas; este es el caso de México que en dos años de pandemia apenas realizó poco más de 12,000 pruebas por cada 100,000 habitantes, cifra similar a Ecuador (15,950) y Egipto (10,770), mientras que Grecia, Reino Unido, Israel y China superaron las 500,000 pruebas por cada 100,000 habitantes.

Respecto al número de pruebas necesarias para detectar un caso positivo se observó que países como China, Australia, Nigeria, Corea del Sur e Israel presentaron los valores más elevados y Ecuador, México, Argentina, Colombia y España fueron los más bajos de los países que presentan esta información. Este valor es importante porque puede ayudar a observar el impacto que tuvieron las campañas para realizar pruebas.

Para la vacunación se registró que los países del grupo amarillo fueron los que en promedio aplicaron más vacunas, seguidos muy de cerca del grupo rojo; para este análisis se observa que el grupo verde presenta resultados muy variables entre los países que lo conforman, pues 9 de ellos han vacunado a más del 50% de su población con al menos una dosis, mientras que de los demás países Egipto tiene el porcentaje más alto con 48.33%.

Al analizar la aplicación del esquema completo (dos dosis) se observó un comportamiento similar en los grupos, en el caso del grupo amarillo la única excepción es Sudáfrica que

presenta resultados similares a los de los países africanos del grupo verde. En el caso de la aplicación de refuerzos se reportó que los países tuvieron una postura más relajada ya que solamente Chile y Japón han aplicado más de 100,000 refuerzos por cada 100,000 habitantes.

Finalmente, para el índice de restricción los países con un puntaje más alto (100 puntos) fueron Argentina, El Salvador e India y los que tuvieron el puntaje más bajo fueron Tanzania (52.78), Japón (55.09) y Yemen (64.81). Este valor representa el puntaje máximo al que llegaron los países durante la pandemia, lo cual permite comparar las restricciones de cada país.

Además de comparar las características generales de los países es posible observar las curvas de casos diarios registrados en cada país. Para esto, se elaboraron las siguientes gráficas en donde se comparan los casos diarios en México, Perú, China, Colombia y España.

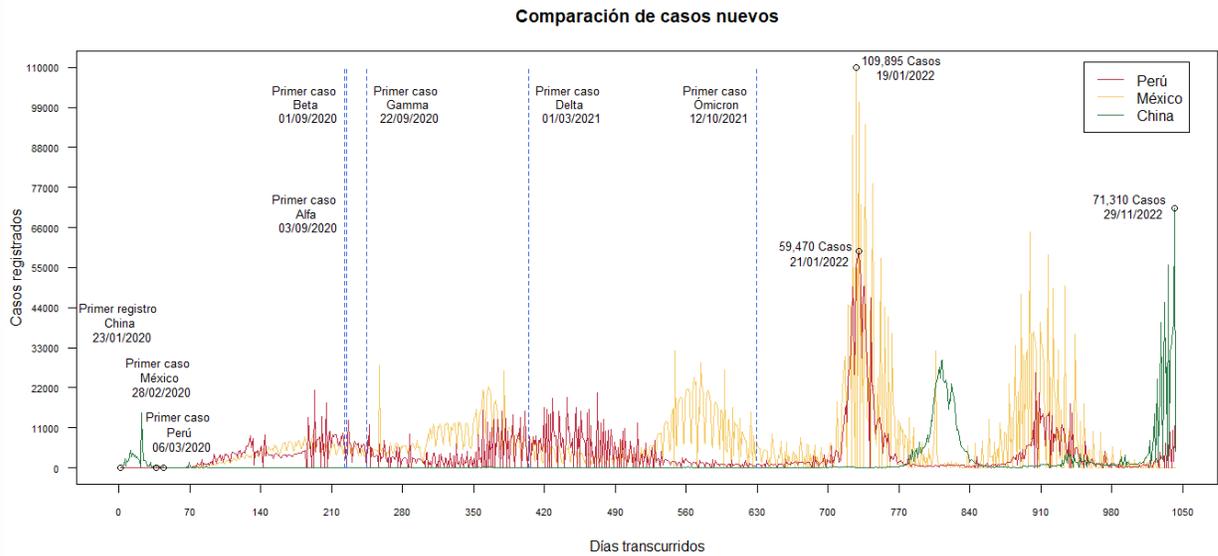


Fig. 26 Casos diarios en Perú, México y China

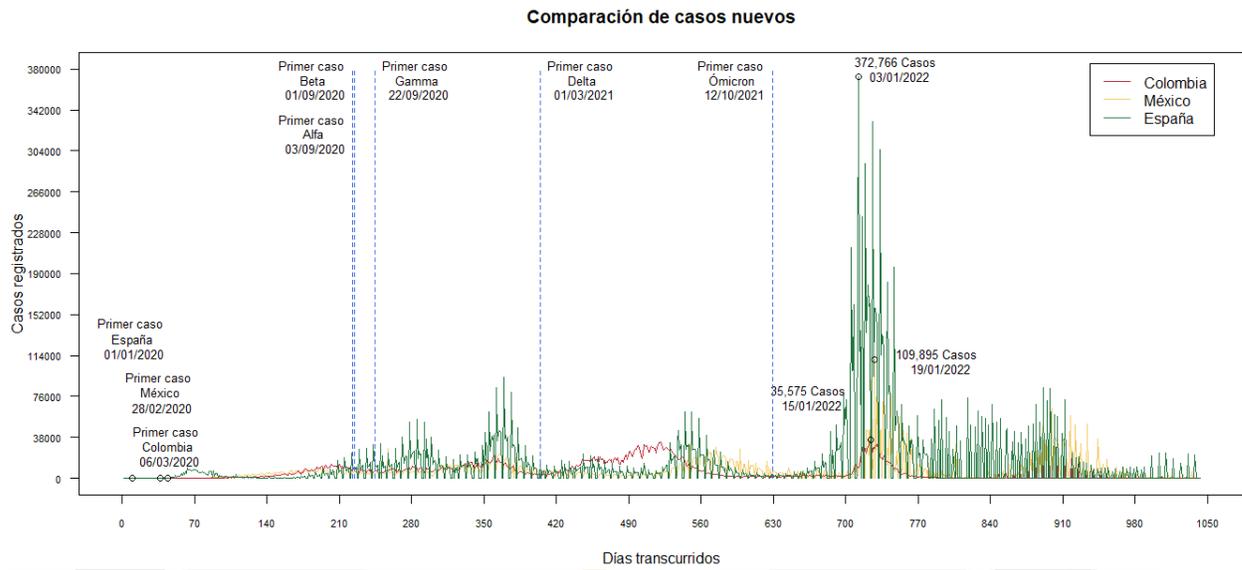


Fig. 27 Casos diarios en Colombia, México y España

Igualmente se realizaron las mismas gráficas para las defunciones diarias observadas en estos países.

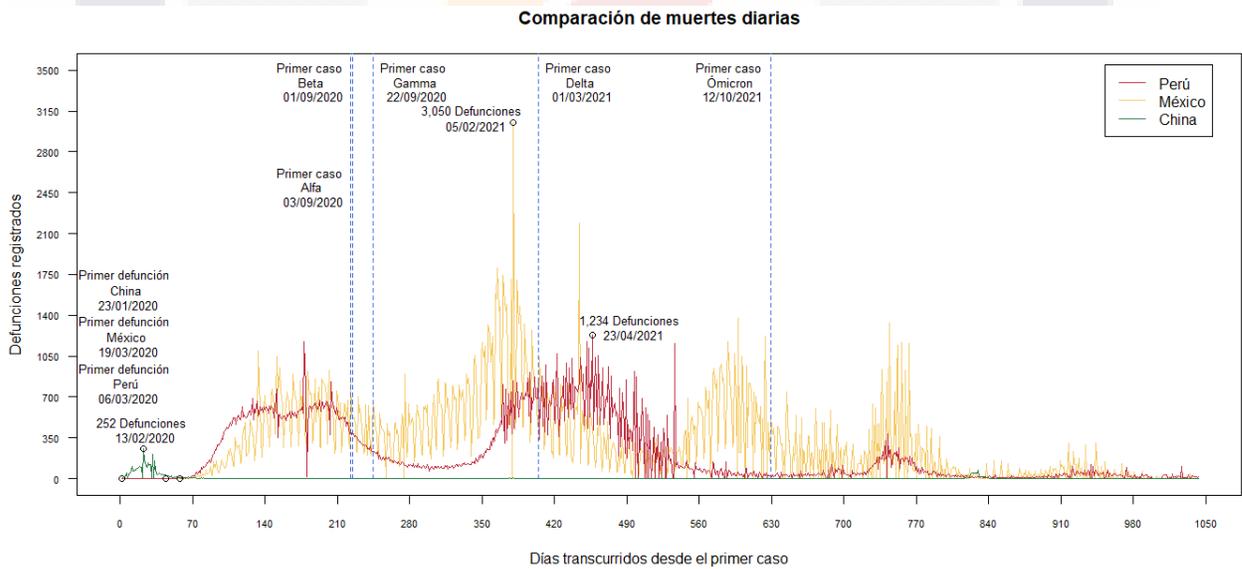


Fig. 28 Defunciones diarias en Perú, México y China

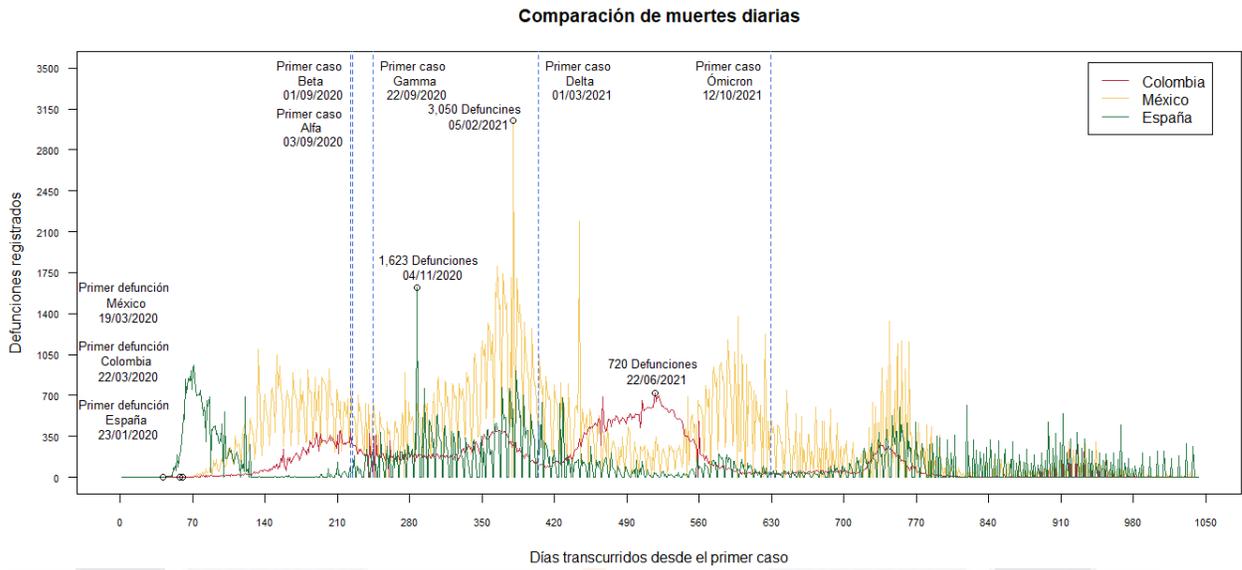


Fig. 29 Defunciones diarias en Colombia, México y España

Otra comparación relevante que se puede realizar es el comportamiento de los casos nuevos y la intensidad de las medidas de restricción en los países. En la Fig. 30 se puede observar esta relación para México.

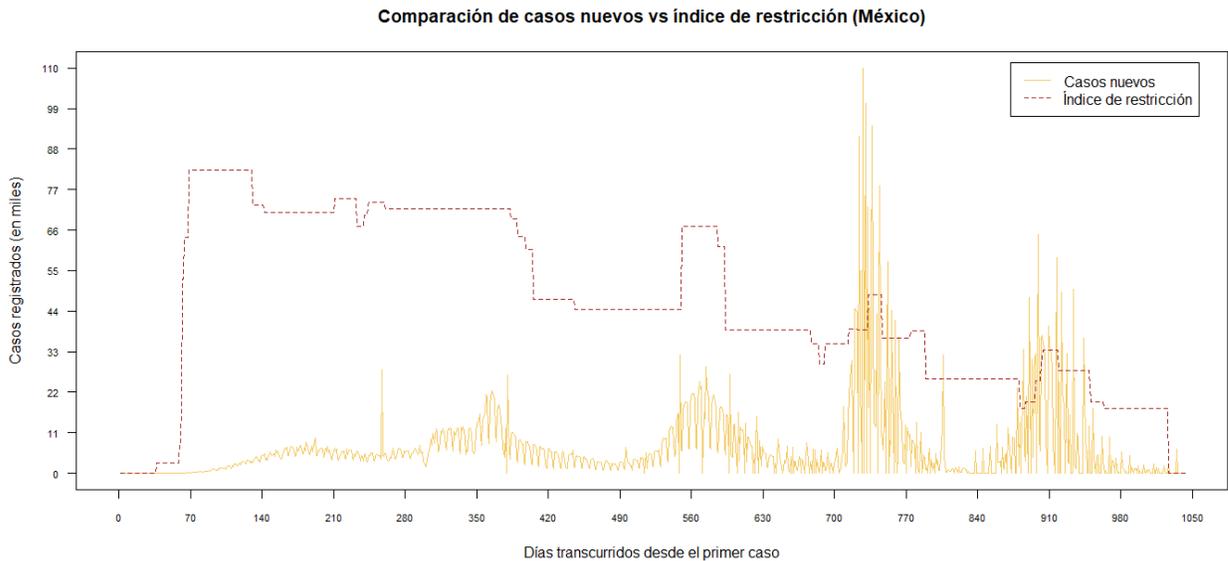


Fig. 30. Comparación entre casos nuevos e índice de restricciones en México

Analizar estas estadísticas es relevante para poder conocer como las condiciones de un país pudieron haber empeorado el resultado de la pandemia. En el caso específico de México se observa que es un país con una población grande y en su mayoría joven (la mitad de la población tiene menos de 30 años), no obstante, hay una fuerte presencia de comorbilidades que se pueden relacionar con casos graves de COVID-19; igualmente se observó que el país siguió un protocolo estricto de vacunación pero fue bastante relajado al momento de realizar pruebas y de mantener las restricciones sugeridas en el mundo. En los siguientes apartados se estudiará el impacto de estas comorbilidades de la población en el desarrollo de la enfermedad.

### 5.1.2 Estadísticas nacionales

Desde el inicio del 2020 al 22 de noviembre de 2022 oficialmente se registraron 7,125,098 casos positivos de COVID-19 y 332,495 defunciones, en el caso de los datos utilizados esta cifra cambia un poco por los filtros considerados en el apartado de *Metodología*. Para analizar las estadísticas nacionales se contabilizaron 7,003,759 casos positivos y 322,428 fallecimientos, como se muestran en las Fig. 31 y 32; además de mantener el registro para los otros niveles de gravedad causados por la enfermedad por COVID-19 como se muestra en la Tabla 13 y 14. Con esta información se obtuvo un porcentaje de letalidad de 4.64%, además, se construyó la Fig. 33, en la cual se observó que fallecieron más hombres que mujeres.

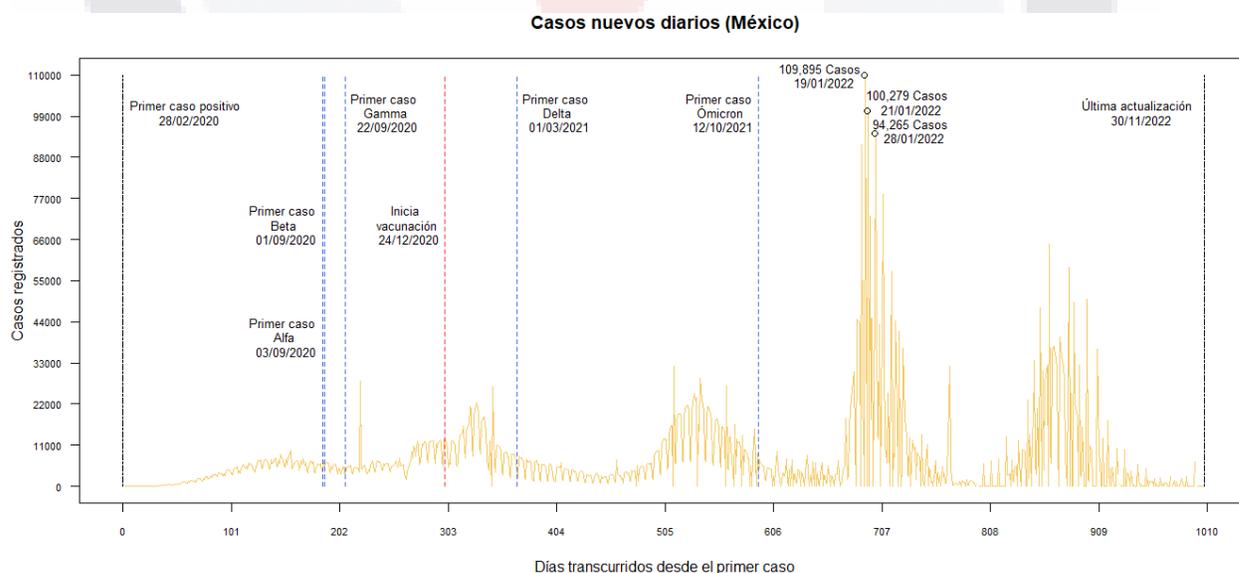


Fig. 31 Casos nuevos en México

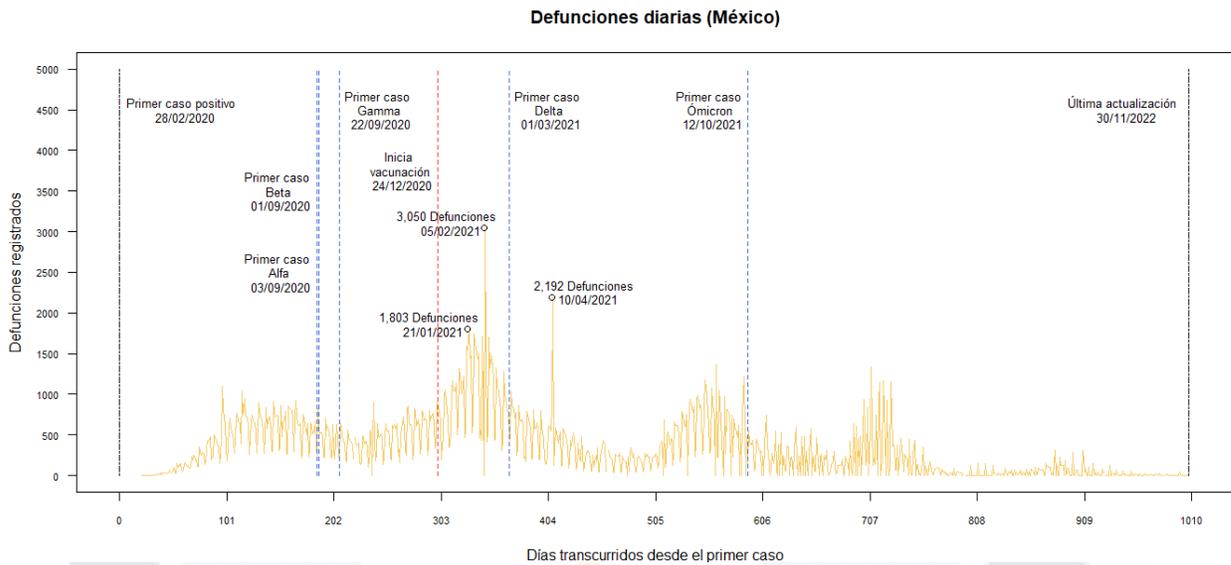


Fig. 32 Defunciones diarias en México

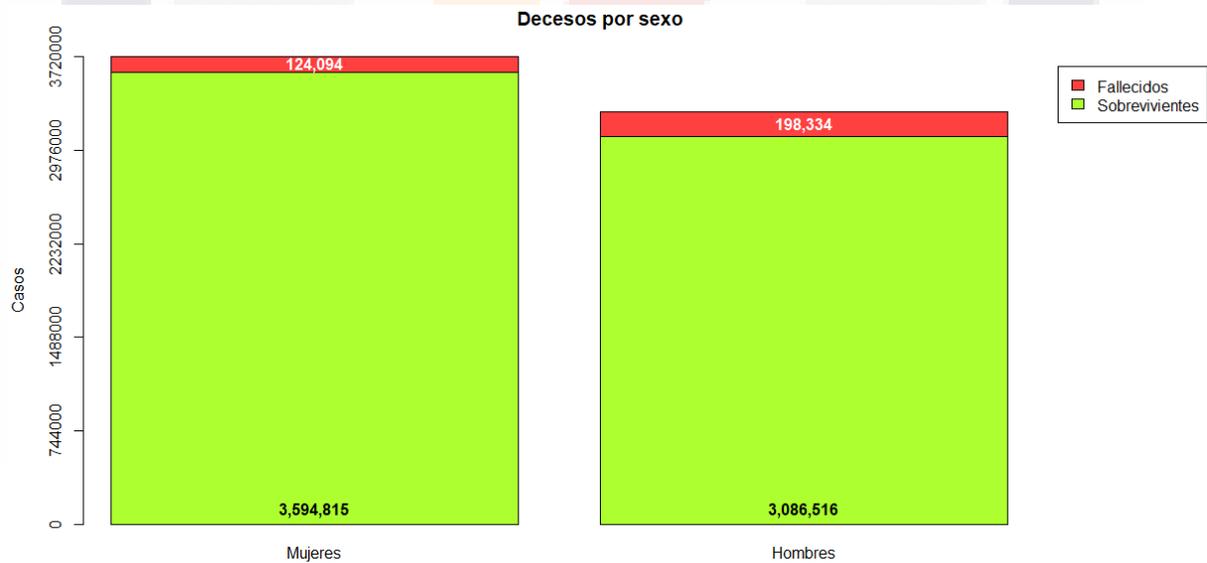


Fig. 33 Defunciones y sobrevivientes por sexo en México

Tabla 13. Casos registrados según el nivel de severidad de la enfermedad de COVID-19

Neumonía	Hospitalizados	Intubados	Ingresados en UCI	Defunciones
506,015	692,986	82,750	52,389	322,428

Tabla 14. Casos por año de las diferentes etapas de la COVID-19

	2020	2021	2022
Hospitalizados	317,736	291,104	84,146
Intubados	48,267	29,647	4,836
Ingreso a UCI	26,742	21,304	4,343
Defunciones	159,850	138,431	24,147

Igualmente, se construyó un gráfico para conocer la distribución de casos por edad y sexo (Fig. 34). En este caso se observó que la mayoría de los casos corresponden a gente joven, principalmente mujeres, y que al aumentar la edad también aumentan las defunciones, siendo más comunes las defunciones en el caso de los hombres.

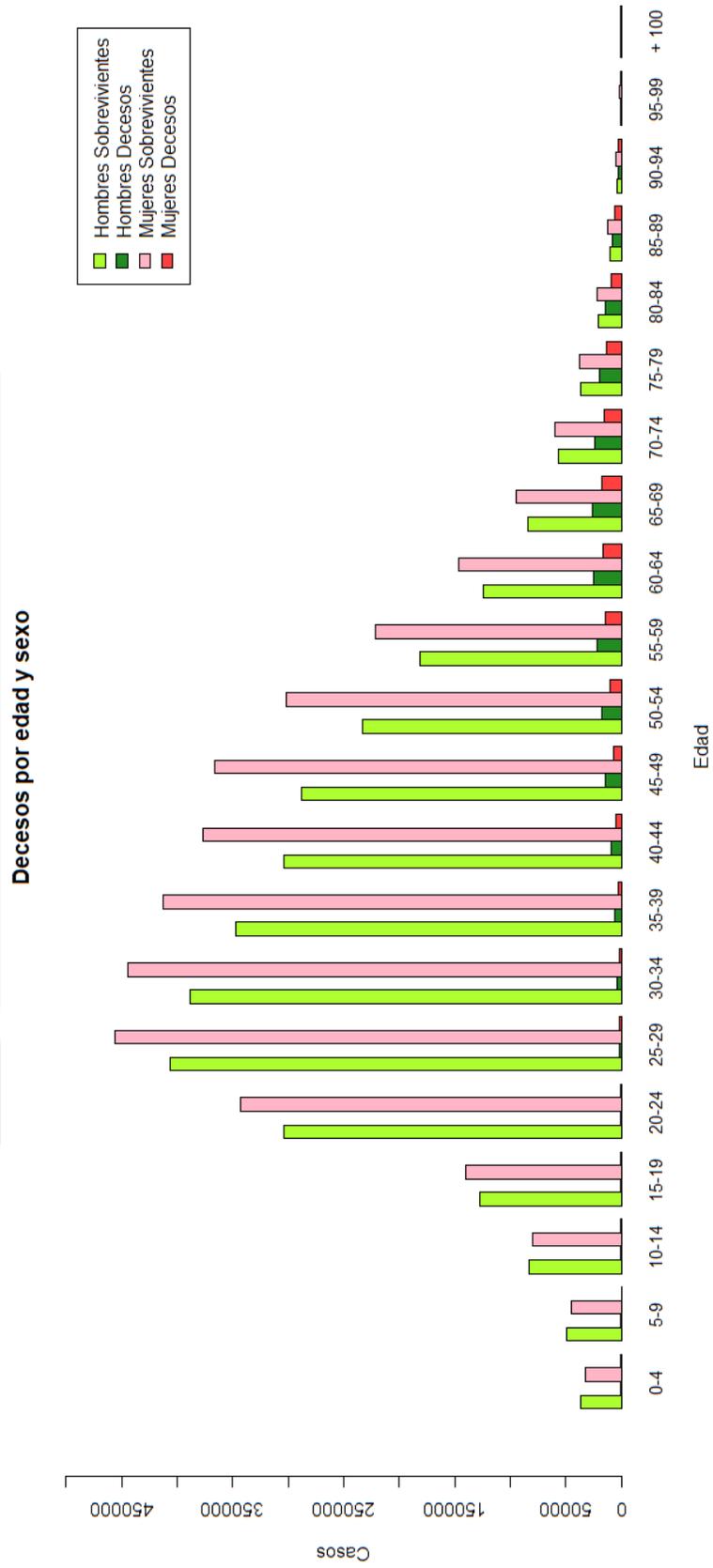


Fig. 34 Número de sobrevivientes y fallecidos por sexo y grupo de edad

Con los gráficos anteriores se observó como la edad y el sexo de las personas influye en el resultado de la enfermedad, por lo que ambas variables pueden ser consideradas factores de riesgo.

Dado que se ha demostrado que las comorbilidades juegan un papel importante para que las personas que se contagian de COVID-19 enfermen de gravedad, se analizó el estado de salud de las personas mediante las siguientes tablas y gráficas: en la Tabla 15 se muestra la comparación entre las personas que fallecieron que padecían al menos una comorbilidad y las que no, esto permitió conocer que el 70.1% de los fallecidos presentaban al menos una comorbilidad.

Tabla 15. Comparación entre fallecimientos con comorbilidades y sin comorbilidades

	Fallecimientos
Con alguna comorbilidad	226,027
Sin comorbilidades	96,401

Complementando la información de la tabla anterior es posible realizar un análisis de riesgo relativo. Para esto se construyó la Tabla 16 considerando la situación: “Fallecer por complicaciones causadas por COVID-19 dado que se presenta alguna comorbilidad”.

Tabla 16. Tabla de contingencia para cálculo de riesgo relativo y riesgo absoluto

	Con comorbilidades	Sin comorbilidades	
Casos que fallecieron	226,027	96,401	322,428
Casos que sobrevivieron	1,826,518	4,854,813	6,681,331
	2,052,545	4,951,214	7,003,759

Retomando la ecuación Ec. IX, se obtuvo que el riesgo relativo de morir por COVID-19 cuando una persona tiene comorbilidades es 2.56, esto indica una relación positiva por lo que se puede afirmar que las comorbilidades en México son un riesgo. Ahora bien, mediante el riesgo absoluto (Ec. X, Ec. XI) se sabe que la probabilidad de fallecer por COVID-19 para el grupo de casos con comorbilidades es del 70.10% mientras que para el otro grupo es del 27.34%; aplicando estos resultados se obtiene que las comorbilidades pueden ser atribuidas exclusivamente del 42.76% de las muertes por complicaciones de COVID-19.

En este contexto es de gran importancia conocer las comorbilidades que tienen un mayor impacto en las defunciones por COVID-19. Para esto, se muestran las estadísticas en la Tabla 17 y además se compararon los porcentajes de defunciones por las diferentes comorbilidades en la Fig. 35 y Fig. 36.

Tabla 17. Casos y defunciones por comorbilidad

Comorbilidad	Casos	Decesos
Embarazo	60,770	365
Diabetes	616,866	118,844
EPOC	47,581	14,198
Asma	130,678	5,463
Inmunosupresión	39,265	7,468
Hipertensión	833,857	143,137
Otra comorbilidad	104,024	16,739
Enfermedad cardiovascular	69,192	16,214
Obesidad	677,211	66,103
Enfermedad renal crónica	66,423	23,313
Tabaquismo	386,886	23,854

La información incluida en la Tabla 17 resulta de gran utilidad para poder obtener los valores de riesgo relativo, absoluto y atribuible para cada una de las comorbilidades. Para esto, es necesario plantear un caso de estudio y sustituir los valores necesarios en la tabla de contingencia (Tabla 1) y aplicar las ecuaciones necesarias (Ec. IX, X, XI y XII). Por ejemplo, se observa que la comorbilidad con más decesos es la hipertensión por lo cual para analizar su riesgo relativo se define el caso como: *el riesgo de fallecer debido a la enfermedad de COVID-19 dado que la persona tiene hipertensión.*

	Persona hipertensa	Persona no hipertensa	
Defunciones	143,137	179,291	322,428
Sobrevivientes	690,720	5,990,611	6,681,331
	833,857	6,169,902	7,003,759

$$RR = \frac{143,137/322,428}{690,720/6,681,331} = \frac{0.4439}{0.1033} = 4.2942 \quad \text{Ec. XXXI}$$

$$R_{Abs_H} = \frac{143,137}{322,428} (100) = 44.39 \% \quad \text{Ec. XXXII}$$

$$R_{Abs_{NH}} = \frac{690,720}{6,681,331} (100) = 10.34 \% \quad \text{Ec. XXXIII}$$

$$RA = 44.39 \% - 10.34 \% = 34.06\% \quad \text{Ec. XXXIV}$$

Estos resultados son muy importantes ya que permiten obtener conclusiones respecto al riesgo que representan las comorbilidades al contraer la COVID-19. Para el valor de riesgo relativo se obtuvo un resultado de 4.29, esto indica que la hipertensión representa un factor de riesgo importante para fallecer de COVID-19, por otro lado el riesgo absoluto permite comparar las probabilidades de fallecer de las personas con y sin hipertensión siendo de 44.39% y 10.33% respectivamente. Finalmente, estos números permiten afirmar que un 34.06% de las muertes pueden ser atribuidas a padecer hipertensión. En la Tabla 18 se incluyen estas mediciones para todas las comorbilidades analizadas.

Tabla 18. Valores de riesgo para cada comorbilidad

Comorbilidad	Riesgo relativo	Riesgo absoluto		Riesgo atribuible
		Expuestos	No expuestos	
<b>Embarazo</b>	0.13	0.11%	0.90%	N/A <sup>16</sup>
<b>Diabetes</b>	4.94	36.86%	7.45%	29.41%
<b>EPOC</b>	8.81	4.40%	0.50%	3.90%
<b>Asma</b>	0.90	1.69%	1.87%	N/A <sup>17</sup>
<b>Inmunosupresión</b>	4.87	2.32%	0.48%	1.84%
<b>Hipertensión</b>	4.29	44.39%	10.34%	34.06%
<b>Otra comorbilidad</b>	3.97	5.19%	1.31%	3.89%
<b>Enfermedad cardiovascular</b>	6.34	5.03%	0.79%	4.24%

<sup>16</sup> En este caso el valor no tiene sentido ya que es negativo, el valor obtenido fue - 0.79%

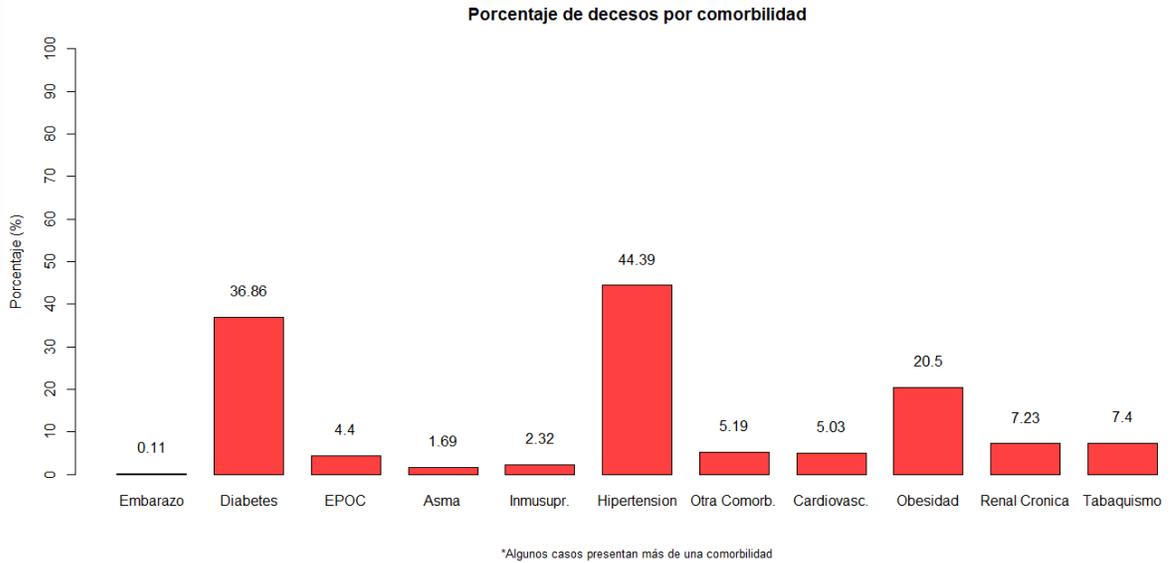
<sup>17</sup> En este caso el valor no tiene sentido ya que es negativo, el valor obtenido fue - 0.18%

<b>Obesidad</b>	2.24	20.50%	9.15%	11.36%
<b>Enfermedad renal crónica</b>	11.21	7.23%	0.65%	6.59%
<b>Tabaquismo</b>	1.36	7.40%	5.43%	1.96%

De la información anterior se puede observar que las comorbilidades que conllevan más riesgo son la enfermedad renal crónica, el EPOC y las enfermedades cardiovasculares, mientras que el tabaquismo, la obesidad y otras comorbilidades son las que tienen un menor riesgo. Es importante resaltar que con estos resultados se observa que el asma y el embarazo tienen un resultado menor a uno, lo cual indica que estas comorbilidades podrían actuar como factores de protección para evitar que las personas fallezcan por COVID-19.

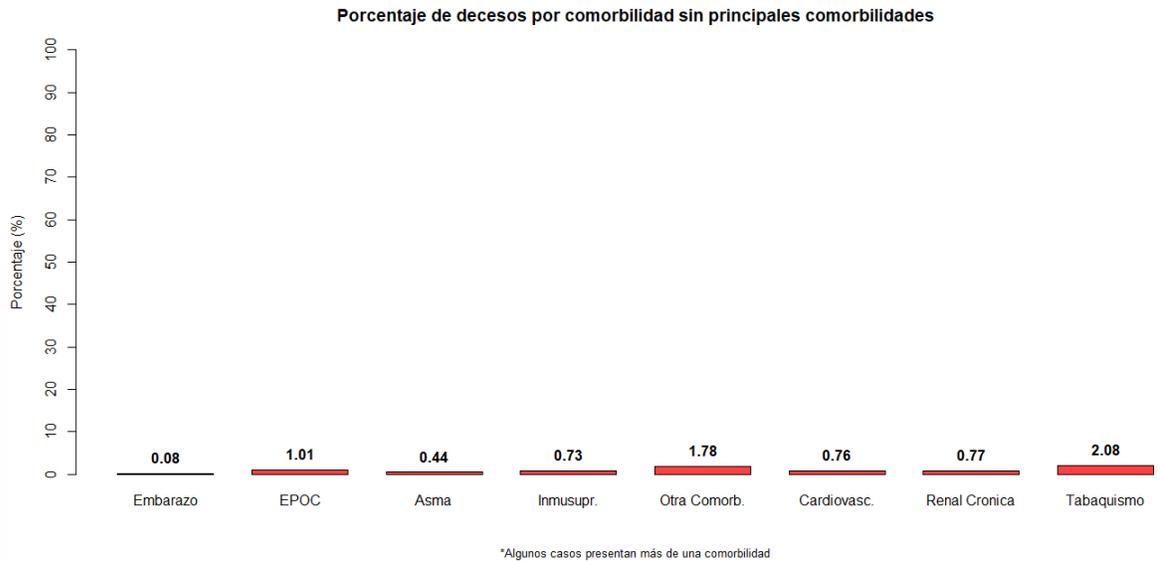
A pesar de no tener un riesgo tan elevado la hipertensión, diabetes y la obesidad son las comorbilidades a las cuales se les puede atribuir un mayor porcentaje de defunciones, por el contrario, el tabaquismo y la inmunosupresión son a las que causan menos muertes.

En la Fig. 35 se muestran los porcentajes de defunción por cada comorbilidad respecto al total de defunciones registradas. Es necesario aclarar que los porcentajes suman más del 100% debido a que existen casos que registraron más de una comorbilidad. A partir de esta ilustración se observa que la diabetes, hipertensión y obesidad son las comorbilidades más comunes entre los casos que fallecieron mientras que el embarazo y el asma son las menos frecuentes. Después en la Fig. 36 se muestran los porcentajes de defunción por



*Fig. 35 Comparación de porcentaje de fallecidos por comorbilidad*

comorbilidad después de eliminar los registros correspondientes a las tres enfermedades con los porcentajes más altos.

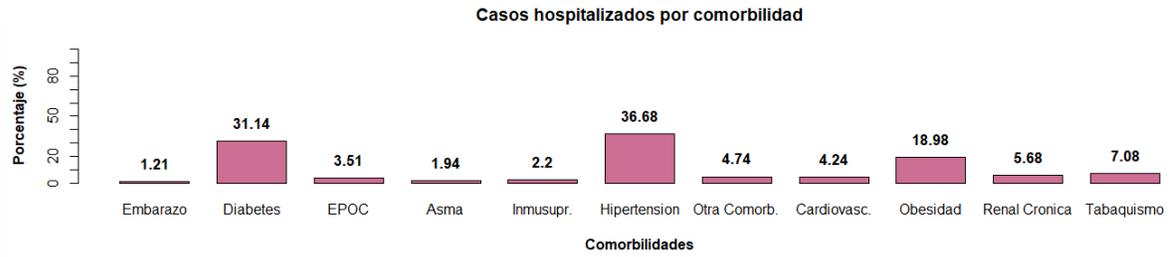
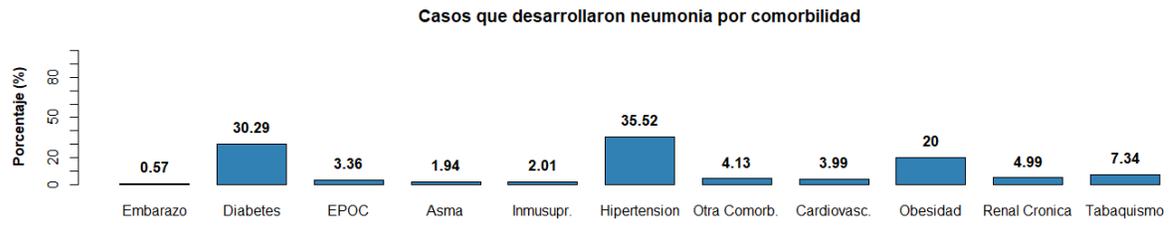


*Fig. 36 Porcentaje de defunciones por comorbilidad sin considerar hipertensión, diabetes y obesidad*

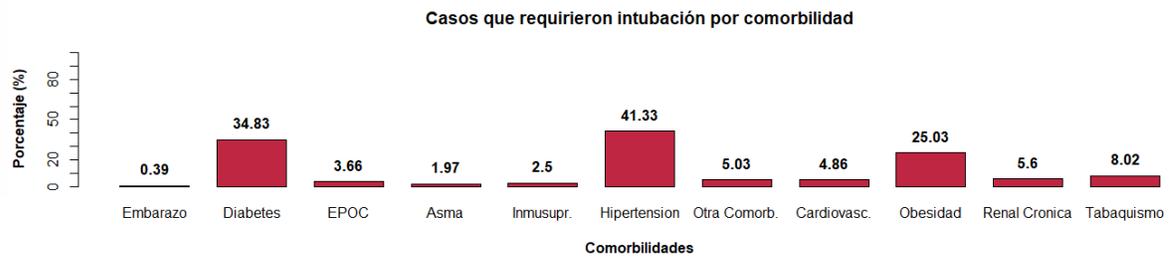
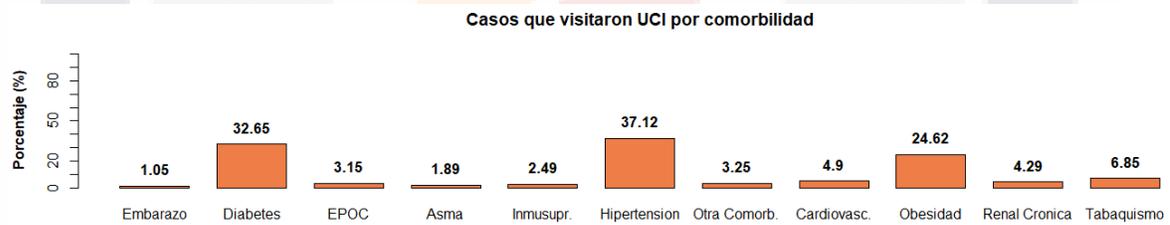
A través de las imágenes anteriores se observa el impacto que tiene la presencia de la hipertensión, diabetes y obesidad ya que al eliminar estas comorbilidades disminuye en gran medida el porcentaje de las demás. Se observó que la enfermedad renal crónica, el tabaquismo y la enfermedad cardiovascular fueron los que presentaron mayor disminución del porcentaje de defunción con 6.46%, 5.32% y 4.27% respectivamente.

Ahora bien, la COVID-19 no es una enfermedad que causa directamente la muerte sino que va evolucionando según el tratamiento que la persona reciba. Las etapas que se puede considerar que presenta esta enfermedad son las siguientes: desarrollo de neumonía, hospitalización de las personas, traslado a cuidados intensivos, intubación y defunción. Dado que la mejor forma de evitar que las personas fallezcan es ofrecerles un tratamiento adecuado en alguna de las etapas anteriores resulta relevante también conocer las características de las personas que se encuentran en las fases previas a la defunción.

La figura 35 resulta muy relevante para comparar el impacto de las comorbilidades observado en las defunciones con el de otras etapas de gravedad de la enfermedad de COVID-19. En las figuras 37 y 38 se muestran los porcentajes de casos por cada comorbilidad, en todos se observa que diabetes, hipertensión y obesidad tienen el mayor porcentaje, además de mantener un porcentaje similar al observado para los decesos.



*Fig. 37 Casos de neumonía y que fueron hospitalizados por comorbilidad*



*Fig. 38 Casos que visitaron UCI y requirieron intubación por comorbilidad*

Con toda la información registrada se observa que, como se ha indicado antes, la edad y las comorbilidades tuvieron un efecto negativo en la severidad con la que se desarrolló la enfermedad por COVID-19 en México.

### 5.1.3 Modelo

Como se mencionó en el capítulo de Metodología, se evaluaron tres posibles algoritmos para desarrollar un modelo que permitiera conocer los principales factores de riesgo para que una persona fallezca de COVID-19. Los algoritmos evaluados fueron máquinas de soporte vectorial, árboles de decisión y bosques aleatorios. Para poder decidir el algoritmo adecuado para el desarrollo del modelo se evaluó el tiempo de ejecución y el desempeño de acuerdo con las métricas de aprendizaje de cada algoritmo, sin considerar métodos para eliminar el sobreajuste o mejorar el desempeño.

Para llevar a cabo esta evaluación se seleccionó un conjunto de datos específico de la matriz reducida y se utilizó el mismo para los tres casos. Para la construcción de la matriz de datos de prueba se utilizaron únicamente las variables *EDAD* y *SEXO*, se consideró como variable objetivo *FECHA\_DEF* y se utilizaron diferentes cantidades de datos. Las cantidades de datos utilizadas fueron: 1%, 2%, 10%, 25%, 50% y 100%.

El tiempo que tomó cada algoritmo en concluir la tarea asignada se obtuvo mediante la función *time()* de *Python* y los resultados de la prueba de tiempo de ejecución se muestran en la Tabla 19. En esta tabla es posible observar que el algoritmo de árboles de decisión es el que requiere menor tiempo de ejecución y que el algoritmo de máquinas de soporte vectorial requirió mucho más tiempo y no pudo concluir para el 100% de los datos.

Tabla 19. Tiempo de ejecución según la cantidad de datos usados

Cantidad de datos	Árboles de decisión	Bosques aleatorios	Máquinas de soporte vectorial
100%	3.84 segundos	102.23 segundos	* 12 horas <sup>18</sup>
50%	1.02 segundos	24.75 segundos	22,753.54 segundos (6 hrs. 19 min. 13 seg.)
25%	0.75 segundos	14.03 segundos	3492.49 segundos (58 min. 12 seg.)
10%	0.46 segundos	3.85 segundos	643.11 segundos (10 min. 43 seg.)
2%	0.48 segundos	1.46 segundos	24.28 segundos
1%	0.80 segundos	1.64 segundos	7.57 segundos
Promedio	1.15 segundos	24.66 segundos	5,384.2 segundos

<sup>18</sup> Se detuvo la ejecución del algoritmo después de 12 horas

La segunda prueba consistió en evaluar las métricas de aprendizaje. En la Tabla 20, se muestra el resultado que tuvo cada algoritmo para las cuatro métricas principales. Es fácil ver que los tres algoritmos presentaron un desempeño muy similar, por lo que esta prueba no fue concluyente para seleccionar uno de los tres algoritmos.

Tabla 20. Valor de las métricas según la cantidad de datos usados

Cantidad de datos	Árboles de decisión		Bosques aleatorios		Máquinas de soporte vectorial	
100%	Exactitud:	79.67%	Exactitud:	79.67%	N/A	
	Precisión:	81.22%	Precisión:	81.22%		
	Exhaustividad:	77.19%	Exhaustividad:	77.19%		
	Valor F:	79.15%	Valor F:	79.15%		
50%	Exactitud:	79.78%	Exactitud:	79.78%	Exactitud:	79.84%
	Precisión:	81.37%	Precisión:	81.36%	Precisión:	81.08%
	Exhaustividad:	77.10%	Exhaustividad:	77.10%	Exhaustividad:	77.96%
	Valor F:	79.17%	Valor F:	79.17%	Valor F:	79.49%
25%	Exactitud:	80.09%	Exactitud:	80.14%	Exactitud:	79.83%
	Precisión:	81.53%	Precisión:	82.19%	Precisión:	81.13%
	Exhaustividad:	78.27%	Exhaustividad:	77.41%	Exhaustividad:	77.80%
	Valor F:	79.86%	Valor F:	79.73%	Valor F:	79.43%
10%	Exactitud:	79.72%	Exactitud:	79.72%	Exactitud:	79.52%
	Precisión:	80.73%	Precisión:	80.73%	Precisión:	80.66%
	Exhaustividad:	78.76%	Exhaustividad:	78.76%	Exhaustividad:	78.12%
	Valor F:	79.73%	Valor F:	79.73%	Valor F:	79.37%
2%	Exactitud:	80.58%	Exactitud:	80.65%	Exactitud:	78.78%
	Precisión:	80.76%	Precisión:	80.74%	Precisión:	78.91%
	Exhaustividad:	81.92%	Exhaustividad:	82.10%	Exhaustividad:	78.81%
	Valor F:	81.34%	Valor F:	81.42%	Valor F:	78.86%
1%	Exactitud:	78.66%	Exactitud:	78.29%	Exactitud:	78.60%
	Precisión:	80.42%	Precisión:	78.59%	Precisión:	79.42%
	Exhaustividad:	75.62%	Exhaustividad:	77.61%	Exhaustividad:	76.11%
	Valor F:	77.95%	Valor F:	78.10%	Valor F:	77.73%
Promedio	Exactitud:	79.75%	Exactitud:	79.71%	Exactitud:	79.31%
	Precisión:	81.01%	Precisión:	80.81%	Precisión:	80.24%
	Exhaustividad:	78.14%	Exhaustividad:	78.36%	Exhaustividad:	77.76%
	Valor F:	79.53%	Valor F:	79.55%	Valor F:	78.98%

Después de realizar las pruebas anteriores se tomó la decisión de utilizar un modelo basado en el algoritmo de árbol de decisión, esto debido a que es el algoritmo que presenta el menor tiempo de ejecución, además de entregar un apoyo visual y presentar un buen desempeño para las métricas de aprendizaje. El siguiente paso en la construcción del modelo fue evaluar las variables que aportan una mayor información y que ayudan a

mejorar las métricas. Para esto, se evaluó el desempeño del modelo al agregar o quitar variables considerando el 100% de los datos de la matriz reducida y se consideró un conjunto de entrenamiento del 75% de los datos y un conjunto de prueba del 25% sin podar el árbol. En la Tabla 21 se muestran los resultados de esta evaluación.

Tabla 21. Resultado de las métricas según las variables utilizadas

Variables	Exactitud	Precisión	Exhaustividad	Valor F	Hojas
<b>Sexo y edad</b>	79.75%	80.31%	78.83%	79.56%	211
<b>Sexo, edad y comorbilidades</b>	81.90%	83.49%	79.52%	81.46%	11,242
<b>Sexo, edad, y comorbilidades (excepto tabaquismo y asma)</b>	82.04%	83.47%	79.90%	81.65%	7,743
<b>Sexo, edad, comorbilidades y neumonía</b>	88.88%	83.47%	79.90%	81.65%	14,053
<b>Sexo, edad, comorbilidades y neumonía (excepto tabaquismo y asma)</b>	88.96%	83.47%	79.90%	81.65%	10,277
<b>Sexo, edad, comorbilidades y hospitalizados</b>	94.79	84.29	76.91	80.43	11,990
<b>Sexo, edad, comorbilidades y hospitalizados (excepto tabaquismo y asma)</b>	94.86	83.47	79.90	81.65	8,960
<b>Sexo, edad, comorbilidades e intubados</b>	84.04	84.29	76.91	80.43	12,196
<b>Sexo, edad, comorbilidades e intubados (excepto tabaquismo y asma)</b>	84.14	83.47	79.90	81.65	8,732
<b>Sexo, edad, comorbilidades y UCI</b>	82.77%	84.29%	76.91%	80.43%	12,585
<b>Sexo, edad, comorbilidades y UCI (excepto tabaquismo y asma)</b>	81.87%	83.47%	79.90%	81.65%	8,972
<b>Todas las variables</b>	95.51%	97.17%	93.74%	95.43%	17,166

<b>Todas las variables (excepto tabaquismo y asma)</b>	95.63%	97.34%	93.81%	95.54%	13,813
--	--------	--------	--------	--------	--------

Con los resultados anteriores, es posible observar que las variables *TABAQUISMO* y *ASMA* entorpecen ligeramente al modelo por lo que más adelante es conveniente omitirlas. De igual forma, se observó que la inclusión de las comorbilidades registradas no mejora el modelo y que las variables que describen el estado del paciente después de contraer la enfermedad son las que resultan más relevantes para obtener una predicción más acertada.

Luego de definir las variables fue necesario ajustar los parámetros del modelo. Los primeros parámetros definidos fueron los valores mínimos de muestras para que se realizara una partición y para que se pueda crear una hoja, en este caso se definió que cada hoja represente al 1% de los datos y por consiguiente los nodos internos deben contener mínimo al 2% de los datos disponibles.

Después de definir los parámetros anteriores fue necesario establecer una profundidad máxima para el árbol. Para esto se evaluó el valor de exactitud del modelo y se observaron los resultados utilizando valores de 5 a 100 y después construyendo la gráfica que se muestra en la Fig. 39. Como se puede observar en la imagen el valor más alto de exactitud se obtuvo al utilizar 80 nodos, por lo que este parámetro se configuró a ese valor.

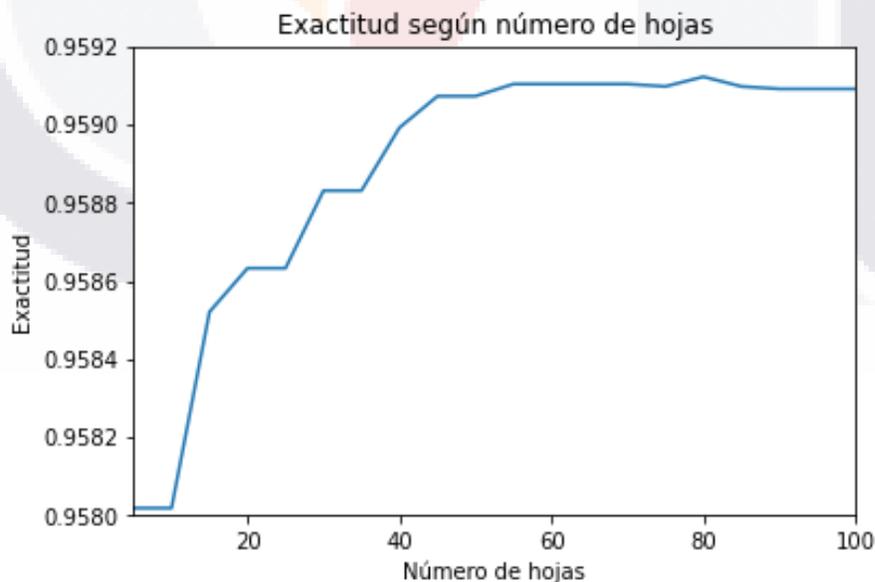


Fig. 39 Cambio en la exactitud según el número de hojas usadas

Por lo anterior, el modelo utilizado es un modelo de árbol de decisión con 644,856 datos, con 80 hojas, límites para nodos y hojas del 2% y 1% respectivamente con las variables: *SEXO*, *EDAD*, *TIPO\_PACIENTE*, *INTUBADO*, *NEUMONIA*, *DIABETES*, *EPOC*, *INMUSUPR*, *HIPERTENSION*, *OTRA\_COM*, *CARDIOVASCULAR*, *REANL\_CRONICA*, *OBESIDAD*, *UCI*. Igualmente, se estableció como variable objetivo *FECHA\_DEF*.

Con estas características el modelo creado se muestra en la Fig. 40, tiene como nodo principal la variable *TIPO\_PACIENTE* y por los límites de muestras para hojas y nodos solamente tiene una profundidad de 54 hojas. De manera general se concluye que si una persona es hospitalizada esta morirá, de lo contrario la vida de la persona no estará en peligro a menos que desarrolle neumonía. En las figuras 41, 42 y 43 se muestran los subárboles que lo componen; la Tabla 22 indica los valores de las métricas para el árbol y la Fig. 44 muestra la matriz de confusión del modelo.

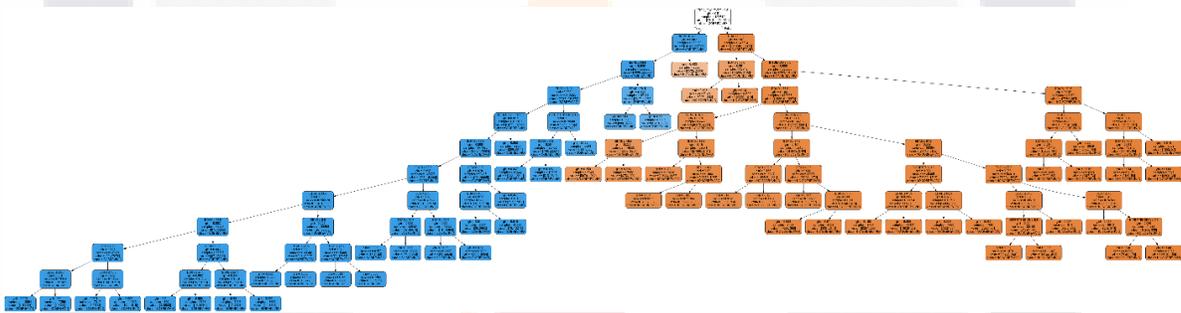


Fig. 40 Modelo completo desarrollado para las variables mencionadas



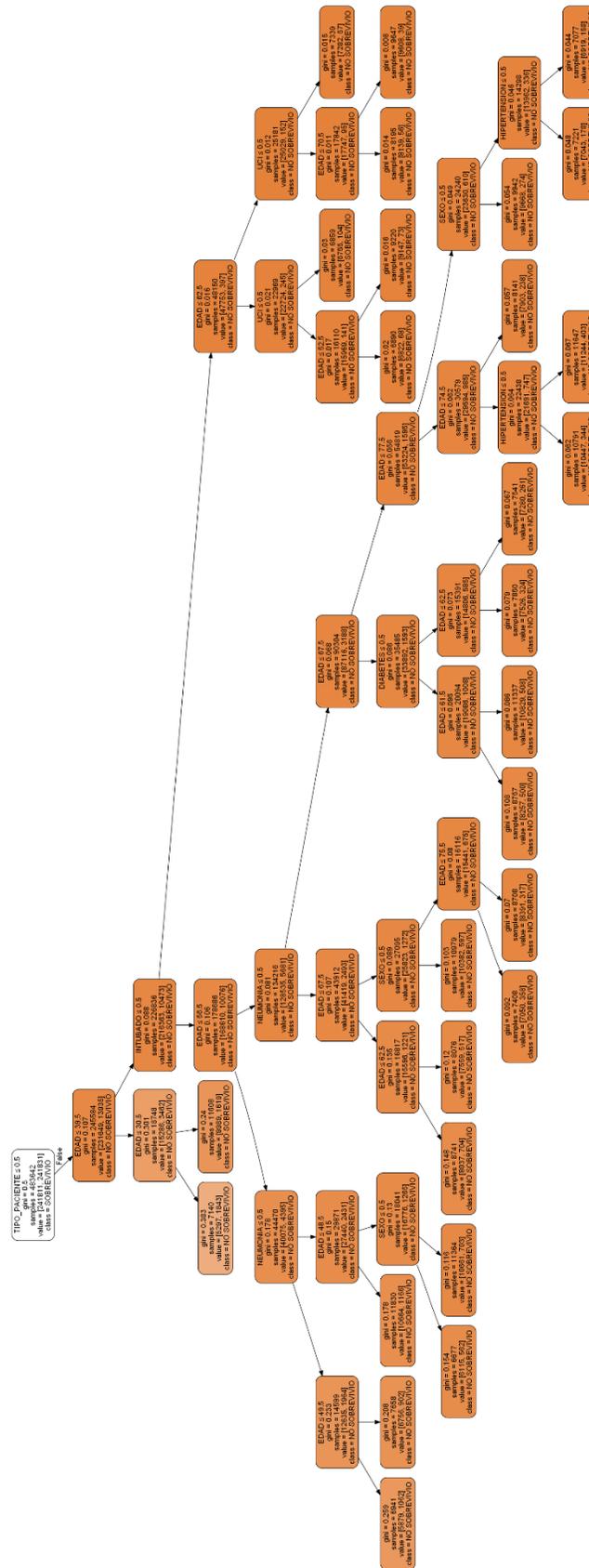


Fig. 42 Subárbol 2. Casos de fallecen por COVID-19

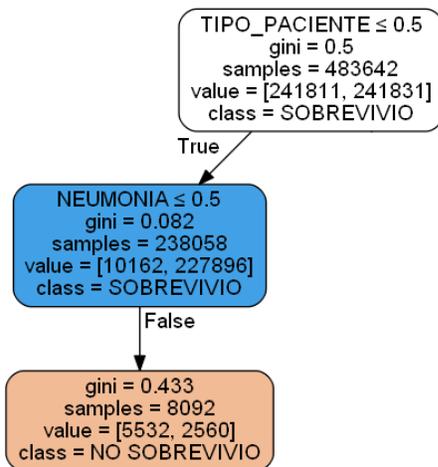


Fig. 43 Subárbol 3. Casos que a pesar de no ser hospitalizado fallecen después de desarrollar neumonía

Tabla 22. Resultados para las métricas del modelo

Métrica	Valor
Exactitud:	95.61%
Precisión:	97.98%
Exhaustividad:	93.15%
Valor F:	95.50%

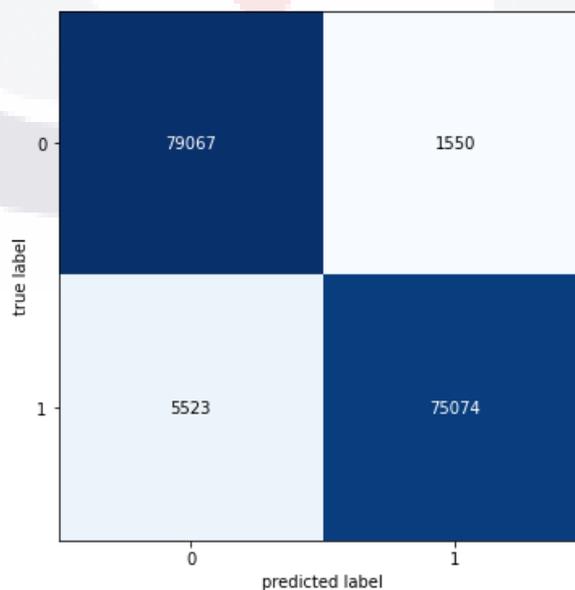


Fig. 44 Matriz de confusión del modelo

#### 5.1.4 Resultados de vacunación

##### Mundial

De acuerdo con el área de salud pública las campañas de vacunación tienen como objetivo controlar, erradicar de ser posible, los virus que afectan al ser humano y para medir el éxito que tienen es necesario enfocarse en aspectos más allá de lo económico. Obtener esta medición de éxito puede estimarse mediante el cálculo de las personas que evitaron la muerte o secuelas graves, además de tener en cuenta los posibles efectos adversos causados por las vacunas [28].

Las reacciones adversas que se presentan después de recibir una vacuna se clasifican según la gravedad con la que se presenta y el efecto que tiene sobre la salud de la persona que fue inoculada, las clases son: reacción adversa (*adverse reaction AR*), evento adverso (*adverse event AE*), evento adverso serio (*serious adverse event SAE*), evento adverso médicamente tratado (*medically-attended adverse event MAAE*). De acuerdo con los protocolos de las vacunas Pfizer, Moderna y Janssen los eventos anteriores pueden definirse de la siguiente forma [10]–[12]:

- *AE*: se define como cualquier ocurrencia médica inesperada que pudiera ser asociada con algún medicamento.
- *AR*: a esta categoría pertenecen todos los *AEs* para los que se ha encontrado evidencia que sugiera una relación causal entre la aplicación de un medicamento, en este caso la vacuna, y el evento observado.
- *SAE*: son tanto los *AEs* como los *ARs* que afectan en gran medida a la calidad de vida de las personas que los presentan. Ejemplos de estos efectos negativos son la muerte, hospitalizaciones o cualquier evento que impida a la persona continuar con su vida cotidiana.
- *MAAE*: en esta categoría se incluyen los *AEs* que requieren que la persona que recibió la vacuna asista a consulta cuando no debería hacerlo sin que los efectos sean peligrosos para la salud de la persona.

También es importante resaltar que en el caso de las *AEs* es necesario asignarles una graduación para conocer el impacto que tienen en la salud de las personas. Para esto se considera la Tabla 23 [11]:

Tabla 23. Grados de severidad de un evento adverso

Grado 0	Grado 1	Grado 2	Grado 3	Grado 4
No se presentó <sup>19</sup> .	Leve. Los síntomas no interfieren con la vida cotidiana de la persona.	Moderados. Los síntomas interfieren con la vida cotidiana de la persona.	Severos. Los síntomas causan que la persona no pueda realizar sus actividades cotidianas.	Potencial amenaza a la vida. Las personas son incapaces de realizar tareas básicas de cuidado personal o necesitan alguna intervención de emergencia.

En el caso específico de las vacunas de COVID-19 su nivel de éxito se podría calcular estimando cuantas personas evitaron enfermarse de gravedad o morir, no obstante, esto no es sencillo por la falta de información disponible y porque cualquiera pudo haber pertenecido al grupo de riesgo de esta enfermedad. A diferencia del caso anterior, conocer los efectos adversos relacionados a las vacunas es más sencillo gracias a las plataformas de vigilancia que hay en el mundo las cuales se encargan de hacer públicos los reportes que reciben.

Al 20 de marzo del 2023 las plataformas internacionales de vigilancia de efectos adversos de las vacunas reportaban las estadísticas incluidas en la Tabla 24 respecto a las vacunas de emergencia de COVID-19.

Tabla 24. Plataformas de vigilancia para reportes de vacunación en el mundo

Plataforma	Región de reporte	Número de reportes
<b>VigiAccess</b>	Mundial. Más de 170 países	4,949,624
<b>VAERS</b>	Estados Unidos	1,533,588
<b>EudraVigilance</b>	Espacio Económico Europeo	2,219,201

Para poder trabajar correctamente con la información disponible en estas plataformas es importante considerar dos cosas: 1) la existencia de estos reportes no significa que la vacuna haya sido la causante de los síntomas reportados, para poder determinar la relación

<sup>19</sup> Este grado únicamente se considera para los eventos adversos que se espera presenten las personas que reciben algún tratamiento médico. Estos eventos esperados se pueden encontrar en los protocolos de las vacunas.

entre la vacuna y los síntomas es necesario recurrir a un comité médico que analice todos los detalles del caso; y 2) hay más síntomas que reportes porque las personas pueden padecer más de una reacción adversa. Después de las consideraciones anteriores la Tabla 25 puede desglosarse de la siguiente forma:

Tabla 25. Eventos más reportados por vacuna en las plataformas de vigilancia

Plataforma	Vacunas para las que se aceptan reportes	Número de reportes	Eventos más reportados
<b>VigiAccess</b>	Acepta reportes para todas las vacunas disponibles	4,949, 624	Desórdenes generales y de aplicación (26%) Desórdenes del sistema nervioso (16%) Desórdenes músculo- esqueléticos y de tejido conectivo (10%)
<b>VAERS</b>	Pfizer, Moderna, Janssen, Novavax.	1,533, 588 (reportes) 4,212,568 (síntomas registrados)	Dolor de cabeza (16.12%) Fiebre (14.07%) Fatiga (14.06%) Dolor (11.82%) Escalofríos (11.37%)
	Pfizer y Moderna bivalentes	24,424 (reportes) 90,104 (Síntomas registrados)	Sin eventos adversos (13.71%) COVID-19 (11.32%) Test positivo (10.29%) Fatiga (8.58%) Fiebre (8.42%)
<b>EudraVigilance</b>	Moderna (Elasomeran)	371,708	Desórdenes generales y de aplicación (238,913) Desórdenes del sistema nervioso (147,941) Desórdenes músculo- esqueléticos y de tejido conectivo (112,563)
	Moderna (Elasomeran, Imelasomeran)	7,452	Desórdenes generales y de aplicación (4,625) Desórdenes del sistema nervioso (3,446) Desórdenes músculo- esqueléticos y de tejido conectivo (2,818)
	Moderna (Elasomeran, Davesomeran)	392	Desórdenes generales y de aplicación (203) Desórdenes del sistema nervioso (106) Infecciones (96)
	Pfizer (Tozinameran)	1,209,065	Desórdenes generales y de aplicación (710,580) Desórdenes del sistema nervioso (421,229) Desórdenes músculo- esqueléticos y de tejido conectivo (310,007)

Pfizer (Tozinameran, Riltozinameran)	5,863	Desórdenes generales y de aplicación (3,916) Desórdenes del sistema nervioso (2,591) Desórdenes músculo- esqueléticos y de tejido conectivo (2,192)
Pfizer (Tozinameran, Famtozinameran)	7,103	Desórdenes generales y de aplicación (4,458) Infecciones (2,332) Desórdenes del sistema nervioso (1,823)
AstraZeneca	544,839	Desórdenes generales y de aplicación (382,452) Desórdenes del sistema nervioso (281,156) Desórdenes músculo- esqueléticos y de tejido conectivo (205,623)
Janssen	71,165	Desórdenes generales y de aplicación (49.652) Desórdenes del sistema nervioso (30,760) Desórdenes músculo- esqueléticos y de tejido conectivo (22,350)
Novavax	1,584	Desórdenes generales y de aplicación (939) Desórdenes del sistema nervioso (743) Desórdenes gastrointestinales (277)
Valneva	27	Lesiones, envenenamiento o complicaciones (20) Procedimientos médicos / cirugías (13) Desórdenes generales y de aplicación (13)
VidPrevtyl Beta	3	Desorden del sistema inmune (2) Desorden gastrointestinal (1)

De la tabla anterior es necesario resaltar que las vacunas que presentaron más reportes fueron Pfizer, AstraZeneca y Moderna lo cual es congruente con las estadísticas que colocan a estas vacunas como las más aplicadas en el mundo [47]. Igualmente hay que mencionar que la mayoría de los síntomas reportados fueron leves y no afectaron la salud de las personas que realizaron los reportes, no obstante, se han publicado algunos artículos que hablan de la relación entre algunas vacunas de COVID-19 con enfermedades más graves como cardiopatías o coágulos sanguíneos dependiendo de la vacuna que recibió la persona.

En el caso de las vacunas de ARN mensajero se menciona que la miocarditis es un SAE poco común que se ha reportado con anterioridad para diferentes vacunas (hepatitis B, influenza y varicela), sin embargo, después de casi 300 millones de dosis de este tipo de

vacunas el sistema de VAERS registraba 1226 posibles casos de miocarditis o pericarditis<sup>20</sup> especialmente en hombres jóvenes que habían recibido la segunda dosis de las vacunas Pfizer o Moderna. A pesar de que estas cifras no pueden ser utilizadas para afirmar que las vacunas de ARN mensajero causan estas enfermedades sí sirven para levantar una alerta que los comités de medicina se encargaran de analizar más adelante [48].

Después de realizar las investigaciones pertinentes un estudio reportó que de los 1226 casos probables de miocarditis o pericarditis solo 323 cumplen con la definición de las enfermedades. También se reportó que el 96% de estos casos confirmados fueron hospitalizados pero la mayoría de ellos recibieron el alta pocos días después sin síntomas. Si bien la tasa estimada es baja, 12.6 casos por millón de personas que recibieron la segunda dosis, se observa una clara afectación de hombres entre 12 y 39 años sin comorbilidades que llama la atención de los investigadores [48]. En la Tabla 26 se muestran las estadísticas de casos sospechosos de miocarditis o pericarditis reportados al sistema de VAERS después de recibir la segunda dosis de alguna vacuna de ARN mensajero y se pueden comparar con las estadísticas que se considerarían normales.

*Tabla 26. Casos de miocarditis o pericarditis esperados y observados después de la segunda dosis de las vacunas de ARN mensajero considerando que los síntomas aparecen en los primeros 7 días<sup>21</sup> [48]*

Grupo de edad	Mujeres			Hombres		
	Dosis administradas	Casos esperados	Casos registrados	Dosis administradas	Casos esperados	Casos registrados
12 a 17	2,189,726	0 a 2	19	2,039,871	0 a 4	128
18 a 24	5,237,262	1 a 6	23	4,337,287	1 a 8	219
25 a 29	4,151,975	0 a 5	7	3,625,574	1 a 7	59
30 a 39	9,356,296	2 a 18	11	8,311,301	2 a 16	61
40 a 49	9,927,773	2 a 19	18	8,577,766	2 a 16	34
50 a 64	18,696,450	4 a 36	18	16,255,927	3 a 31	18
65+	21,708,975	4 a 42	10	18,041,547	3 a 35	11

Como se puede observar en la tabla anterior existe una clara afectación entre las personas menores de 50 años, principalmente hombres, que reciben la segunda dosis de las vacunas

<sup>20</sup> La CDC define a la miocarditis como la inflamación del músculo cardiaco (miocardio) y a la pericarditis como la inflamación de la membrana que recubre al corazón (pericardio). Ambos casos ocurren cuando el sistema inmune reacciona negativamente a una infección o a otro desencadenante.

<sup>21</sup> Hubo 9 casos en los que no fue identificado el grupo de edad al que pertenecían las personas.

de ARN mensajero. Esta observación no es exclusiva de los Estados Unidos sino que ya ha sido reportada en otros países como Israel y Corea del Sur.

En el caso de Israel el Ministerio de Salud reportó que entre diciembre del 2020 y mayo del 2021 se registraron 275 casos de miocarditis de los cuales 148 pueden estar relacionados con las vacunas por el tiempo en el que se hizo el reporte, Tabla 27 [49].

Tabla 27. Reportes de miocarditis en Israel [49]

Periodo en que se hizo el reporte	Vacunas aplicadas	Casos reportados	Observaciones
Después de la primera dosis	5,401,150	27	De los casos reportados 11 tenían alguna condición médica previa
Después de la segunda dosis	5,049,424	121	De los casos reportados 60 tenían alguna condición médica previa

La información disponible lleva al gobierno de Israel a concluir que probablemente existe una conexión entre la segunda dosis de las vacunas de ARN mensajero y la aparición de miocarditis leve en hombres entre 16 y 19 años por lo cual consideran que debe analizarse la decisión de aplicar esta vacuna a jóvenes [49].

En el caso del estudio realizado para la población de Corea del Sur, al 31 de mayo del 2022, se confirmaron un total de 480 casos de miocarditis relacionados a las vacunas de 1,533 casos sospechosos registrados por la KDCA (*Korean Disease Control and Prevention Agency*). De estos casos confirmados se observó que la mayoría de los casos eran de hombres que habían recibido alguna dosis de una vacuna de ARN mensajero. Esos resultados indican que en la población coreana la miocarditis relacionada con las vacunas se presenta con una frecuencia de 1.08 casos por cada 100,000 personas vacunadas, no obstante, en el caso de los hombres este valor aumenta a 1.35 casos por 100,000 vacunados y para las mujeres disminuye a 0.82 casos por 100,000 vacunados [50].

Al igual que para los países anteriores se observó una mayor incidencia de miocarditis en hombres jóvenes siendo el grupo de edad de 12 a 17 años el más afectado (5.29 casos por 100,000 vacunados). Si bien la mayoría de los casos fueron leves se reportó que el 19.8% de los casos fueron severos y el 4.4% de todos los casos fallecieron. En el caso de los fallecidos se observó que 8 de ellos presentaron muerte súbita cardiaca relacionada con la miocarditis, tenían menos de 45 años y habían recibido alguna dosis de una vacuna de ARN [50]. De forma general los casos de miocarditis relacionada con las vacunas se presentaron

como se muestra en la Tabla 28, mientras que las características de los casos severos y los 8 casos que presentaron muerte súbita cardiaca se encuentran en las Tablas 29 y 30 respectivamente.

Tabla 28. Casos de miocarditis registrados por grupo de edad [50]

Grupo de edad	Mujeres			Hombres		
	Dosis administradas	Casos registrados	Incidencia por 100,000 vacunados	Dosis administradas	Casos registrados	Incidencia por 100,000 vacunados
12 a 17	1,105,176	16	1.45	1,171,550	62	5.29
18 a 29	3,544,543	46	1.30	3,920,277	115	2.93
30 a 39	2,977,337	34	1.14	3,381,709	52	1.54
40 a 49	3,797,674	36	0.95	3,954,215	27	0.68
50 a 59	4,173,776	36	0.86	4,245,958	19	0.45
60 a 69	3,431,287	8	0.23	3,297,193	15	0.46
70+	3,089,034	5	0.16	2,186,975	9	0.41

Tabla 29. Características de los casos severos y no severos de miocarditis [50]

	Casos severos de miocarditis (n = 95)	Casos no severos de miocarditis (n = 385)
<b>Grupo de edad</b>		
12 a 17	15 (16%)	63 (16%)
18 a 29	23 (24%)	138 (36%)
30 a 39	14 (15%)	72 (19%)
40 a 49	18 (19%)	45 (12%)
50 a 59	11 (12%)	44 (11%)
60 a 69	10 (11%)	13 (3%)
70+	4 (4%)	10 (3%)
<b>Sexo</b>		
Mujeres	46 (48%)	135 (35%)
Hombres	49 (52%)	250 (65%)

Tabla 30. Características de los 8 pacientes con muerte súbita cardíaca [50]

Edad	Sexo	Días que transcurrieron desde la aplicación de la vacuna	Tipo de vacuna	Dosis que recibió
22	H	6	BNT162b2	1
30	M	3	BNT162b2	1
45	H	3	BNT162b2	2
25	H	3	BNT162b2	2
45	H	3	BNT162b2	2
36	M	2	mRNA-1273	1
33	H	1	mRNA-1273	2
33	H	3	mRNA-1273	2

De las tablas anteriores se puede observar que los casos de miocarditis son más comunes entre jóvenes y que el riesgo de padecer esta enfermedad disminuye a mayor edad; en el caso de las personas que fallecieron el promedio de edad era de 33.6 años, 6 de ellos eran hombres, cinco de ellos fueron vacunados con Pfizer y tres con Moderna y el 62.5% falleció en promedio 2.6 días después de haber recibido la segunda dosis de la vacuna. Esto último es relevante porque concuerda con lo observado en otros países sobre el aumento del riesgo de padecer miocarditis después de recibir la segunda dosis de una vacuna de ARN mensajero, en la Tabla 31 se muestran los casos por tipo de vacuna registrado y de acuerdo con la dosis que recibió la persona.

Tabla 31. Casos de miocarditis por vacuna y dosis recibida [50]

Tipo de vacuna	General		Primera dosis		Segunda dosis		Tercera dosis	
	n	Incidencia por 100,000 vacunados	n	Incidencia por 100,000 vacunados	n	Incidencia por 100,000 vacunados	n	Incidencia por 100,000 vacunados
<b>BNT162b2</b> <sup>22</sup>	306	1.23	133	0.54	140	0.60	33	0.30
<b>mRNA-1273</b> <sup>23</sup>	156	2.30	68	1.00	77	1.16	11	0.19
<b>ChAdOx1</b> <sup>24</sup>	15	0.14	5	0.05	10	0.09	-	-
<b>Ad26</b> <sup>25</sup>	3	0.20	3	0.20	-	-	0	0.00
<b>Total</b>	480	1.08	209	0.47	227	0.55	44	0.24

<sup>22</sup> Nombre que recibe la vacuna Pfizer en los estudios

<sup>23</sup> Nombre de la vacuna Moderna en estudios clínicos

<sup>24</sup> Nombre que recibe la vacuna AstraZeneca en estudios. También se le conoce como AZD1222

<sup>25</sup> Nombre para los estudios clínicos de la vacuna Johnson & Johnson

Estos no han sido los únicos eventos adversos severos reportados, para el caso específico de la vacuna de AstraZeneca diversos periódicos como el New York Times y CNN han advertido de la posible relación entre la formación de coágulos sanguíneos y la aplicación de esta vacuna. Estos efectos han sido reportados principalmente en países europeos quienes han decidido detener su aplicación o restringir su uso en personas jóvenes para evitar estos casos [51].

Entre los países que han estado vigilando la aparición de este efecto adverso se encuentran el Reino Unido, Alemania, Noruega y Dinamarca. En el primer caso se reportaron 30 casos de personas que desarrollaron algún tipo de coágulo sanguíneo poco tiempo después de haber recibido la vacuna de AstraZeneca, además, la Agencia Reguladora de Medicinas y Productos Sanitarios (*MHRA*) comunicó que se registraron cuatro muertes por coágulos sanguíneos. Cabe resaltar que al momento que se escribió la noticia los 30 reportes que se mencionan se presentaron después de haber vacunado a 15.8 millones de personas con esta vacuna y hasta el 31 de agosto del 2022 su uso no había sido suspendido en el Reino Unido [34], [51].

En el caso de Alemania se reportaron una cifra similar de casos a pesar de haber administrado menos dosis de vacunas que el Reino Unido. En este país se tuvieron 31 reportes, después de haber vacunado a 2.7 millones de personas, de los cuales 9 fallecieron [51].

Para Noruega y Dinamarca no se publican las cifras de casos reportados pero sí se indica que se presentaron defunciones en estos países a causa de la aparición de estos coágulos (tres en Noruega y una en Dinamarca) [51]. En este caso, Dinamarca detuvo la aplicación de esta vacuna después de publicar que la incidencia de estos casos era más alta de la esperada y reportó una cifra de 1 caso en 40,000 personas vacunadas [52].

A diferencia de lo observado con las vacunas de ARN mensajero para este evento adverso se reportó que la población más afectada fueron mujeres jóvenes (entre 20 y 65 años), por ejemplo, en Alemania de los 31 reportes registrados 29 corresponden a mujeres [51].

Si bien es esperado que una vacuna presente efectos adversos, los reportes recibidos por la plataforma VigiAccess<sup>26</sup> para estas vacunas son mucho mayores a los recibidos para otras vacunas del esquema nacional de vacunación (Tabla 32) por lo cual es conveniente

---

<sup>26</sup> Datos al 7 de septiembre del 2023

estar alerta de más posibles casos. Igualmente aunque el número de efectos adversos reportado parece minúsculo en comparación con todas las dosis de vacunas aplicadas ha sido una cantidad suficiente para que estos efectos raros sean agregados como contraindicación en los protocolos de las vacunas.

Tabla 32. Reportes de efectos adversos registrados en VigiAccess para las vacunas del esquema nacional de vacunación [53]

Vacuna	Reportes
BCG <sup>27</sup>	39,925
Hepatitis B	111,387
Neumococo	266,455
Rotavirus	79,765
SRP <sup>28</sup>	171,883
SR <sup>29</sup>	10,646
Td <sup>30</sup>	50,797
DPT	235,659
Tdpa <sup>31</sup>	235,659
VPH	127,292
Influenza	311,980
Poliomielitis	135,633
Varicela	218,069
Hepatitis A	51,490

### México

Al 7 de octubre del 2022 la página de *OWID* reportaba en México la aplicación de más de 220 millones de dosis de vacunas desde que se autorizaron en el país y se estima que 97.18 millones de personas han recibido al menos una dosis de alguna de las vacunas disponibles. Igualmente, se reporta que del total de vacunas aplicadas el 74.45% de ellas corresponden a alguna de las dosis del protocolo de vacunación inicial, mientras que el porcentaje restante corresponde a vacunas de refuerzo.

<sup>27</sup> Tuberculosis.

<sup>28</sup> Sarampión, rubéola, parotiditis o paperas, comúnmente llamada triple viral. En inglés MMR.

<sup>29</sup> Sarampión, rubéola. También conocida como doble viral.

<sup>30</sup> Tétanos, difteria.

<sup>31</sup> Tétanos, difteria y tos ferina. Tdap en inglés.

Tomando como referencia la publicación realizada por el portal de periodismo independiente Serendipia del día 23 de diciembre del 2022, se observan las siguientes cifras de aplicación por el tipo de vacuna en el país:

Tabla 33. Dosis aplicadas en México por vacuna [54]

Vacuna	Dosis aplicadas (en millones)	Porcentaje
<b>AstraZeneca</b>	54.78	36.44%
<b>Pfizer</b>	53.48	35.57%
<b>SinoVac</b>	18.6	12.37%
<b>Sputnik V</b>	11	7.34%
<b>Cansino</b>	7.85	5.22%
<b>Moderna</b>	3.35	2.23%
<b>Janssen</b>	1.25	0.83%

Con el fin de conocer datos referentes a eventos adversos causados por las vacunas y defunciones por COVID-19 de personas vacunadas se realizó una solicitud de acceso a la información a la Secretaría de Salud mediante la Plataforma de Transparencia, dicha solicitud fue registrada con el folio: 330026923002342. Esta solicitud fue turnada a la Dirección General de Información en Salud y la Dirección General de Epidemiología, las cuales respondieron que en las bases de datos disponibles no tienen registros de eventos adversos y se compartió un archivo en formato .XLSX en donde se desglosan los datos de defunciones.

Este archivo de *Excel* ha servido como una base de datos complementaria a las dos principales con las que se estuvo trabajando y de esta fue posible obtener gráficas para conocer los decesos por COVID-19 de personas vacunadas. En la Fig. 45 se observa como de los 20,746 registros de defunciones considerados el 60.3% de ellos corresponde a hombres.

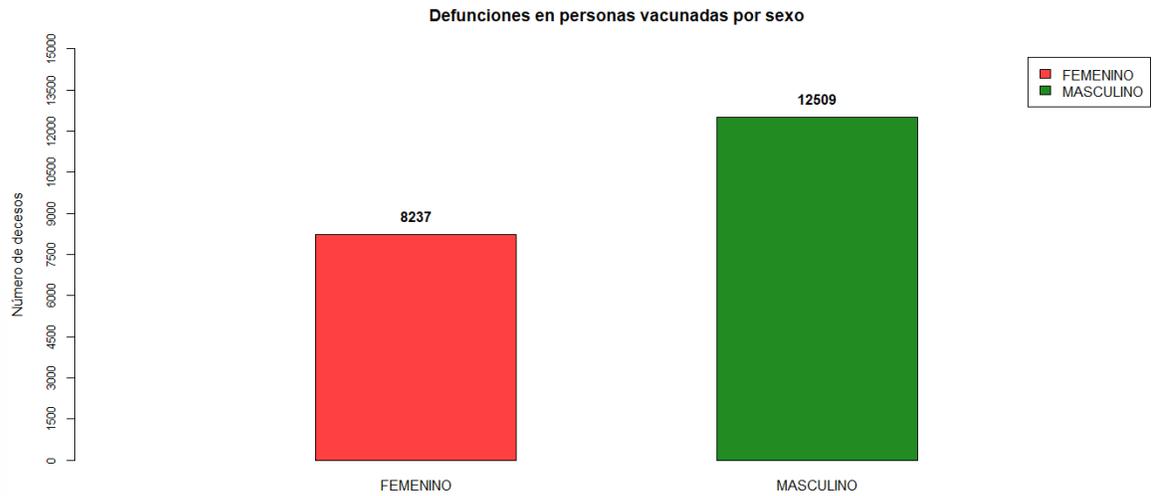


Fig. 45 Defunciones por COVID-19 de personas vacunadas en México por sexo

En el caso de la Fig. 46 se observó que las defunciones de personas vacunadas se presentaron principalmente en personas de la tercera edad.

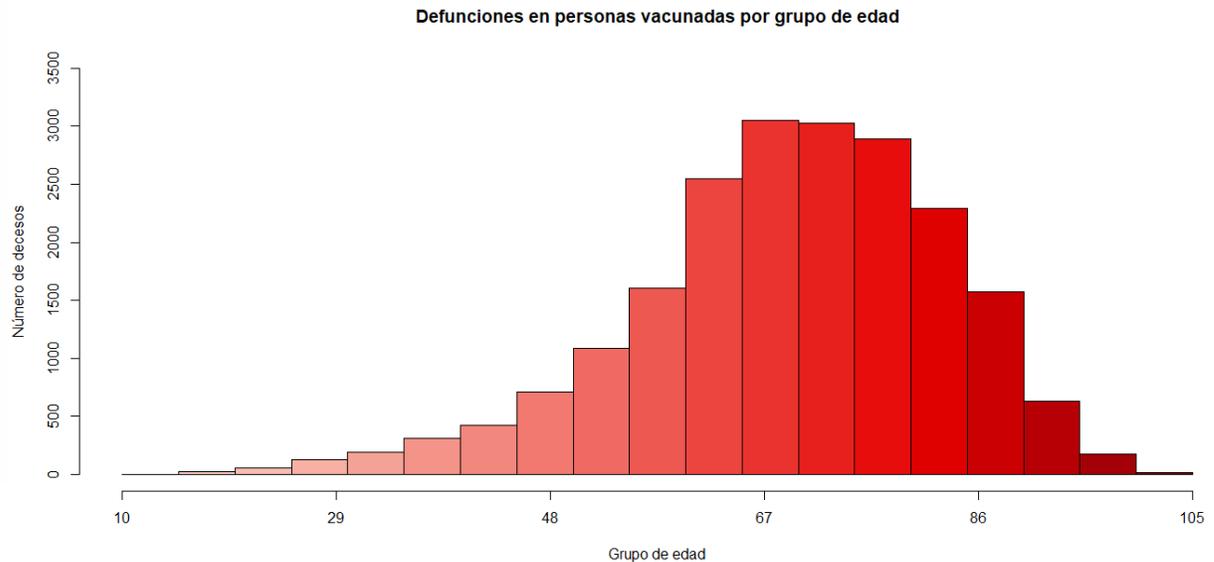


Fig. 46 Defunciones por COVID-19 de personas vacunadas en México por grupo de edad

Finalmente la Fig. 47 es un desglose de las defunciones de personas según el tipo de vacuna que recibieron. Se observó que las vacunas más comunes entre los difuntos fueron las vacunas Pfizer BioNTech, AstraZeneca y Sinovac, estos resultados fueron congruentes con los números compartidos por el sitio de *Serendipia* debido a que estas tres fueron las

vacunas más aplicadas en todo el país siendo que el 84.38% de la población vacunada recibió alguna de estas tres vacunas. Por el contrario, la vacuna Cansino fue recibida por el 15.03% de los fallecidos siendo que únicamente se aplicó al 5.22% de la población vacunada. Finalmente, es necesario remarcar que para el 5.67% se desconoce qué vacuna recibieron.

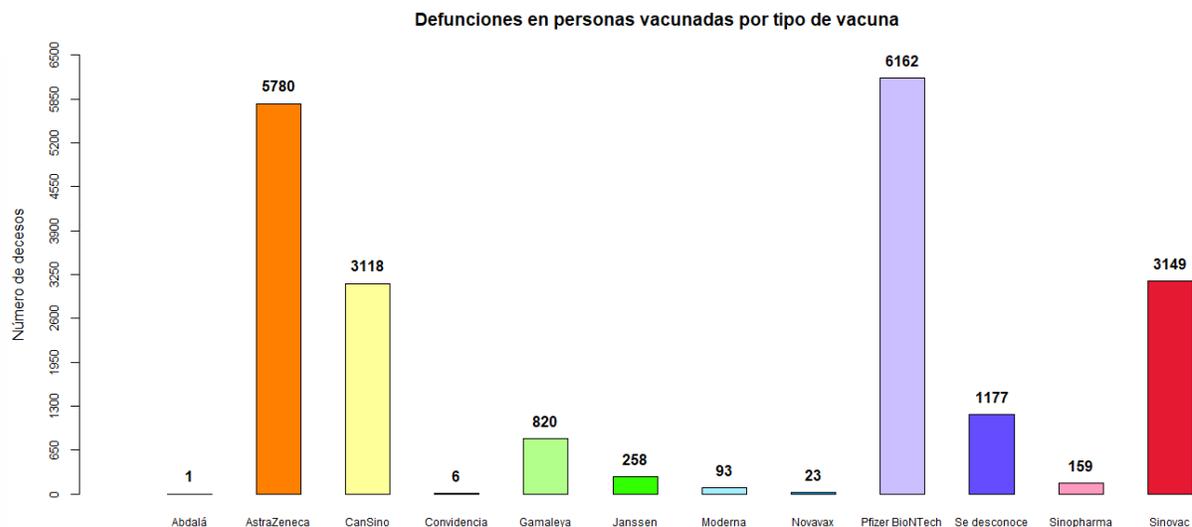


Fig. 47 Defunciones por COVID-19 de personas vacunadas en México por tipo de vacuna que recibieron

## Discusión de resultados

Para el apartado de discusiones se realizará un análisis más detallado de los resultados obtenidos en la sección anterior, además de compararlos con la información disponible en diversos artículos. Esta sección se divide de acuerdo con el tipo de información que se compara. Las secciones que se revisaran a continuación son: estadísticas mundiales, modelo y estadísticas nacionales y vacunación.

### 5.1.5 Estadísticas mundiales

A nivel mundial las estadísticas obtenidas permiten conocer que, como se ha mencionado en otros estudios [3]–[5], factores como la edad y las comorbilidades tienen un impacto en la mortalidad. El efecto de las comorbilidades se puede observar a nivel mundial en los países que, a pesar de tener una población joven presentan una mortalidad elevada,

TESIS TESIS TESIS TESIS TESIS

mientras que, el impacto de la edad se observa sobre todo en los países europeos que suelen tener las poblaciones con edades más avanzadas.

Otra observación importante es que la letalidad no resultó un concepto relevante para medir el impacto que tuvo la pandemia en los países debido a que es una medición que depende en gran medida de la capacidad que tienen los países para detectar casos positivos lo cual, en el caso de esta pandemia, equivale a la capacidad de realizar pruebas. La CDC indica que existen dos tipos diferentes de pruebas para detectar el virus: de PCR y de antígenos; ambas pruebas se realizan mediante la obtención de muestras nasofaríngeas sin embargo, la primera de estas es más recomendada para la detección tanto de casos sintomáticos como asintomáticos debido a que presenta los resultados más precisos aunque es más cara y toma más tiempo obtener los resultados [55], [56].

Considerando las características de las pruebas disponibles los países establecieron diferentes estrategias para la aplicación de estas y mediante la Fig. 19 es posible observar cómo los países con la mortalidad más baja fueron, en su mayoría, los que realizaron menos pruebas. Un ejemplo de esta situación es la comparación entre Yemen y Estados Unidos, inicialmente estos países presentaron una letalidad de 18.07% (la más alta del mundo) y 1.09% respectivamente, no obstante, al analizar la mortalidad los resultados cambiaron a 6.41 y 319.41 defunciones por cada 100,000 habitantes respectivamente; es importante resaltar la enorme diferencia entre el total de pruebas aplicadas por 100,000 habitantes en cada país, mientras que Estados Unidos aplicó 269,820 pruebas Yemen solamente aplicó 1,000.

En las Fig. 26 y 27 se comparan los casos diarios observados en México, Perú, China, Colombia y España. Esto permite observar que los picos de contagios en estos países se presentaron después de la aparición de la variante Ómicron, lo cual concuerda con la información publicada por la BBC que la describe como la variante más contagiosa registrada. Otra observación importante es que, después de España, México tuvo los picos de casos más altos de este grupo. Para las muertes diarias, Fig. 28 y Fig. 29, se observa que los picos de defunciones en estos países se atribuyen a las primeras variantes del virus que fueron más agresivas que la variante Ómicron [57].

Mediante la información de la base de datos de OWID también es posible analizar si existe alguna relación entre la severidad de las restricciones y los casos diarios observados en los países. En la Fig. 30 se observa el comportamiento de las restricciones respecto a los casos

diarios en México; inicialmente las restricciones fueron estrictas (un puntaje de 82.41 sobre 100) pero a partir del 2 de marzo del 2021 estos se comienzan a relajar considerablemente y desde ese momento continuaron disminuyendo excepto cuando se observa un aumento de casos.

Respecto a la vacunación, las gráficas parecen indicar que la campaña mundial no ha tenido el efecto deseado ya que los países con menor mortalidad también tienen menor porcentaje de vacunación, no obstante, al ser información que no considera otras características de esta campaña no se pueden utilizar para llegar a una conclusión.

#### **5.1.6 Modelo y estadísticas nacionales**

En el caso específico de México, en la Fig. 31 se observa el comportamiento de los casos nuevos registrados diariamente en el país y tiene la característica de aumentar el valor máximo con cada ola de contagios que se presenta desde el año 2020. Por el contrario, el comportamiento de los decesos diarios reportados (Fig. 32) muestra picos cada vez más pequeños a partir de la segunda ola de contagios. Este comportamiento puede ser causado por diferentes factores, por ejemplo, que las variantes más nuevas resultan menos agresivas pero más contagiosas, la reducción de restricciones o el efecto protector de las vacunas.

Estas estadísticas también permiten observar un problema importante dentro del país durante esta pandemia: la falta de equipo médico para tratar a quienes lo necesitan. De acuerdo con un comunicado publicado por la UAM, en abril del 2020 el país contaba únicamente con poco más de 5,000 ventiladores mecánicos para atender a toda la población y si bien las estadísticas podrían sugerir que no es un problema grave (de acuerdo con la Tabla 14, en el 2020 se requirieron en promedio 4,827 ventiladores por mes), otros países con más disponibilidad de equipo médico se declararon en crisis por no poder atender a todas las personas que lo requirieron [58].

Ahora bien, los resultados del modelo resaltan tres variables principales: edad, sexo y si la persona fue hospitalizada. Estos resultados resultan coherentes con las estadísticas observadas a través de los datos abiertos de gobierno donde la mayoría de las defunciones corresponden a personas de la tercera edad y hombres, además se observa que el 44.57% de las personas que fueron hospitalizadas fallecieron. En la Tabla 34 se muestran las estadísticas del grupo de fallecidos y se observa como la hospitalización, neumonía y comorbilidades son comunes entre las personas que fallecen.

Tabla 34. Características de las defunciones

Comorbilidades	Neumonía	Hospitalizados	UCI	Intubados
226,027	230,040	308,860	30,149	69,413

A pesar de que las comorbilidades no son tan relevantes en el modelo para predecir si una persona va a fallecer o no, es posible observar cómo afectan a las defunciones a través de las estadísticas que entrega la base de datos publicada por el gobierno de México y los conceptos de epidemiología como riesgo relativo, absoluto y atribuible.

De acuerdo con los conceptos estadísticos se observa que las comorbilidades no afectan de la misma manera a las personas, algunas presentaron un nivel de riesgo relativo muy alto mientras que otras, por el resultado obtenido, son consideradas como un factor de protección. Dentro de este análisis se observa que si bien la enfermedad renal crónica, cardiovascular y EPOC son las que tienen el riesgo más elevado de fallecer son la hipertensión, diabetes y obesidad las que causan más defunciones por lo comunes que son en la sociedad mexicana.

Estos resultados obtenidos concuerdan con otros estudios, por ejemplo en la Universidad Autónoma de San Luis Potosí se encontró que la elevada prevalencia de obesidad y diabetes en México pueden estar relacionadas con la alta mortalidad observada al inicio de la pandemia [7]. En el caso de un estudio publicado por la Universidad Veracruzana también obtuvieron que la variable más informativa era *TIPO\_PACIENTE* seguida por *NEUMONIA* y finalmente *RESULTADO* [9]. Para esta investigación no se consideró la última variable mencionada debido a que desde un inicio se buscaba que los registros considerados pertenecieran únicamente a las personas que tenían un resultado positivo a COVID-19 por lo cual, para este estudio *EDAD* y *SEXO* son otras variables muy relevantes para decidir cuánto riesgo tiene la persona de fallecer. Algo que tienen en común ambas investigaciones es que destaca el poco impacto que tiene el tabaquismo en las defunciones, en la investigación de la Universidad Veracruzana se indica que no es un factor concluyente para poder predecir una defunción y se cita otro estudio en donde se menciona que el tabaquismo no se relaciona con la gravedad con la que se desarrolla la enfermedad [9]; para esta investigación se observó un riesgo relativo de 1.36 y solamente se le pueden atribuir un 1.96% de las defunciones, esto sin considerar que las personas puedan tener una segunda comorbilidad. Respecto al elevado riesgo asociado a la enfermedad renal

crónica los resultados obtenidos concuerdan con lo que se menciona respecto a ser la enfermedad con el riesgo más elevado de evolucionar desfavorablemente [59].

Finalmente es importante mencionar que el árbol resultante en la investigación mencionada anteriormente tenía 3,612 nodos y una precisión de 91.33% mientras que el que se obtuvo en esta ocasión tenía 53 hojas y una precisión de 97.98% por lo que los resultados obtenidos fueron mejores.

### **5.1.7 Vacunación**

Como se observó en apartados anteriores, en el área de salud pública, las campañas de vacunación se han convertido en la opción más ética y rápida para inmunizar a una población sin que los individuos se vean expuestos a complicaciones por dicha enfermedad [19], por lo cual, es lógico que se haya buscado crear una vacuna para solucionar la emergencia sanitaria causada por la COVID-19.

Sobre la campaña de vacunación y los resultados que se tienen al momento se pueden rescatar varios puntos. El primero de ellos es respecto a los cambios que se tuvieron que hacer al protocolo para poder tener la vacuna en tiempo para ser de ayuda ya que, como se mencionó con anterioridad, el proceso de desarrollo de una vacuna suele ser extremadamente riguroso y largo para asegurar la calidad y seguridad de esta, sin embargo, en el caso de las vacunas de emergencia de COVID-19 las etapas de desarrollo de la vacuna se encimaron lo cual pudo haber puesto en riesgo a los voluntarios que participaron en las pruebas.

Otro punto muy relevante es que para crear una vacuna es necesario identificar el virus que se desea controlar y la población en riesgo para luego continuar con el protocolo para crear una vacuna. En este caso específico se sabe que el SARS-COV-2 es un virus respiratorio que puede afectar a toda la población, no obstante, para obtener su autorización de emergencia se puede observar que los protocolos únicamente consideraron a personas mayores de edad y que estuvieran clínicamente sanas, como se puede observar en la siguiente tabla:

Tabla 35. Criterios de exclusión para participar en el estudio de las vacunas

Vacuna	Rango de edad	Criterios de exclusión
<b>Pfizer [10]</b>	18 a 85 años	<ul style="list-style-type: none"> <li>• Condiciones médicas o psiquiátricas que puedan poner en riesgo la participación de los sujetos.</li> <li>• Padecer VIH, Hepatitis C o B.</li> <li>• Historial de alergias a los componentes de las vacunas.</li> <li>• Haber usado medicamentos para prevenir la COVID-19</li> </ul>
<b>Moderna [12]</b>	Mayores de 18 años	<ul style="list-style-type: none"> <li>• Embarazo o lactancia.</li> <li>• Historial de infección por COVID-19</li> <li>• Historial de alergias a los componentes de la vacuna.</li> <li>• Condiciones que afecten al sistema inmune.</li> <li>• Que haya recibido un tratamiento experimental en los días previos a este estudio.</li> <li>• Que se demuestre que no puede terminar el estudio</li> </ul>
<b>Johnson &amp; Johnson [11]</b>	18 a 60 años	<ul style="list-style-type: none"> <li>• Padecer alguna enfermedad severa.</li> <li>• Tener historial de alergias a los componentes de la vacuna.</li> <li>• Funcionamiento anormal del sistema inmune</li> <li>• Que hayan recibido algún medicamento experimental o alguna otra vacuna de COVID-19</li> <li>• Embarazo o lactancia.</li> <li>• Enfermedades psiquiátricas severas o adicciones.</li> <li>• Padecer VIH.</li> </ul>

Si bien estas restricciones sirven para tener un control sobre las posibles reacciones de las vacunas sobre los individuos aceptar su uso general sin antes haber realizado estudios para conocer cómo afectan a personas con comorbilidades los deja en una posición vulnerable. Por ejemplo, en un estudio para diferentes grupos de personas inmunocomprometidas realizado en Suecia se encontró que si bien la aplicación de la vacuna Pfizer no representa ningún riesgo para la salud de estas personas, la aparición de los anticuerpos es más baja en comparación con las personas sanas por lo que estos

grupos necesitan dosis diferentes para desencadenar la respuesta inmunológica deseada [60].

Otro problema muy serio observado durante este proceso fue que las muestras no fueron lo suficientemente extensas para poder detectar eventos adversos raros, como la vacuna Pfizer que se ha relacionado con la aparición de miocarditis o pericarditis en hombres jóvenes. Revisando el estudio de la vacuna Pfizer se afirma que para los resultados solamente se realizó un seguimiento de la evolución de los participantes dos meses después de la segunda dosis y que el estudio puede detectar eventos adversos siempre que su incidencia sea mayor o igual al 0.01%. Igualmente se comentó que a pesar de ser un estudio diseñado para mantener la vigilancia en ambos grupos durante dos años se le ofreció la vacuna al grupo que recibió el placebo después que los gobiernos dieron su autorización para la aplicación [10], esta situación no se observó con las vacunas Johnson & Johnson y Moderna por lo que los posibles eventos adversos que se puedan presentar después de dos meses de la aplicación de la vacuna Pfizer quedan pendientes de evaluación.

Respecto a los eventos adversos severos si bien es cierto que tienen una incidencia más elevada de la que se esperó, en el caso de las vacunas de ARN mensajero no parecen presentar una incidencia mayor a la que se ha visto con otras vacunas o en situaciones normales. Por ejemplo, tomando como referencia un artículo de 2003 en donde se realizó un seguimiento a un grupo de militares que comenzaron a presentar miocarditis después de recibir la vacuna de la varicela se encontró una incidencia de 7.8 por cada 100,000 vacunados, mientras que para esa fecha se estimaba una incidencia de 1 a 10 por cada 100,000 habitantes [61]. Igualmente en un estudio publicado en 2018 se analizaron las hospitalizaciones por pericarditis grave de personas mayores a 65 años durante 13 años (1999 a 2012) y se encontró una incidencia constante de 26 hospitalizaciones por cada 100,000 habitantes [62]. Estas cifras son similares a las observadas en Estados Unidos de 12.6 casos por cada millón de vacunados con la segunda dosis [48] y en Corea del sur de 1.08 por 100,000 vacunados [50]. Cabe resaltar que en todos los casos estos padecimientos afectan principalmente a los hombres.

A partir de la información anterior se han desarrollado estudios en donde se aborda un posible análisis costo beneficio para conocer el impacto de esta campaña de vacunación. El 23 de junio del 2021 en Estados Unidos la CDC presentó durante la reunión del ACIP

TESIS TESIS TESIS TESIS TESIS

(*Advisory Committee on Immunization Practices*) un análisis en donde a través de estimaciones basadas en incidencia de efectos adversos y estadísticas de hospitalización, intubación y defunción para esas fechas concluían que los casos severos que se evitaron superaron a los casos de miocarditis y pericarditis por lo cual, en este caso mencionan que se tuvo un efecto positivo [63]. Un segundo estudio enfocado a personas mayores de 50 años que han recibido una vacuna de ARN mensajero habla de cómo la vacuna es capaz de prevenir en un 89% la hospitalización, mientras que para UCI y urgencias los porcentajes suben a 90% y 91% respectivamente [64]. Igualmente en España se maneja que la vacunación hasta abril del 2021 pudo haber evitado 4,396 defunciones de adultos mayores [65], a pesar de sus estadísticas positivas los estudios recomiendan considerar más cuidadosamente el riesgo para poblaciones más jóvenes que se ha visto son los principales afectados.

A pesar de que organizaciones de salud como la OMS y la CDC consideran que las vacunas son seguras y se debe continuar con la inmunización de las personas, algunos doctores consideran que se debe suspender la aplicación de estas y evaluar su seguridad. Tal es el caso del Dr. Aseem Malhotra, un cardiólogo en el Reino Unido, quien atribuye el exceso de mortalidad por enfermedades cardíacas a las vacunas de ARN mensajero y resalta que los estudios iniciales de estas vacunas indican que es más posible presentar un efecto adverso serio que ser hospitalizado por COVID-19. Por lo anterior, hace un llamado para que las farmacéuticas entreguen todos los datos disponibles para realizar un análisis adecuado [66]. Estos argumentos también fueron replicados en México por el Dr. Alejandro Díaz Villalobos, quien en una plática ante el Senado mencionó que los datos publicados al momento por VAERS indicaban 1,521,347 reportes de efectos adversos y 34,384 defunciones, además de citar al Dr. Malhotra como referencia [67].

Para el caso específico de México la falta de datos públicos limita poder realizar más investigaciones respecto al impacto que tuvieron sobre las defunciones de la población y los posibles efectos adversos, no obstante las estadísticas que se tienen permiten conocer que solamente el 6.43% de los fallecidos habían sido vacunados, que la mayoría de las defunciones fueron de hombres y que igualmente esta situación afecta más a las personas mayores de 60 años. Igualmente las vacunas que más defunciones registraron fueron Pfizer, AstraZeneca y SinoVac, lo cual es un resultado esperado dado que fueron las vacunas más aplicadas en el país.

## 6 Conclusiones y trabajo a futuro

Este apartado muestra brevemente los resultados más importantes obtenidos mediante esta investigación e incluye sugerencias para mejorar los resultados en otros trabajos que se realicen más adelante.

### 6.1 Conclusiones

A través de los resultados obtenidos antes se realizan las siguientes observaciones: en promedio los países que tuvieron mortalidades más elevadas cumplían con las características de tener poblaciones con edades avanzadas y una presencia importante de algunas comorbilidades. En caso específico de México las estadísticas nacionales indican que las personas que más fallecen por COVID-19 suelen ser hombres mayores de 60 años y que la enfermedad renal crónica, EPOC y las enfermedades cardiovasculares son las que implican un riesgo más elevado para fallecer de COVID-19 aunque la hipertensión, diabetes y obesidad son las que causan más muertes.

Después de la creación de un modelo de árboles de decisión para predecir quienes son las personas que están en riesgo de fallecer es posible observar como el factor de riesgo más importante es ser hospitalizado, seguido de desarrollar neumonía, la edad y el sexo. Este árbol puede ser resumido en tres puntos: 1) si una persona es internada va a fallecer, 2) si una persona no es hospitalizada pero desarrolla neumonía esta va a fallecer y 3) si una persona no es hospitalizada y no desarrolla neumonía va a sobrevivir.

Respecto a la campaña de vacunación mundial es importante resaltar que la decisión de aplicar una vacuna se tomó en base al área de salud pública y que hasta la fecha diversas organizaciones de salud a nivel mundial respaldan el uso de estas, no obstante, la aparición de efectos adversos serios ha creado incertidumbre en algunos doctores respecto a su uso. Estadísticamente no es posible realizar un análisis preciso respecto al impacto que estas vacunas han tenido, por lo cual es ideal analizar la situación individualmente.

### 6.2 Trabajos a futuro

Como posible trabajo a futuro se podrían considerar los siguientes puntos: desarrollar un modelo enfocado a las causas que llevan a la hospitalización de las personas e implementar un algoritmo que pueda realizar el aprendizaje aunque las clases sean desiguales.

Igualmente es importante mencionar que la falta de información pública y el seguimiento de efectos adversos causados por las vacunas en México es un problema muy grave ya que impide identificar si hay alguna situación de riesgo para la población o si los resultados obtenidos son los esperados.



## Referencias

- [1] INEGI, "ESTADÍSTICA DE DEFUNCIONES REGISTRADAS 2020 (NOTA TÉCNICA)", [https://www.inegi.org.mx/contenidos/programas/mortalidad/doc/defunciones\\_registradas\\_2020\\_nota\\_tecnica.pdf](https://www.inegi.org.mx/contenidos/programas/mortalidad/doc/defunciones_registradas_2020_nota_tecnica.pdf).
- [2] INEGI, "ESTADÍSTICAS DE DEFUNCIONES REGISTRADAS 2021", [inegi.org.mx/contenidos/saladeprensa/boletines/2022/EDR/EDR2021\\_10.pdf](https://inegi.org.mx/contenidos/saladeprensa/boletines/2022/EDR/EDR2021_10.pdf).
- [3] E. L. de Moura *et al.*, "Comorbidities increase the risk of severity and mortality in COVID-19 patients: a systematic review and meta-analysis", *Research, Society and Development*, vol. 10, núm. 2, feb. 2021, doi: 10.33448/rsd-v10i2.12533.
- [4] D. Petrakis *et al.*, "Obesity - a risk factor for increased COVID-19 prevalence, severity and lethality (Review)", *Mol Med Rep*, vol. 22, núm. 1, may 2020, doi: 10.3892/mmr.2020.11127.
- [5] F. Zhou *et al.*, "Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study", *The Lancet*, vol. 395, núm. 10229, mar. 2020, doi: 10.1016/S0140-6736(20)30566-3.
- [6] O. Y. Bello-Chavolla *et al.*, "Predicting Mortality Due to SARS-CoV-2: A Mechanistic Score Relating Obesity and Diabetes to COVID-19 Outcomes in Mexico", *J Clin Endocrinol Metab*, vol. 105, núm. 8, ago. 2020, doi: 10.1210/clinem/dgaa346.
- [7] D. E. Noyola, N. Hermosillo-Arredondo, C. Ramírez-Juárez, y A. Werge-Sánchez, "Association between obesity and diabetes prevalence and COVID-19 mortality in Mexico: an ecological study", *The Journal of Infection in Developing Countries*, vol. 15, núm. 10, oct. 2021, doi: 10.3855/jidc.15075.
- [8] C. C. John, V. Ponnusamy, S. Krishnan Chandrasekaran, y N. R., "A Survey on Mathematical, Machine Learning and Deep Learning Models for COVID-19 Transmission and Diagnosis", *IEEE Rev Biomed Eng*, vol. 15, 2022, doi: 10.1109/RBME.2021.3069213.
- [9] N. Cruz-Ramírez, G.-J. Hoyos-Rivera, S.-L. Mestizo-Gutierrez, y H. Tapia-McClung, "Data Science: A Useful Tool for Understanding SARS-CoV-2 Information Facts", en *Moving From COVID-19 Mathematical Models to Vaccine Design: Theory, Practice and Experiences*, BENTHAM SCIENCE PUBLISHERS, 2022. doi: 10.2174/9789815051902122010011.
- [10] F. P. Polack *et al.*, "Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine", *New England Journal of Medicine*, vol. 383, núm. 27, dic. 2020, doi: 10.1056/NEJMoa2034577.
- [11] J. Sadoff *et al.*, "Safety and Efficacy of Single-Dose Ad26.COV2.S Vaccine against Covid-19", *New England Journal of Medicine*, vol. 384, núm. 23, jun. 2021, doi: 10.1056/NEJMoa2101544.

- [12] L. R. Baden *et al.*, “Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine”, *New England Journal of Medicine*, vol. 384, núm. 5, feb. 2021, doi: 10.1056/NEJMoa2035389.
- [13] K. Beattie, *Worldwide Bayesian Causal Impact Analysis of Vaccine Administration on Deaths and Cases Associated with COVID-19: A BigData Analysis of 145 Countries*. 2021. doi: 10.13140/RG.2.2.34214.65605.
- [14] Hernández A. Mauricio y Lazcano P. Eduardo, *Salud Pública. Teoría y práctica*, 1era ed. México: Manual Moderno, 2013.
- [15] Hernández A. Mauricio, *Epidemiología: diseño y análisis de estudios*, 1era ed. México: Editorial Médica Panamericana, 2007.
- [16] A. Moreno Altamirano, S. López Moreno, y A. Corcho Berdugo, “Principales medidas en epidemiología”, *Revista cubana de higiene y epidemiología*, Cuba, pp. 337–348, 2000.
- [17] J. D. Lemus, *Epidemiología y salud comunitaria*. Corpus Editorial, 2008. [En línea]. Disponible en: <https://elibro.net/es/lc/uaa/titulos/67144>
- [18] R. P. Henquin, *Epidemiología y estadística para principiantes*. Corpus Editorial, 2013. [En línea]. Disponible en: <https://elibro.net/es/lc/uaa/titulos/76940>
- [19] L. Salleras Sanmartí, *Vacunas preventivas, principios y aplicaciones*. Barcelona: Masson, 1998.
- [20] OMS, “Coronavirus”. Consultado: el 30 de abril de 2023. [En línea]. Disponible en: [https://www.who.int/es/health-topics/coronavirus#tab=tab\\_1](https://www.who.int/es/health-topics/coronavirus#tab=tab_1)
- [21] “Se cumple un año del primer caso confirmado de Covid-19 en el mundo”. Consultado: el 8 de noviembre de 2021. [En línea]. Disponible en: <https://www.forbes.com.mx/noticias-se-cumple-un-ano-del-primer-caso-de-coivid-19/>
- [22] V. Suárez, M. Suarez Quezada, S. Oros Ruiz, y E. Ronquillo De Jesús, “Epidemiología de COVID-19 en México: del 27 de febrero al 30 de abril de 2020”, *Rev Clin Esp*, vol. 220, núm. 8, pp. 463–463–471, 2020, doi: 10.1016/j.rce.2020.05.007.
- [23] “DOF - Diario Oficial de la Federación”. Consultado: el 8 de noviembre de 2021. [En línea]. Disponible en: [https://dof.gob.mx/nota\\_detalle.php?codigo=5603264&fecha=21/10/2020](https://dof.gob.mx/nota_detalle.php?codigo=5603264&fecha=21/10/2020)
- [24] A. Urrutia y N. Jiménez, “La Jornada Nacional de Sana Distancia se extiende al 30 de mayo”, *La Jornada*. Consultado: el 24 de septiembre de 2023. [En línea]. Disponible en: <https://www.jornada.com.mx/2020/04/17/politica/002n1pol>
- [25] “Seguimiento de las variantes del SARS-CoV-2”. Consultado: el 5 de junio de 2022. [En línea]. Disponible en: <https://www.who.int/es/activities/tracking-SARS-CoV-2-variants>
- [26] “Updated working definitions and primary actions for SARSCoV2 variants”. Consultado: el 30 de abril de 2023. [En línea]. Disponible en: <https://www.who.int/publications/m/item/updated-working-definitions-and-primary-actions-for--sars-cov-2-variants>

- [27] O’Toole Á, Hill V, Pybus OG, y et al, “Cov-Lineages”. Consultado: el 4 de junio de 2022. [En línea]. Disponible en: [https://cov-lineages.org/lineage\\_list.html](https://cov-lineages.org/lineage_list.html)
- [28] A. Escobar Gutiérrez, *Vacunas, ciencia y salud*. México: Secretaría de Salud, 1992.
- [29] OMS, “¿Cómo se desarrollan las vacunas?”, <https://www.who.int/es/emergencias/diseases/novel-coronavirus-2019/covid-19-vaccines/how-are-vaccines-developed>.
- [30] ONU, “Eficacia teórica, eficacia real y protección de las vacunas”, <https://www.who.int/es/news-room/feature-stories/detail/vaccine-efficacy-effectiveness-and-protection>.
- [31] “Gráfico: ¿Qué tan eficaces son las vacunas contra la COVID-19? | Statista”. Consultado: el 18 de mayo de 2023. [En línea]. Disponible en: <https://es.statista.com/grafico/23569/eficacia-estimada-de-los-candidatos-a-vacuna-en-la-prevencion-de-la-covid-19/>
- [32] “Vacuna Cansino: ¿efectiva vs variante delta? ¿requiere segunda dosis? | Tecnológico de Monterrey”. Consultado: el 18 de mayo de 2023. [En línea]. Disponible en: <https://conecta.tec.mx/es/noticias/nacional/salud/vacuna-cansino-efectiva-vs-variante-delta-requiere-segunda-dosis>
- [33] Gobierno de México, “Información de la vacuna”. Consultado: el 18 de mayo de 2023. [En línea]. Disponible en: <https://vacunacovid.gob.mx/informacion-de-la-vacuna/>
- [34] C. Zimmer, J. Corum, S.-L. Wee, y M. Kristoffersen, “Coronavirus Vaccine Tracker”, *The New York Times*. Consultado: el 29 de agosto de 2023. [En línea]. Disponible en: <https://www.nytimes.com/interactive/2020/science/coronavirus-vaccine-tracker.html>
- [35] J. Dunstan, A. Maas, y F. Tobar, *Una mirada a la era de los datos*. Editorial Universitaria de Chile, 2022. [En línea]. Disponible en: <https://elibro.net/es/lc/uaa/titulos/221785>
- [36] E. Soria Olivas, P. Rodriguez Belenguer, y Q. Garcia Vidal, *Inteligencia Artificial: casos practicos con aprendizaje profundo*. RA-MA Editorial, 2022. [En línea]. Disponible en: <https://elibro.net/es/lc/uaa/titulos/222670>
- [37] Information Resources Management Association (IRMA), *Machine Learning: Concepts, Methodologies, Tools and Applications*, 1st ed., vol. 1. Pennsylvania: IGI Global, 2011.
- [38] J. Bobadilla, *Machine Learning y Deep Learning: Usando Python, Scikit y Keras*. RA-MA Editorial, 2020. [En línea]. Disponible en: <https://elibro.net/es/lc/uaa/titulos/222698>
- [39] A. Shmilovici, “Support Vector Machines”, en *Data Mining and Knowledge Discovery Handbook*, O. Maimon y L. Rokach, Eds., Boston, MA: Springer US, 2005, pp. 257–276. doi: 10.1007/0-387-25465-X\_12.

- [40] G. A. Betancourt, "LAS MÁQUINAS DE SOPORTE VECTORIAL (SVMs)", *Scientia et Technica* Año XI, No 27. Consultado: el 1 de agosto de 2023. [En línea]. Disponible en: <https://revistas.utp.edu.co/index.php/revistaciencia/article/view/6895/4139>
- [41] Information Resources Management Association, *Machine Learning: Concepts, Methodologies, Tools and Applications*, vol. Volume II. 2011.
- [42] L. Rokach y O. Maimon, "Decision Trees", en *Data Mining and Knowledge Discovery Handbook*, O. Maimon y L. Rokach, Eds., Boston, MA: Springer US, 2005, pp. 165–192. doi: 10.1007/0-387-25465-X\_9.
- [43] J. M. Ortega Candel, *Big data, machine learning y data science en Python*. RA-MA Editorial, 2022. [En línea]. Disponible en: <https://elibro.net/es/lc/uaa/titulos/230290>
- [44] A. Bosch Rue, J. Casas Roma, y T. Lozano Bagen, *Deep learning: principios y fundamentos*. Editorial UOC, 2019. [En línea]. Disponible en: <https://elibro.net/es/lc/uaa/titulos/126167>
- [45] M. Vannucci, V. Colla, S. Cateni, y M. Sgarbi, "Artificial Intelligence Techniques for Unbalanced Datasets in Real World Classification Tasks", en *Machine Learning*, vol. 1, IGI Global, 2012. doi: 10.4018/978-1-60960-818-7.ch304.
- [46] Scikit Learn, "Decision Tree Classifier", <https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html#sklearn.tree.DecisionTreeClassifier>.
- [47] J. Holder, "Tracking Coronavirus Vaccinations Around the World", *The New York Times*. Consultado: el 29 de agosto de 2023. [En línea]. Disponible en: <https://www.nytimes.com/interactive/2021/world/covid-vaccinations-tracker.html>
- [48] B. Bozkurt, I. Kamat, y P. J. Hotez, "Myocarditis With COVID-19 mRNA Vaccines", *Circulation*, vol. 144, núm. 6, ago. 2021, doi: 10.1161/CIRCULATIONAHA.121.056135.
- [49] Ministry of Health of Israel, "Surveillance of Myocarditis (Inflammation of the Heart Muscle) Cases Between December 2020 and May 2021 (Including)". Consultado: el 29 de agosto de 2023. [En línea]. Disponible en: <https://www.gov.il/en/departments/news/01062021-03>
- [50] J. Y. Cho *et al.*, "COVID-19 vaccination-related myocarditis: a Korean nationwide study", *Eur Heart J*, vol. 44, núm. 24, jun. 2023, doi: 10.1093/eurheartj/ehad339.
- [51] A. Dewan y R. A. Greene, "Esto es lo que debes saber sobre el riesgo de coágulos de sangre y la vacuna de AstraZeneca contra el covid-19", *CNN en español*. Consultado: el 30 de agosto de 2023. [En línea]. Disponible en: <https://cnnespanol.cnn.com/2021/04/03/esto-es-lo-que-debes-saber-sobre-el-riesgo-de-coagulos-de-sangre-y-la-vacuna-astrazeneca-trax/>
- [52] BBC, "AstraZeneca vaccine: Denmark stops rollout completely", *BBC*. Consultado: el 30 de agosto de 2023. [En línea]. Disponible en: <https://www.bbc.com/news/world-europe-56744474>

- TESIS TESIS TESIS TESIS TESIS
- [53] Secretaría de Salud, “Esquema de Vacunación”, Gobierno de México. Consultado: el 6 de septiembre de 2023. [En línea]. Disponible en: <https://www.gob.mx/salud/articulos/esquema-de-vacunacion>
- [54] C. Escobar, “¿Cuál es la vacuna contra COVID-19 más aplicada en México?”, Serendipia. Consultado: el 30 de agosto de 2023. [En línea]. Disponible en: [https://serendipia.digital/covid-19/vacuna-covid-mas-aplicada-en-mexico/#google\\_vignette](https://serendipia.digital/covid-19/vacuna-covid-mas-aplicada-en-mexico/#google_vignette)
- [55] R. Treviño, “Pruebas COVID: ¿cuál debo hacerme? ¿cuántas existen?”, CONECTA. Consultado: el 3 de octubre de 2023. [En línea]. Disponible en: <https://conecta.tec.mx/es/noticias/nacional/salud/pruebas-covid-cual-debo-hacerme-cuantas-existen>
- [56] CDC, “Pruebas de detección del COVID-19: información importante”. Consultado: el 3 de octubre de 2023. [En línea]. Disponible en: <https://espanol.cdc.gov/coronavirus/2019-ncov/symptoms-testing/testing.html#:~:text=Pruebas%20de%20ant%C3%ADgeno&text=Los%20resultados%20positivos%20son%20muy,especialmente%20cuando%20no%20hay%20s%C3%ADntomas.>
- [57] BBC News Mundo, “Ómicron: cuáles son los síntomas hasta ahora conocidos de la nueva variante de covid”. Consultado: el 3 de octubre de 2023. [En línea]. Disponible en: <https://www.bbc.com/mundo/noticias-59671740>
- [58] Boletines UAM, “INVESTIGADORES CREAN VENTILADOR INVASIVO ÚTIL PARA ATENDER A PACIENTES CON COVID-19”, UAM. Consultado: el 4 de octubre de 2023. [En línea]. Disponible en: <https://www.comunicacionsocial.uam.mx/boletinesuam/241-20.html#:~:text=En%20relaci%C3%B3n%20con%20el%20d%C3%A9ficit,necesarios%20otros%20diez%20mil%2C%20como>
- [59] V. Sancha-Escudero, J. L. Cobo-Sánchez, N. Mancebo-Salas, R. Pelayo-Alonso, y Z. Gancedo-Gonzalez, “Vacuna contra el SARS-CoV-2 (COVID-19) y enfermedad renal crónica”, *Enfermería Nefrológica*, vol. 24, núm. 2, jun. 2021, doi: 10.37551/S2254-28842021012.
- [60] P. Bergman *et al.*, “Safety and efficacy of the mRNA BNT162b2 vaccine against SARS-CoV-2 in five groups of immunocompromised patients and healthy controls in a prospective open-label clinical trial”, *EBioMedicine*, vol. 74, dic. 2021, doi: 10.1016/j.ebiom.2021.103705.
- [61] J. S. Halsell, “Myopericarditis Following Smallpox Vaccination Among Vaccinia-Naive US Military Personnel”, *JAMA*, vol. 289, núm. 24, jun. 2003, doi: 10.1001/jama.289.24.3283.
- [62] P. Mody, B. Bikdeli, Y. Wang, M. Imazio, y H. M. Krumholz, “Trends in acute pericarditis hospitalizations and outcomes among the elderly in the USA, 1999–2012”, *Eur Heart J Qual Care Clin Outcomes*, vol. 4, núm. 2, abr. 2018, doi: 10.1093/ehjqcco/qcx040.
- [63] M. Wallace y S. Oliver, “COVID-19 mRNA vaccines in adolescents and young adults: Benefit-risk discussion”, CDC, 2021. Consultado: el 10 de septiembre de 2023. [En línea]. Disponible en: <https://www.cdc.gov/vaccines/acip/meetings/downloads/slides-2021-06/05-COVID-Wallace-508.pdf>

- [64] M. G. Thompson *et al.*, “Effectiveness of Covid-19 Vaccines in Ambulatory and Inpatient Care Settings”, *New England Journal of Medicine*, vol. 385, núm. 15, oct. 2021, doi: 10.1056/NEJMoa2110362.
- [65] J. J. G. Marco, M. J. Á. Pasquín, y S. M. Martín, “Efectividad y seguridad de las vacunas para el SARS-CoV-2 actualmente disponibles”, *FMC*, vol. 28, núm. 8, oct. 2021, doi: 10.1016/j.fmc.2021.07.001.
- [66] Dr. Aseem Malhotra, “Dr Malhotra on calling for vaccine data transparency: ‘The information has evolved considerably’”, GBNews, Reino Unido, el 11 de agosto de 2022. Consultado: el 5 de octubre de 2023. [En línea]. Disponible en: <https://www.youtube.com/watch?v=DWYdDKKTXc>
- [67] Dr. Alejandro Díaz Villalobos, “Conferencia magistral ‘Pandemia y vacunas, lecciones aprendidas’, del 28 de febrero de 2023”, Senado de México, México, el 28 de febrero de 2023. Consultado: el 5 de octubre de 2023. [En línea]. Disponible en: <https://www.youtube.com/watch?v=3NZSb02B53I>
- [68] CDC, “Clasificaciones y definiciones de las variantes del SARS-CoV-2”, <https://espanol.cdc.gov/coronavirus/2019-ncov/variants/variant-classifications.html>.
- [69] COVID-19 Prevention Network, “Vacunas contra la COVID-19:¿Cómo las conseguimos tan rápido?”, <https://espanol.coronaviruspreventionnetwork.org/rapidez-de-desarrollo-de-vacunas-covid19>.
- [70] IBM, “Árboles de decisión”, <https://www.ibm.com/es-es/topics/decision-trees>.

## Recursos de plataformas

- EMA. (s.f.). *EudraVigilance*. Obtenido de <https://www.ema.europa.eu/en/human-regulatory/research-development/pharmacovigilance/eudravigilance>
- Epidemiología, D. G. (s.f.). *Datos Abiertos Dirección General de Epidemiología*. Obtenido de Gobierno de México: <https://www.gob.mx/salud/documentos/datos-abiertos-152127>
- Mathieu E., R. H.-G.-O. (s.f.). *COVID-19 Data*. Obtenido de Our World In Data: <https://github.com/owid/covid-19-data/tree/master/public/data>
- Mathieu, E., Ritchie, H., Ortiz-Ospina, E. *et al.* A global database of COVID-19 vaccinations. *Nat Hum Behav* (2021). <https://doi.org/10.1038/s41562-021-01122-8>
- Hasell, J., Mathieu, E., Beltekian, D. *et al.* A cross-country database of COVID-19 testing. *Sci Data* **7**, 345 (2020). <https://doi.org/10.1038/s41597-020-00688-8>
- OMS. (s.f.). *VigiAccess*. Obtenido de <https://www.vigiaccess.org/>
- VAERS. (s.f.). *VAERS*. Obtenido de <https://vaers.hhs.gov/reporteventspan.html>





**SALUD**  
SECRETARÍA DE SALUD

Oficina del Abogado General  
Unidad de Transparencia



Ciudad de México, a 17 de julio de 2023.

**Estimado (a) Solicitante:**

Me refiero a la solicitud de información con número de folio **330026923002342**, ingresada por Usted a esta Unidad de Transparencia de la Secretaría de Salud, en los términos siguientes:

**Modalidad preferente de entrega de información**

*Entrega por internet en la PNT*

**Descripción clara de la solicitud de información**

*"Solicito los datos abiertos (sin datos personales) relacionados con la vacunación de COVID-19 en la cual se incluya, de ser posible, el número de decesos de personas vacunadas desglosadas por tipo de vacuna, edad y sexo; además, el número de reportes adversos de cada vacuna desglosado por tipo de vacuna, edad, sexo y evento adverso. En caso de no estar disponible esta información me gustaría recibir la base de datos solicitada en el reporte con folio: 330026922011725. "(Sic)*

Sobre el particular, con fundamento en lo dispuesto por los artículos 6 de la Constitución Política de los Estados Unidos Mexicanos; 121 y 133 al 136 de la Ley Federal de Transparencia y Acceso a la Información Pública; 39 de la Ley Orgánica de la Administración Pública Federal; en el artículo 24 y 32 BIS 2 del Reglamento Interno de la Secretaría de Salud; le informo que la solicitud fue turnada a la **Dirección General de Información en Salud**; y a la **Dirección General de Epidemiología**; mismas que en el ámbito de su competencia, dieron respuesta en los términos de los oficios que se anexan.

Por otro lado, se le informa que de conformidad con lo dispuesto por el artículo 142 de la Ley General de Transparencia y Acceso a la Información Pública (LGTAIP), así como el 147 de la LFTAIP, usted cuenta con un término de quince días hábiles siguientes a la fecha de notificación de la respuesta a la solicitud de información, para la interposición del recurso correspondiente.

Finalmente, le comento que de tener alguna duda o aclaración respecto de la respuesta que se otorga, puede comunicarse al teléfono 5550621600 extensión 53003# y 53005#; o bien, mediante el correo electrónico: [unidadenlace@salud.gob.mx](mailto:unidadenlace@salud.gob.mx).

Sin otro particular, le envío un cordial y afectuoso saludo.

**Atentamente**  
**Unidad de Transparencia**  
**Secretaría de Salud**

Av, Marina Nacional No. 60, Planta Baja, Col. Tacuba, Alcaldía Miguel Hidalgo, C.P. 11410, Tel: 55-5062-1600 [www.gob.mx/](http://www.gob.mx/)



EYAR



Subsecretaría de Prevención y Promoción de la Salud  
Dirección General de Epidemiología

**Ciudad de México a 11 de julio de 2023**  
**DGE-DG-9616-2023**

Asunto: Respuesta a la solicitud de Información  
Número 330026923002342

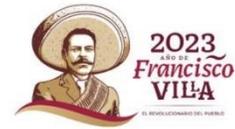
**Unidad de Transparencia de la  
Oficina de la Abogada General  
Marina Nacional número 60, Planta baja,  
Colonia Tacuba, Demarcación Territorial Miguel Hidalgo  
C. P. 11410, Ciudad de México**

Para dar cumplimiento a lo establecido en la Ley Federal de Transparencia y Acceso a la Información Pública y en atención a la solicitud de información número 330026923002342 que a continuación se transcribe:

*Solicito los datos abiertos (sin datos personales) relacionados con la vacunación de COVID-19 en la cual se incluya, de ser posible, el número de decesos de personas vacunadas desglosadas por tipo de vacuna, edad y sexo; además, el número de reportes adversos de cada vacuna desglosado por tipo de vacuna, edad, sexo y evento adverso. En caso de no estar disponible esta información me gustaría recibir la base de datos solicitada en el reporte con folio: 330026922011725..*

Con fundamento en los artículos 6, 13, 15 y 133, de la Ley Federal de Transparencia y Acceso a la Información Pública y en interpretación armónica al artículo 32 BIS 2 del Reglamento Interior de la Secretaría de Salud; me permito informar que la Dirección General de Epidemiología cuenta con el Sistema de Vigilancia Epidemiológica de Enfermedad Respiratoria Viral (SISVER), en el cual se recopila información relacionada con COVID-19, por lo anterior se comparte la información disponible en formato Excel de las defunciones registradas por edad, sexo y tipo de vacuna, reportadas durante el periodo 2021-2023 (corte de información al 26 de junio del 2023).





Subsecretaría de Prevención y Promoción de la Salud  
Dirección General de Epidemiología

Sin más por el momento, quedo de usted.

**Atentamente**

**MSP. Humberto Macías Gamiño**  
**Jefe de Departamento de Apoyo Gerencial**  
**y Enlace de Transparencia**

C.c.p: Dr. Hugo López-Gatell Ramírez. -Subsecretario de Prevención y Promoción de la Salud. – Para su superior conocimiento.  
Dr. Gabriel García Rodríguez. - Director General de Epidemiología. Para su conocimiento.  
Dr. Christian Arturo Zaragoza Jiménez. - Director de Información Epidemiológica. Para su conocimiento.  
Dra. Santa Elizabeth Ceballos Liceaga. Directora de Vigilancia Epidemiológica de Enfermedades Transmisibles. Para su conocimiento.

Sección/Serie: 12C.6





Secretaría de Salud  
Subsecretaría de Prevención y Promoción de la Salud  
Dirección General de Información en Salud  
Dirección de Coordinación, Normatividad y Difusión  
Subdirección de Integración y Difusión de la Información

Ciudad de México, a 04 de julio del 2023

**Oficio No. DGIS-DCND-SIDI-481-2023**

**Asunto:** Respuesta a la Solicitud de información 330026923002342

**Lic. Daniel Luna Alamilla**  
**Director de Área**  
**Oficina del Abogado General**  
**Avenida Marina Nacional No. 60, Colonia**  
**Tacuba, Delegación Miguel Hidalgo C. P.**  
**11410, Ciudad de México**

**P r e s e n t e**

Con referencia al oficio número OAG-UT-2887-2022 enviado por correo electrónico, a través del cual señala haber recibido en el Sistema Integral de Solicitudes de Información (INFOMEX), la solicitud de acceso a la información pública con número de folio 330026923002342, que a la letra señala:

“Solicito los datos abiertos (sin datos personales) relacionados con la vacunación de COVID-19 en la cual se incluya, de ser posible, el número de decesos de personas vacunadas desglosadas por tipo de vacuna, edad y sexo; además, el número de reportes adversos de cada vacuna desglosado por tipo de vacuna, edad, sexo y evento adverso. En caso de no estar disponible esta información me gustaría recibir la base de datos solicitada en el reporte con folio: 330026922011725.”

Otros datos para su localización:

“Vacunas de COVID-19.” [sic]

Motivo por el cual solicita a esta Dirección General de Información en Salud (DGIS) en su ámbito de competencia proporcione la información relacionada con dicha petición.

Al respecto, con fundamento en los artículos 6 inciso A), 80 y 90 de la Constitución Política de los Estados Unidos Mexicanos; 26 y 39 de la Ley Orgánica de la Administración Pública Federal; 130 y 132 de la Ley Federal de Transparencia y Acceso a la Información Pública; 2 y 24 del Reglamento Interior de la Secretaría de Salud y en mi calidad de





Secretaría de Salud  
Subsecretaría de Prevención y Promoción de la Salud  
Dirección General de Información en Salud  
Dirección de Coordinación, Normatividad y Difusión  
Subdirección de Integración y Difusión de la Información

Enlace de Transparencia se comunica lo siguiente:

Por lo que hace a "Solicito los datos abiertos (sin datos personales) relacionados con la vacunación de COVID-19 en la cual se incluya, de ser posible, el número de decesos de personas vacunadas desglosadas por tipo de vacuna, edad y sexo;" se comunica que si bien es cierto la DGIS cuenta con información de defunciones generales "INEGI/SS", la cual es una réplica de la emitida por el Instituto Nacional de Estadística y Geografía (INEGI) ya que de conformidad con el artículo 6, 46 y 78 de la Ley del Sistema Nacional de Información Estadística y Geográfica y el numeral Segundo del Acuerdo por el que se determina Información de Interés Nacional al Proyecto de generación de estadística básica denominado Estadísticas de Defunciones Registradas, es la instancia encargada de emitir la información oficial de defunciones, dicha base únicamente cuenta con información de la causa básica de la defunción, en ese sentido, se aclara que la afección que refiere es considerada como una causa externa a una defunción, motivo por el cual no se cuenta con registros bajo dicha afección. Motivo por el cual se invoca el criterio SO/018/2013 emitido por el Instituto Nacional de Transparencia y Acceso a la Información y Protección de Datos Personales (INAI) que a la letra señala:

"Respuesta igual a cero. No es necesario declarar formalmente la inexistencia. En los casos en que se requiere un dato estadístico o numérico, y el resultado de la búsqueda de la información sea cero, éste deberá entenderse como un dato que constituye un elemento numérico que atiende la solicitud, y no como la inexistencia de la información solicitada. Por lo anterior, en términos del artículo 42 de la Ley Federal de Transparencia y Acceso a la Información Pública Gubernamental, el número cero es una respuesta válida cuando se solicita información cuantitativa, en virtud de que se trata de un valor en sí mismo.

En cuanto a "el número de reportes adversos de cada vacuna desglosado por tipo de vacuna, edad, sexo y evento adverso En caso de no estar disponible esta información me gustaría recibir la base de datos solicitada en el reporte con folio: 330026922011725". Se comunica que dentro de los sistemas de información que administra esta Dirección General, no captan y/o general información sobre "reportes adversos", "base de datos solicitada en el reporte con folio: 330026922011725", motivo por el cual no es posible proporcionar la información requerida.

Lo anterior, con la finalidad de que esa Unidad de Transparencia pueda dar oportuna respuesta a la solicitud en comento.

Sin otro particular por el momento, aprovecho la ocasión para enviarle un cordial saludo.





Secretaría de Salud  
Subsecretaría de Prevención y Promoción de la Salud  
Dirección General de Información en Salud  
Dirección de Coordinación, Normatividad y Difusión  
Subdirección de Integración y Difusión de la Información

**Atentamente**  
**Subdirectora de Integración y Difusión de la Información**

**Lic. Marian Selene Camacho Muñoz**

**Clasificación:** 12 C / 12C.6  
**C.c.p.:** Erick Yussel Álvarez Ruiz - Soporte Administrativo C  
MSCM/

Firmante: MARIAN SELENE CAMACHO MUÑOZ  
Fecha de la firma: 2023/07/04 18:22:38 -0500  
SHA: 7b1b193defb2a4e559c704bf379cb9d760a0288b4094702b128f33be37502dc6  
Homero 213, piso 11, Col. Chapultepec Morales.  
Tel: (55) 5062 1600 [www.gob.mx/salud](http://www.gob.mx/salud)

