UNIVERSIDAD PERUANA
CAYETANO HEREDIA

“MODELOS DE PREDICCIÓN PARA LA LETALIDAD POR COVID-19: ANÁLISIS SECUNDARIO DEL SISTEMA NACIONAL DE VIGILANCIA EPIDEMIOLÓGICA DEL MINISTERIO DE SALUD DE PERÚ”

TESIS PARA OPTAR EL GRADO DE MAESTRO EN CIENCIAS EN INVESTIGACIÓN EPIDEMIOLÓGICA

WENDY CAROLINA NIETO GUTIERREZ

LIMA – PERÚ

2023

ASESOR

Cesar Ugarte-Gil; MD, MSc, PhD

CO-ASESOR

Steev Loyola; TM, MSc

JURADO DE TESIS

DR. CESAR PAUL EUGENIO CARCAMO CAVAGNARO

PRESIDENTE

MG. GABRIEL CARRASCO ESCOBAR

VOCAL

MG. DIEGO ALEJANDRO FANO SIZGORICH

SECRETARIO

DEDICATORIA.

A mi familia que me enseñaron la perseverancia y me brindan su apoyo constante. A todas las personas que me brindaron las oportunidades para aprender sobre la investigación desde pre-grado, y que siempre han tenido una vocación de enseñanza hacia mi persona.

AGRADECIMIENTOS.

A Jaid Campos-Chambergo y Steev Loyola que me brindaron el acceso a la base de datos y apoyaron tanto en el análisis como revisión del manuscrito final

FUENTES DE FINANCIAMIENTO.

Autofinanciada

MODELOS DE PREDICCIÓN PARA LA LETALIDAD POR COVID-19: ANÁLISIS SECUNDARIO DEL SISTEMA NACIONAL DE VIGILANCIA EPIDEMIOLÓGICA DEL MINISTERIO DE SALUD DE PERÚ

INFORME DE ORIGINALIDAD

20%	18%	15%	14%
INDICE DE SIMILITUD	FUENTES DE INTERNET	PUBLICACIONES	TRABAJOS DEL ESTUDIANTE

FUENTES PRIMARIAS

1	www.dge.gob.pe Fuente de Internet	<1%
2	ouci.dntb.gov.ua Fuente de Internet	<1%
3	repositorio.upch.edu.pe Fuente de Internet	<1%
4	Submitted to Universidad Autónoma de Bucaramanga,UNAB Trabajo del estudiante	<1%
5	Submitted to Monash University Trabajo del estudiante	<1%
6	www.wjgnet.com Fuente de Internet	<1%
7	Submitted to Queen's University of Belfast Trabajo del estudiante	<1%
	qmro.qmul.ac.uk	

TABLA DE CONTENIDOS

RESUMEN

Objetivo: Constuir modelos y evaluar su performance para la predicción de letalidad por COVID-19 considerando datos poblacionales del sistema de vigilancia de la enfermedad en el Perú. **Metodología:** Estudio de tipo cohorte retrospectiva. La población de estudio estuvo conformada por los casos registrados y confirmados de COVID-19 en el sistema de vigilancia de nueve provincias de Lima. La población total fue dividida en una muestra aleatoria de 80%, en donde se realizó la construcción de modelos utilizando estrategias diferentes para seleccionar variables (1: estudios previos; 2: método de Lasso; 3: significancia estadística ; 4: post-hoc). En el 20% restante se realizó la validación interna. La selección de los modelos finales se basó en la comparación del performance obtenido y la coherencia biológica de las asociaciones. **Resultados:** La información de 17 678 casos fue utilizada para la creación de modelos y 4 420 para la validación. Los modelos construidos tuvieron performances comparables; sin embargo, se decidió seleccionar el modelo 1 (13 ítems), debido a su menor cantidad de variables y ligero mayor performance, y el modelo resumido obtenido de la estrategia 4 (3 ítems). Los performance de ambos modelos fueron consistentes cuando se realizó la validación, inclusive, cuando se evaluó en diferentes escenarios. **Conclusión:** Los modelos construidos tuvieron performances comparables; sin embargo, recomendamos dos modelos de predicción, cuyas áreas bajo la curva fueron óptimas y mostraron superioridad debido a su plausibilidad biológica y al menor número de variables incluidas. Futuros estudios deberán corroborar el performance y validar la utilidad en contextos acutales.

Palabras claves: COVID-19;SARS-CoV-2;mortalidad;pronóstico (fuente: DeCs)

ABSTRACT

Aim: To develop and evaluate models for predicting COVID-19 fatality, we considered population data from the disease surveillance system in Peru. **Methods:** Retrospective cohort study. The study population consisted of registered and confirmed COVID-19 cases in the surveillance system of nine provinces in Lima. The total population was divided into an 80% random sample, where models were constructed using different strategies for variable selection (1: previous studies; 2: Lasso method; 3: statistical significance; 4: post-hoc). The remaining 20% underwent internal validation. Selection of final models was based on comparing the achieved performance and biological coherence of associations. **Results:** Information from 17,678 cases was used for model creation, and 4,420 for validation. Constructed models had comparable performances; however, model 1 (13 items) was selected due to its lower number of variables and slightly higher performance, along with the summary model obtained from strategy 4 (3 items). The performances of both models remained consistent during validation, even across different scenarios. **Conclusion:** The constructed models exhibited comparable performances; nevertheless, we recommend two prediction models with optimal area under the curve values that demonstrated superiority due to their biological plausibility and fewer included variables. Future studies should corroborate the performance and validate utility in current contexts.

Keywords: COVID-19;SARS-CoV-2;mortality;prognosis (source: MeSH)

I. INTRODUCCION

Planteamiento del problema

La pandemia por COVID-19 ha generado una crisis sanitaria, económica, y humanitaria a nivel global, ocasionando hasta julio del 2021 más de 100 millones de casos y más de 3.9 millones de muertes (1). La muerte por COVID-19 fue estimada como menor al 3% durante los primeros meses de pandemia (2); no obstante, estimaciones posteriores reportaron cifras entre el 0.1 al 5.0% (3), las cuales variaron entre continentes y países . Los cambios en la letalidad se han atribuido a las características propias de cada localidad; específicamente, la temperatura, epidemias concomitantes de otras infecciones respiratorias (tuberculosis, etc.) patrones culturales y económicos, y la disponibilidad de recursos médicos, entre otros (4).

Frente a los altos índices de letalidad se han realizado múltiples iniciativas por parte de la Organización Mundial de Salud (5) para promover la creación de modelos predictivos de muerte por COVID-19, con el objetivo de facilitar a los tomadores de decisiones la realización de estrategias de decisión, planeamiento, y formulación de políticas públicas para la priorización de los grupos más vulnerables. Diferentes estudios han utilizado datos poblacionales para evaluar si las características clínicas, sociodemográficas, y de laboratorio influyen en la ocurrencia de desenlaces negativos en casos con COVID-19 (6-8). Sin embargo, es necesario validar dichos modelos en contextos diferentes al que fue desarrollado debido a las diferencias de las tasa de letalidad y la distribución de las características relacionadas con este desenlace (9).

Perú ha experimentado una variable y dinámica incidencia de COVID-19 (10), la cual muy probablemente haya estado influenciada por los diferentes cuadros de presentación de la enfermedad, protocolos de prevención y tratamiento, distribución de recursos sanitarios, y patrones socio-culturales específicos. Estos factores han sido marcadamente diferentes a los observados en países de altos ingresos (11), e incluso entre países de la misma región (12).

A nivel global, Perú se posicionó como el país con el mayor número de muertes por cada 100.000 habitantes durante la primera y segunda ola (13), y, hasta julio del 2023, con la mayor letalidad por COVID-19 (3). Si bien algunos estudios han reportado factores asociados a muerte en la población peruana (14-17), estos presentan limitaciones importantes como; bajo tamaño de muestra, falta de temporalidad, y sesgo de selección, por lo que sus resultados podrían ser poco precisos y representativos. En ese sentido, evaluar los factores predictivos de muerte en casos con COVID-19 sigue siendo necesario en contextos poco estudiados y para poblaciones no antes caracterizadas. El presente estudio tuvo como objetivo construir múltiples modelos predictivos de muerte entre casos con COVID-19 atendidos en nueve provincias urbano-rurales utilizando datos poblacionales del sistema de vigilancia del Perú.

Justificación del estudio

Desde el inicio de la pandemia por COVID-19, se han impulsado diferentes estrategias para frenar la trasmisión de los casos, que a su vez han ido de la mano con la instauración de estrategias para la priorización de grupos vulnerables, con el fin de prever desenlaces fatídicos de la enfermedad y evitar el colapso del sistema de salud (18, 19).

En respuesta a la pandemia, la Organización Mundial de Salud promovió la creación de modelos de predicción para ser instaurados dentro de las estrategias de priorización de grupos vulnerables a la enfermedad, permitiendo así la decisión, planeamiento, y formulación de políticas públicas en todo el mundo (5). Si bien, diferentes estudios han realizado modelamientos de desenlaces fatídicos para la enfermedad de COVID-19 (20), muchos de estos han reportado falencias metodológicas, principalmente correspondiente a la selección de la muestra, y, sólo algunos, utilizan datos poblacionales (6-8). Lamentablemente, estos modelos, en la mayoría de casos, incluyen variables de laboratorio que no suelen encontrarse disponibles en Perú, más aún en provincias, y que su obtención podría retrasar la priorización de los pacientes por falta de disponibilidad y escases de recursos (21). Además, para la extrapolación de cualquier modelo predictivo es necesario realizar validaciones del performance en contextos diferentes (9). Más aún, porque se hipotetiza que estos modelos podrían no ser útiles en países en vías de desarrollo, como es el caso de Perú, debido a diferencias en patrones socio-culturales (que acarrean mayores tendencias de aglomeración y exposición viral) (22), económicos (que acarrean desigualdades en el acceso sanitario) (12), geográficos (factores climáticos y niveles de altitud) (23, 24), distribución de comorbilidades (25), y factores genéticos que podrían modificar la carga de la enfermedad y sus tendencias (26).

A pesar que se han realizado algunos estudios de modelos de predicción para la letalidad de COVID-19 en Perú (14-17), estos presentan limitaciones que afectarían importantemente sus estimaciones, siendo necesario construir nuevos

modelos con datos poblacionales peruanas o inclusive realizar validaciones externas de modelos previamente construidos. Más aún en provincias, donde la diferencia es aún más marcada frente a las poblaciones de estudios previos y a la de la capital del Perú (27).

Perú es un país centralizado, motivo por el cual es poca la información, investigación, e implementación de estrategias públicas en provincias, en comparación a Lima metropolitana, la cual es la capital (28). Estas diferencias podrían conllevar a diferenciales tendencias y patrones de predisposición a letalidad, dado que es ampliamente conocido que las provincias; tienen un escaso acceso a servicios de salud de alto nivel, enfrentan diversas barreras geográficas que limitan la movilización a la capital, enfrentan una baja disponibilidad de recurso sanitario y por ende menores inversiones sanitarias (28), tuvieron una mejor ejecución de planes y estrategias de contención de la enfermedad (29), y cuentan con una mayor vulnerabilidad social (30) que los predispondría a una mayor exposición frente al virus. Inclusive estas diferencias se ve en el mismo departamento de la capital del Perú, pues se han observado diferencias entre las provincias de Lima y Lima metropolitana, donde las provincias tiene patrones similares a otros departamentos del país durante la pandemia (29, 30), y que su evaluación podría representar una población particular poco estudiada.

II. OBJETIVOS

Objetivo general

Construir múltiples modelos y evaluar su performance para la predicción de letalidad por COVID-19 en pacientes atendidos en nueve provincias urbano-rurales utilizando datos poblacionales del sistema de vigilancia del Perú.

Objetivos específicos

- Estimar las curvas de sobrevida de la enfermedad en los pacientes atendidos en las provincias de Lima-Perú.
- Determinar las características sociodemográficas predictivas de letalidad en pacientes con COVID-19 en provincias de Lima-Perú.
- Determinar las características clínicas predictivas de letalidad en pacientes con COVID-19 en provincias de Lima-Perú
- Estimar los puntos de corte de probabilidad de los modelos construidos para predecir la letalidad en pacientes con COVID-19 en provincias de Lima-Perú
- Validar los modelos construidos para predecir la letalidad en pacientes con COVID-19 en provincias de Lima-Perú

III. MARCO TEÓRICO

Antecedentes del estudio

Yadaw et. al, en el año 2020, realizó un estudio con el objetivo de desarrollar un modelo de predicción de la letalidad por COVID-19 mediante métodos computacionales imparciales, para identificar las características clínicas predictivas de muerte. Para ello se aplicaron técnicas de aprendizaje automático a los datos clínicos de una gran cohorte de pacientes con COVID-19 tratados en el Mount Sinai Health System en la ciudad de Nueva York. La población estuvo compuesta por los pacientes con diagnóstico confirmado de COVID-19 que asistieron a algún centro de salud de la ciudad entre el 9 de marzo y el 6 de abril del 2020. En este caso, se diseñaron modelos de predicción basados en las características clínicas y las características del paciente, en los cuales se evaluó el performance mediante el área bajo la curva (AUC). Se desarrolló un modelo de predicción para la letalidad por COVID-19 que mostró una alta precisión (AUC=0.91), el cual fue consistente en diferentes muestras de validación que incluyó variables como la edad del paciente, la saturación mínima de oxígeno en el transcurso de su encuentro médico y el tipo de encuentro con el paciente. Se concluyó que el modelo era preciso y parsimonioso, pero que debían realizarse validaciones externas en otras poblaciones (7).

En el año 2021, Banoei et. al realizó un estudio con el objetivo de identificar las variables predictivas para la letalidad por COVID-19. Para lo cual, se evaluó entre 108 características clínicas, comorbilidades y marcadores sanguíneos en una población de pacientes con diagnóstico de COVID-19 atendidos un hospital de

Florida. Se aplicó una modificación inspirada en un modelo basado en el mínimo cuadrado parcial para predecir la muerte hospitalaria. Asimismo, se aplicó un análisis de clases latentes para agrupar a los pacientes con COVID-19 en bajo o alto riesgo. El modelo construido tuvo un poder predictivo moderado (Q2 = 0.24) y alta precisión (AUC > 0.85) tanto en los datos de entrenamiento y validación. El modelo final estuvo constituido por 18 predictores clínicos y de comorbilidades y 3 marcadores bioquímicos. Se concluyó que el modelo construido era preciso para predecir muerte por COVID-19 entre los pacientes hospitalizados utilizando datos clínicas y comorbilidades, pudiendo desempeñar un papel beneficioso en el entorno clínico y consecuentemente un mejor manejo de los pacientes con COVID-19 (31).

En el año 2022, Akama et. al. realizó un estudio con el fin de identificar las variables asociadas a los resultados graves en pacientes con COVID-19. Para lo cual realizó un estudio retrospectivo que incluyó pacientes con COVID-19 que asistieron a un centro de salud entre el 10 de marzo de 2020 y el 13 de octubre de 2020. Se encontró que, del total de la población 149 fallecieron a causa de la enfermedad. Así mismo, se identificó que la frecuencia respiratoria se relación directamente proporcional a la gravedad de la enfermedad (Rho de Spearman = -0.56). Utilizando un método de inteligencia artificial se identificó que las características demográficas a de los pacientes, los resultados de laboratorio, los medicamentos, las comorbilidades, los signos y síntomas y los signos vitales eran capaces de predecir la letalidad de los pacientes con COVID-19, con un buen performance (AUC = 0.82). Se concluyó que era necesario realizar validaciones del modelo para corroborar lo reportado (8).

Jong et. al, en el 2022, realizó una revisión sistemática sobre los modelos de predicción para la letalidad de COVID-19 a corto plazo y describió las validaciones externas realizadas para estos modelos. En este caso se incluyó un total de 46 914 pacientes de 18 países, ingresados a un hospital con diagnóstico confirmado de COVID-19 entre noviembre de 2019 a abril de 2021. Sólo ocho modelos fueron validados con diversos predictores. Se realizó un metaanálisis usando información individual de cada participante, y se consideraron múltiples análisis; estadística de concordancia estimada del modelo, la pendiente de calibración, la calibración en general y el ratio de datos observados y esperados (O:E). Las estimaciones agrupadas oscilaron entre 0.67 y 0.80 (estadística C), 0.22 y 1.22 (pendiente de calibración) y 0.18 a 2.59 (relación O:E). La puntuación de letalidad 4C de Knight et al (estadística C: 0.80; IC95%: 0.75 - 0.84) y el modelo clínico de Wang et al (estadística C: 0.77; IC95% 0.73 - 0.80), fueron las que tuvieron mayor capacidad discriminativa para el desenlace (20).

Bases teóricas

Virus SARS-CoV-2

El SARS-COV-2 es un virus de ARN de cadena positiva de la familia Coronaviridae causante de la enfermedad del COVID-19 en humanos. Este virus fue identificado por primera vez en el año 2019 en Wuhan, China. El análisis del genoma viral muestra que el SARS-CoV-2 se encuentra relacionado con el coronavirus SARS (SARSr-CoV) encontrado en los murciélagos, por lo que se encuentra categorizado en el subgénero Sarbecovirus del género Betacoronavirus. Frente a esto se conoce que el huésped natural de este virus es potencialmente el

murciélago chino de herradura (*Rhinolophus affinis*), por lo cual es considerado un agente causante de zoonosis (32).

El SARS-COV-2 comparte un 79 % de la secuencia del genoma del SARS-CoV y el 50 % con el MERS-CoV (33). Se conoce que la mayoría de las proteínas codificadas por este virus tienen una morfología y longitud similar a las proteínas correspondientes a su antecesor, el SARS-CoV, así mismo, entre ellos se comparte más del 90% de los genes estructurales, con excepción el gen S (34), y el 85% de las proteínas no estructurales (35).

Enfermedad por coronavirus-2019 (COVID-19)

La enfermedad por coronavirus-2019 o COVID-19 se caracteriza por presentar síntomas respiratorios y poco específicos, siendo los más frecuentes la fiebre (78%; IC95%: 75 – 81), tos (57%; IC95%: 54 – 60), y la fatiga (31%; IC95%: 27 – 35) (28). Revisiones sistemáticas han estimado que aproximadamente el 23% de los pacientes con COVID-19 desarrollaran una presentación grave de la enfermedad (36) y el 7% fallece a causa de la enfermedad (37). Así mismo, se ha estimado que de los pacientes hospitalizados, principalmente por neumonía grave por COVID-19, el 19% requieren ventilación no invasiva, el 17% cuidados intensivos, y el 9% ventilación invasiva (36). Sin embargo, estas prevalencias pueden variar según la zona geográfica.

Estudios han reportado que la patogénesis de la enfermedad se basa en cuatro posibles puntos de entrada: 1) los receptores ACE2, 2) la mediación de Furina, 3) mediación con GRP78, y 4) mediación con CD147 (36). Para el primer punto de entrada, se ha reportado que existe una unión de las proteínas espiga del coronavirus

que conlleva a la reducción de la expresión de ACE2 en las células, y consecuentemente una mayor conversión de la angiotensina II en el heptapéptido vasodilatador angiotensina, lo que contribuye al desarrollo de hipertensión y a una lesión pulmonar grave. Por otro lado, el SARS-CoV-2 puede reconocer y unirse al GRP78 SBDβ, facilitando la entrada viral, así como, con el CD147, el cual se expresa altamente en los tejidos tumorales, los tejidos inflamados y las células infectadas por patógenos. Por último, se conoce que la furina se expresa principalmente en las células del tracto respiratorio lo que promueve la entrada del SARS-CoV-2 y su consecuente colonización (36).

El diagnóstico de la enfermedad se basa en la sospecha clínica. En este caso, se debe considerar la posibilidad de COVID-19 en cualquier persona con fiebre de nueva aparición y/o síntomas respiratorios (38). Si bien no hay características clínicas específicas que puedan distinguir de forma fiable la COVID-19 de otras infecciones respiratorias virales, se han reportado algunas características que podrían proveer cierto nivel de certeza, como la pérdida del gusto y la pérdida del olfato; sin embargo, ninguno de estos establece definitivamente el diagnóstico de COVID-19 (39).

De manera general, el diagnóstico laboratorial debería ser realizado en todo paciente sintomático con sospecha de COVID-19. Sin embargo, en contextos de recursos limitados, se ha sugerido que los departamentos de salud locales pueden tener criterios específicos para la priorización de casos, como pacientes hospitalizados (especialmente pacientes críticos con enfermedades respiratorias inexplicables), personal que trabaja en el sector de salud, entre otros (40).

Existen diferentes técnicas diagnósticas: 1) la detección del ARN viral con pruebas de amplificación de ácidos nucleicos, como la reacción en cadena de la polimerasa con transcripción reversa en tiempo real (rRT-PCR); o 2) la detección de antígenos virales con ensayos de flujo lateral o también llamados pruebas de diagnóstico rápido (Ag-RDT). Sin embargo, se ha recomendado que el diagnóstico confirmatorio se realice en base a la detección directa del ARN del virus mediante pruebas de amplificación de ácido nucleico (41).

Pandemia por COVID-19

Según el repositorio de datos de COVID-19 del Centro de Ciencia e Ingeniería de Sistemas de la Universidad Johns Hopkins, hasta el 15 de diciembre del 2022, un total de 652 280 628 casos con diagnóstico de COVID-19 han sido reportados, de los cuales han fallecido un total de 6 663 373 personas a causa de la enfermedad (34).

En general, dentro de los países con mayor ratio de muerte en los últimos 28 días y casos totales (hasta 16 de diciembre del 2022), Japón, Corea del Sur, Estados Unidos, Francia, y China son los países que lideran la lista a nivel mundial. Sin embargo, el ratio de letalidad de la COVID-19 es liderado por Corea del Norte, Zaandam, Yemen, Sudan, Siria, Somalia, Perú, Egipto, y México (42).

En los países de Sudamérica se han reportado un total de 66 013 185 casos, siendo Perú quien ocupa el quinto puesto con el mayor número de casos (4 399 073) y muertes (217 782) para una población de 33 684 208 (42).

IV. METODOLOGÍA

Diseño del estudio

Estudio observacional, analítico, tipo cohorte abierta retrospectiva. El presente estudio realizó un análisis secundario de los datos recolectados por el Sistema Nacional de Vigilancia Epidemiológica para COVID-19 (NotiWeb) del Centro Nacional de Epidemiología, Prevención y Control de enfermedades (CDC) del Perú.

Contexto

El Sistema Nacional de Vigilancia Epidemiológica de COVID-19 realiza un proceso pasivo de identificación de casos y el seguimiento de los mismos hasta la resolución de la enfermedad o muerte. La notificación es de carácter obligatorio para toda institución prestadora de servicios de salud. Por lo que, al atenderse un caso sospechoso o probable, según la directiva del Ministerio de Salud del Perú, el individuo es captado por los equipos de respuesta rápida y notificado utilizando la plataforma NotiWeb (43).

La Dirección Regional de Salud de Lima Provincias (DIRESA-LIPRO) administra la información de los casos de COVID-19 reportados por las instituciones prestadoras de servicios de salud pertenecientes a alguno de los 128 distritos ubicados en alguna de las nueve provincias del departamento de Lima-Perú; Barranca, Cajatambo, Canta, Cañete, Huaral, Huarochirí, Huaura, Oyón y Yauyos *(Anexo 1)*. Según el último informe nacional previo al 2020, las provincias de Lima albergan al 3.1% de la población Peruana, la cual corresponde

aproximadamente al 9% de la población total del departamento de Lima, y se caracteriza por tener un 83% de población urbana (44).

Población de estudio

El Sistema Nacional de Vigilancia Epidemiológica para COVID-19 incluye información de individuos notificados como casos sintomáticos sospechosos, probables y confirmados de COVID-19 de todo el Perú (43). Sin embargo, para el presente estudio se incluyó información anonimizada de casos sintomáticos confirmados y notificados por la DIRESA-LIPRO. Los casos analizados fueron registrados entre el 01 de marzo del 2020 y el 30 de setiembre del 2020, periodo en el cual la vacunación contra COVID-19 no había sido implementada en el país. La definición de caso confirmado, utilizada en este estudio, estuvo basada en los lineamientos estipulados por el Ministerio Nacional de Salud del Perú (43). Para este análisis se excluyó los casos no confirmados, con nacionalidad diferente a la peruana, y casos menores de 18 años. Estos últimos fueron excluidos por su vulnerabilidad, y menor y diferente tasa de letalidad en comparación a los adultos (45). En caso un participante haya sido reportado como un caso en más de una oportunidad (identificado mediante el ID único de cada sujeto), se incluyó únicamente la información del primer reporte.

Muestreo y potencia estadística

Los casos registrados en el Sistema Nacional de Vigilancia Epidemiológica peruano son enrolados mediante un muestreo censal, abordando a todos aquellos que se reporten como casos sospechosos o probables de COVID-19. Luego, las muestras biológicas recolectadas son procesadas para la detección del SARS-CoV-2 o de los

anticuerpos producto de la infección. De encontrar resultados positivos en los exámenes de laboratorio, el caso es reclasificado y notificado como confirmado.

Debido a que el presente estudio realizó un análisis secundario de una base de datos de vigilancia epidemiológica, se decidió estimar un tamaño de muestra mínimo tomando en cuenta los cuatro criterios para la estimación de muestras para modelos de predicción descritos por Smeden y Riley (46). Para la estimación del tamaño muestra se utilizó una calculadora interactiva (https://riskcalc.org/pmsamplesize/). Tomando en cuenta un número de 35 parámetros potenciales para su inclusión en el modelo, una letalidad del 11.0% estimada para países de Latinoamérica (47), un valor esperado del R^2 de 0.1, un nivel de shrinkage (una medida de sobreajuste) de 0.9, y una estadística C informada en un estudio de predicción previo de 0.94 (7), el tamaño de muestra obtenido más grande fue de 2972 casos para alcanzar una potencia del 80.0%. Considerando que el tamaño de muestra analizado en este estudio supera el tamaño muestral más grande obtenido, se concluyó que nuestro poder estadístico es adecuado para la construcción del modelo de predicción.

Operacionalización de variables

El desenlace fue definido como la muerte de un caso confirmado cuya causa registrada fue la de COVID-19 o alguna complicación relacionada (ej. insuficiencia respiratoria, orgánica, etc.). Esta variable fue obtenida de los datos registrados en el sistema de vigilancia hasta enero del 2021, y corroborado de forma individual con el registro del Sistema Nacional de Defunciones (SINADEF) del Perú (48). De esta manera se obtuvo una variable categórica dicotómica de “no” y “sí”.

Se consideraron variables predictoras como edad (años), sexo (femenino y masculino), características del cuadro clínico al momento del registro en el sistema de vigilancia (fiebre, tos, entre otros), severidad de la enfermedad según los síntomas (sin y con síntomas de severidad), comorbilidades (hipertensión, diabetes, entre otras), y número de comorbilidades (ninguna, 1 – 2, y ≥3).

Procedimientos

La recolección de los datos fue realizada utilizando una ficha clínico-epidemiológica y estuvo a cargo del personal de salud de los centros de atención de salud de las provincias del departamento de Lima. Los casos sospechosos de COVID-19 fueron identificados, principalmente, por su atención en algún centro de salud y, minoritariamente, por comunicación a través de la central telefónica (113 Infosalud, 107 EsSalud, etc.), página web, o aplicativo móvil. Luego, los casos sospechosos que cumplieron con la definiciones propuestas por las autoridades de salud fueron registrados en el sistema de vigilancia (43). La DIRESA-LIPRO estuvo a cargo de la supervisión del correcto y completo llenado de las fichas epidemiológicas, así como del análisis diario de la información registrada en el sistema de vigilancia.

Consideraciones éticas

Se solicitó los permisos para el acceso y uso de la base de datos con fines científicos a la DIRESA-LIPRO. Previo al análisis de los datos, el protocolo del estudio fue evaluado y aprobado por el comité de ética institucional de la Universidad Peruana Cayetano Heredia. Por último, el presente estudio fue registrado en la plataforma de proyectos de investigación en salud (código: F8DEB3AE-B12E-4CA5-A8E5-

44623325D244), según lo especificado en el Decreto Supremo N° 014-2020-SA. El análisis fue realizado con datos codificados.

Análisis estadístico

El análisis de los datos fue realizado utilizando Stata v.16 (StataCorp. 2019. Stata Statistical Software: Release 16. College Station, TX: StataCorp LLC.) y Phyton v.3.4.3. En primera instancia, la base de datos fue dividida en dos; una primera muestra aleatoria correspondiente al 80% del total y una muestra restante del 20%. En la primera muestra se realizó la construcción del modelo de predicción, mientras que, en el 20% restante se realizó validación de los modelos construidos.

Se realizó un análisis descriptivo de las características de los participantes en las tres bases de datos (global, dataset para creación de los modelos, y dataset para validación del modelo) utilizando frecuencias absolutas y relativas para las variables categóricas, y medidas de dispersión y de tendencia central para las variables numéricas.

Considerando el desbalance de las categorías del desenlace se decidió ajustar el conjunto de datos para realizar las estimaciones de todos los modelo predictivos (49), mediante la ponderación de las categorías de muerte por COVID-19, asignándole un mayor peso a la categoría minoritaria para aumentar su importancia durante el entrenamiento (50). Dicha ponderación fue introducida a las estimaciones mediante el comando *svy*.

Se realizó regresiones logísticas para la construcción de los modelos considerando que es uno de los métodos más usados y de los más sencillos para interpretar el performance predictivo (51). Se construyeron cuatro modelos de predicción

tomando en cuenta diferentes estrategias de selección de variables: estrategia 1) utilizando variables reportadas como predictivas en modelos de machine learning previamente publicados (7, 8, 31); estrategia 2) según el método de least absolute shrinkage and selection operator [Lasso] *(Anexo 2)*; estrategia 3) según significancia estadística ($p<0.05$) observada en el análisis de regresión logística bivariado; y una estrategia 4) basada en una estimación post-hoc, en la cual se incluyó variables consistentemente incluidas en los modelos creados por las tres primeras estrategias. Se realizó un análisis de sensibilidad de las variables predictores incluidas en cada modelo utilizando regresiones de poisson con varianzas robustas, para evaluar las diferencias entre los estimados reportados como riesgo relativo y odds ratio.

Para los modelo creados con regresión logística se estimó la sensibilidad y especificidad para un punto de corte de probabilidad (seleccionado por obtener el performance más balanceado entre la sensibilidad y especificidad), valores predictivos (positivos y negativos), likelihood (positivo y negativo), y áreas bajo la curva (AUC). Se decidió seleccionar el modelo de predicción final utilizando los siguientes criterios; mejor performance obtenido (sensibilidad, especificidad, y AUC), parsimonia (menor número de variables incluidas), y diferencia clínicamente relevante (diferencias del 10% en el performance).

Para la validación del modelo seleccionado se comparó los AUC y las curvas ROC graficadas en el dataset para la creación de los modelos y en el dataset para la validación. Por último, se realizó un análisis de sensibilidad para la validación del modelo, considerando subgrupos de casos; en aquellos con diagnóstico

confirmatorio de COVID-19 mediante RT-PCR, según periodos de la pandemia (primer periodo: desde enero hasta junio del 2020; segundo periodo: desde julio del 2020 hasta setiembre del 2020; y periodo entre picos de incidencia de COVID-19: desde junio del 2020 hasta el 9 de julio del 2020), para evaluar la robustez y consistencia en diferentes escenarios.

V. RESULTADOS

Un total de 23 742 casos confirmados de COVID-19 con alguna prueba positiva y con fecha de inicio de síntomas hasta el 30 de setiembre del 2020 conformaron la base de datos, de los cuales se excluyeron 1404 por ser menores de 18 años, 169 por no tener nacionalidad peruana, y 71 por ser procedentes de un departamento diferente de Lima *(Anexo 3)*, resultando en un total de 22 098 casos (edad media 45.96 ± 16.82; 53.41% sexo femenino). Todas las características fueron comparables entre la muestra total y los datasets utilizados para la creación y validación de los modelos de predicción *(Tabla 1)*. Asimismo, las funciones de supervivencia fueron comparables entre la base total y las muestras obtenidas *(Anexo 4)*. La base para la creación del modelo de predicción estuvo constituida por una muestra de 17 678 casos (edad media 45.99 ± 16.86; 53.25% sexo femenino), y la base para la validación de modelos por un total de 4 420 casos (edad media 45.85 ± 16.64; 54.05% sexo femenino) *(Tabla 1)*.

Tabla 1. Diferencias en las características de la población entre los dataset de creación y validación del modelo

Variable	Muestra general (N= 22 098)	Dataset para creación de los modelos (N= 17 678)	Dataset para validación del modelo (N= 4 420)
	N (%)	N (%)	N (%)
Edad*	45.96 ± 16.82	45.99 ± 16.86	45.85 ± 16.64
Sexo			
Femenino	11803 (53.41)	9414 (53.25)	2389 (54.05)
Masculino	10295 (46.59)	8264 (46.75)	2031 (45.95)
Dirección fiscal			
Barranca	3525 (15.95)	2849 (16.12)	676 (15.29)
Cajatambo	29 (0.13)	23 (0.13)	6 (0.14)
Canta	259 (1.17)	195 (1.1)	64 (1.45)
Cañete	7215 (32.65)	5775 (32.67)	1440 (32.58)

Huaral	4279 (19.36)	3409 (19.28)	870 (19.68)
Huarochirí	583 (2.64)	464 (2.62)	119 (2.69)
Huara	5526 (25.01)	4427 (25.04)	1099 (24.86)
Oyón	143 (0.65)	111 (0.63)	32 (0.72)
Yauyos	183 (0.83)	142 (0.8)	41 (0.93)
Fuera de Lima provincia**	356 (1.61)	283 (1.60)	73 (1.65)
Tipo de prueba para el diagnóstico de la enfermedad			
RT-PCR	4291 (19.44)	3454 (19.56)	837 (18.95)
Prueba serológica	17756 (80.45)	14183 (80.33)	3573 (80.91)
Prueba antigénica	25 (0.11)	19 (0.11)	6 (0.14)
Establecimiento donde se realizó la notificación			
Ministerio de Salud	19754 (89.39)	15795 (89.35)	3959 (89.57)
Seguro Social de Salud	1315 (5.95)	1061 (6)	254 (5.75)
Otros (privado o sanidades)	1029 (4.66)	822 (4.65)	207 (4.68)
Cuadro clínico			
Fiebre	9238 (41.8)	7368 (41.68)	1870 (42.31)
Malestar general	11356 (51.39)	9111 (51.54)	2245 (50.79)
Tos	14450 (65.39)	11588 (65.55)	2862 (64.75)
Dolor de garganta	13012 (58.88)	10385 (58.75)	2627 (59.43)
Congestión nasal	5898 (26.69)	4714 (26.67)	1184 (26.79)
Sensación de dificultad respiratoria	5321 (24.08)	4305 (24.35)	1016 (22.99)
Diarrea	2690 (12.17)	2146 (12.14)	544 (12.31)
Náuseas y vómitos	1742 (7.88)	1401 (7.93)	341 (7.71)
Cefalea	8676 (39.26)	6884 (38.94)	1792 (40.54)
Confusión	239 (1.08)	182 (1.03)	57 (1.29)
Dolor muscular	4952 (22.41)	3940 (22.29)	1012 (22.9)
Dolor abdominal	591 (2.67)	455 (2.57)	136 (3.08)
Dolo de tórax	2987 (13.52)	2356 (13.33)	631 (14.28)
Dolor articular	648 (2.93)	522 (2.95)	126 (2.85)
Disosmia y disgeusia	982 (4.44)	786 (4.45)	196 (4.43)
Dolor de oído	12 (0.05)	9 (0.05)	3 (0.07)
Severidad según síntomas			
Sin síntomas de severidad	18613 (85.56)	14935 (85.79)	3678 (84.61)
Con síntomas de severidad	3142 (14.44)	2473 (14.21)	669 (15.39)
Comorbilidades			
Enfermedad cardiovascular	1574 (7.12)	1278 (7.23)	296 (6.7)
Hipertensión arterial	396 (1.79)	321 (1.82)	75 (1.7)
Dislipidemia	77 (0.35)	68 (0.38)	9 (0.2)
Diabetes	1133 (5.13)	883 (4.99)	250 (5.66)

Tiroidopatía	163 (0.74)	121 (0.68)	42 (0.95)
Hepatopatía	115 (0.52)	91 (0.51)	24 (0.54)
Enfermedad neurológica	124 (0.56)	100 (0.57)	24 (0.54)
Inmunodeficiencia	35 (0.16)	29 (0.16)	6 (0.14)
Enfermedad renal	104 (0.47)	91 (0.51)	13 (0.29)
Enfermedad pulmonar	242 (1.1)	189 (1.07)	53 (1.2)
Asma	343 (1.55)	275 (1.56)	68 (1.54)
Cáncer	111 (0.5)	89 (0.5)	22 (0.5)
Obesidad	863 (3.91)	691 (3.91)	172 (3.89)
Tuberculosis	145 (0.66)	120 (0.68)	25 (0.57)
Número de comorbilidades			
Sin ninguna comorbilidad	17817 (80.63)	14254 (80.63)	3563 (80.61)
Con una o dos comorbilidades	4129 (18.68)	3303 (18.68)	826 (18.69)
Más de tres comorbilidades	152 (0.69)	121 (0.68)	31 (0.7)
Muerte			
No	20433 (92.47)	16346 (92.47)	4087 (92.47)
Sí	1665 (7.53)	1332 (7.53)	333 (7.53)

**Media ± desviación estándar*
***Reside en Lima provincia, pero su domicilio fiscal registra una dirección fuera de esta*

Inicialmente se construyeron modelos utilizando tres estrategias diferentes, de los cuales, la estrategia 1 (considerando variables incluidas en modelos de predicción previamente descritos) incluyó 13 variables, la estrategia 2 (según el método Lasso) incluyó 9 variables, y la estrategia 3 (según significancia estadística en la regresión logística bivariada) incluyó 22 variables *(Tabla 2)*. Interesantemente, entre los modelos construidos, tres variables estuvieron consistentemente incluidas y fueron significativas; edad, sexo, y dificultad respiratoria *(Tabla 2)*. Al observar este patrón se procedió a realizar la estrategia post-hoc. Al realizar el análisis de sensibilidad de las variables predictoras estimando los riesgo relativos, no identificamos diferencias en las asociaciones estadísticamente significativas encontradas en la regresión logística, pero sí una ligera diferencia en los valores

puntuales y los intervalos de confianza, siendo ligeramente menores en las estimaciones reportadas como riesgo relativo que con odds ratio *(Anexo 5)*.

Tabla 2. Variables predictoras para la letalidad en pacientes con COVID-19 (N= 17 678)

Variables	Análisis bivariado	Construcción de modelos			
		Modelo de la estrategia 1	Modelo de la estrategia 2	Modelo de la estrategia 3	Modelo de la estrategia 4
	OR (IC95%)	OR (IC95%)	OR (IC95%)	OR (IC95%)	OR (IC95%)
Edad	**1.09 (1.09 – 1.10)**	**1.08 (1.08 – 1.09)**	**1.08 (1.08 –1.09)**	**1.08 (1.07 – 1.09)**	**1.08 (1.08 – 1.09)**
Sexo					
Femenino	Ref.	Ref.	Ref.	Ref.	Ref.
Masculino	**2.23 (1.99 – 2.51)**	**1.82 (1.55 – 2.13)**	**1.71 (1.46 – 2.00)**	**1.80 (1.53 – 2.11)**	**1.78 (1.52 – 2.08)**
Cuadro clínico					
Fiebre	**1.43 (1.29 – 1.61)**	1.11 (0.95 – 1.30)	–	1.09 (0.92 – 1.28)	–
Malestar general	**1.71 (1.53 – 1.93)**	–	**1.41 (1.20 – 1.66)**	**1.43 (1.22 – 1.68)**	–
Tos	**2.31 (2.01 – 2.66)**	–	**1.55 (1.29 – 1.88)**	**1.62 (1.34 – 1.96)**	–
Dolor de garganta	**0.78 (0.70 – 0.88)**	–	–	0.96 (0.81 – 1.12)	–
Congestión nasal	**0.73 (0.63 – 0.83)**	–	–	0.91 (0.76 – 1.09)	–
Sensación de dificultad respiratoria	**9.87 (8.71 – 11.18)**	**7.36 (6.32 – 8.57)**	**7.14 (6.12 – 8.33)**	**6.96 (5.96 – 8.14)**	**7.59 (6.53 – 8.81)**
Diarrea	0.93 (0.78 – 1.11)	–	–	–	–
Náuseas y vómitos	1.20 (0.99 – 1.46)	–	–	–	–
Cefalea	**0.57 (0.51 – 0.65)**	–	**0.61 (0.51 – 0.72)**	0.59 (0.50 – 0.70)	–
Confusión	**3.53 (2.48 – 5.04)**	–	–	**2.81 (1.36 – 5.83)**	–
Dolor muscular	1.01 (0.88 – 1.15)	–	–	–	–
Dolor abdominal	1.02 (0.72 – 1.45)	–	–	–	–
Dolor de tórax	1.13 (0.96 – 1.32)	–	–	–	–
Dolor articular	**1.40 (1.05 – 1.87)**	–	–	0.91 (0.63 – 1.31)	–
Disosmia y disgeusia	**0.09 (0.04 – 0.20)**	–	**0.19 (0.79 – 0.45)**	–	–
Dolor de oído	1.53 (0.19 – 12.28)	–	–	–	–
Severidad según síntomas (n=270)					
Sin síntomas de severidad	Ref.		Ref.	Ref.	
Con síntomas de severidad	**1.33 (1.15 – 1.54)**	–	1.14 (0.93 – 1.39)	1.14 (0.93 – 1.40)	–
Comorbilidades					
Enfermedad cardiovascular	**3.63 (3.13 – 4.22)**	1.25 (0.97 – 1.61)	–	0.99 (0.70 – 1.39)	–
Hipertensión arterial	**2.72 (2.03 – 3.65)**	0.89 (0.55 – 1.44)	–	0.72 (0.43 – 1.20)	–
Dislipidemia	0.37 (0.09 – 1.52)	–	–	–	–
Diabetes	**2.54 (2.10 – 3.06)**	**1.54 (1.17 – 2.02)**	**–**	1.15 (0.81 – 1.63)	–

Tiroidopatía	0.87 (0.42 – 1.78)	–	–	–	–
Hepatopatía	**2.64 (1.53 – 4.54)**	–	–	0.90 (0.37 – 2.24)	–
Enfermedad neurológica	**2.91 (1.76 – 4.80)**	–	–	1.21 (0.42 – 3.51)	–
Inmunodeficiencia	0.91 (0.22 – 3.83)	4.45 (0.74 – 26.75)	–	–	–
Enfermedad renal	**4.99 (3.15 – 7.89)**	**3.10 (1.48 – 6.52)**	**–**	**2.93 (1.42 – 6.06)**	–
Enfermedad pulmonar	**2.94 (2.04 – 4.25)**	1.37 (0.78 – 2.39)	–	0.97 (0.53 – 1.78)	–
Asma	0.91 (0.57 – 1.45)	1.14 (0.57 – 2.28)	–	–	–
Cáncer	**2.30 (1.30 – 4.09)**	1.16 (0.52 – 2.57)	–	0.79 (0.35 – 1.81)	–
Obesidad	1.29 (0.99 – 1.67)	**1.93 (1.36 – 2.74)**	**1.88 (1.32 – 2.68)**	–	–
Tuberculosis	0.99 (0.50 – 1.97)	–	–	–	–
Número de comorbilidades					
Sin ninguna comorbilidad	Ref.			Ref.	
Con una o dos comorbilidades	**2.60 (2.31 – 2.94)**	–	–	**1.51 (1.12 – 2.03)**	–
Más de tres comorbilidades	**5.03 (3.30 – 7.69)**	–	–	2.09 (0.84 – 5.20)	–

***Modelo de la estrategia 1:** variables incluidas considerando aquellas previamente reportadas como predictores de muerte en otros estudios (7, 8, 27); **Modelo de la estrategia 2:** variables incluidas considerando selección mediante método de Lasso; **Modelo de la estrategia 3**: variables incluidas según significancia estadística en análisis bivariado; **Modelo de la estrategia 4:** modelo creado post-hoc con variables que fueron consistentes en los tres modelos creados previamente.*

Cuando se evaluó el performance de los modelos construidos se identificó que para un punto de corte de probabilidad de muerte de 52% se obtuvo estimaciones similares, con una ligera mayor sensibilidad y especificidad para la estrategia 3 (Sensibilidad (S): 83.08%; Especificidad (E): 82.30%), seguido por la estrategia 1 (S: 82.58%; E: 82.07%), la estrategia 2 (S: 80.00%; E: 82.35%), y la estrategia 4 (S: 80.78%; E: 81.75%) *(Tabla 5)*. De la misma manera, se observó similares áreas bajo la curva para los modelos obtenidos mediante las cuatro estrategias de selección de variables *(Anexo 6)*.

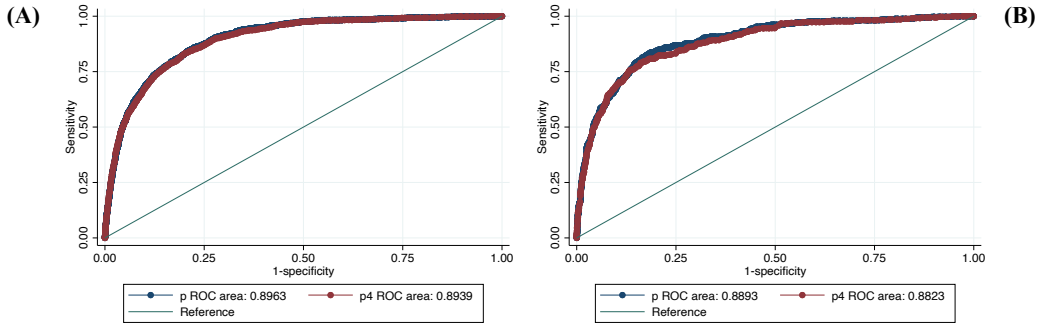
Tabla 3. Performance de los modelos de regresión construidos para la predicción de muerte en casos diagnosticados de COVID-19 en el dataset de validación (N= 4 420)

Performance diagnóstico	Modelo de la estrategia 1	Modelo de la estrategia 2	Modelo de la estrategia 3	Modelo de la estrategia 4
	Estimado (IC95%)*	**Estimado (IC95%)***	**Estimado (IC95%)***	**Estimado (IC95%)***
Sensibilidad	82.58%	80.06%	83.08%	80.78%
Especificidad	82.07%	82.35%	82.30%	81.75%
Área bajo la curva	0.89 (0.87 – 0.91)	0.88 (0.87 – 0.90)	0.89 (0.87 – 0.91)	0.88 (0.86 – 0.90)
Likelihood positivo	4.60	4.54	4.69	4.43
Likelihood negativo	0.21	0.24	0.21	0.24

**Punto de corte de probabilidad de tener el desenlace de muerte: 52%*

A pesar de que las cuatro estrategias de selección de variables resultaron en modelos con performances comparables, se optó por seleccionar la estrategia 1 y 4 de acuerdo a los criterios establecidos en este estudio. El modelo de la estrategia 1 incluyó un menor número de variables en comparación a la estrategia 3 (13 vs. 22) y, en comparación a la estrategia 2, el performance fue ligeramente mejor (AUC 0.89 vs 0.88). Por otro lado, el modelo de la estrategia 4 es un modelo resumido y tiene un performance comparable a las demás estrategias. Interesantemente, el área bajo de la curva de los modelos seleccionados en el dataset de validación y en el de creación fue comparable para las estrategias de los modelos seleccionados (Estrategia 1: 0.89 vs 0.89 / Estrategia 4: 0.89 vs 0.88; *Figura 1)*. Inclusive, utilizando el dataset de validación, el performance del modelo de la estrategia 1 y 4 fue estable cuando se realizó un análisis de sensibilidad incluyendo únicamente casos confirmados por RT-PCR (Estrategia 1: 0.89 vs 0.89/ Estrategia 4: 0.86 vs 0.88) y según periodos de la pandemia (Estrategia 1: primer periodo= 0.89 vs 0.89; segundo periodo= 0.89 vs 0.89; picos entre incidencia= 0.88 vs 0.89/ Estrategia 4: primer periodo= 0.88 vs 0.88; segundo periodo= 0.90 vs 0.88; picos entre incidencia= 0.87 vs 0.88) *(Anexo 7)*.

Figura 1. Validación del modelo seleccionado para la predicción de la muerte en casos diagnosticados de COVID–19



(A) Performance de los modelos seleccionados con la estrategia 1 y 4 en el dataset para creación de modelos (N =17 678); (B) Performance de los modelos seleccionado con la estrategia 1 y 4 en el dataset para validación de modelo (N = 4 420)

VI. DISCUSIONES

Letalidad por COVID-19

A lo largo de la pandemia, Perú se posicionó como el país con el mayor número de muertes por COVID-19 por cada 100 000 habitantes, con una tasa de letalidad de aproximadamente 5%, superando a países como México y Estados Unidos (52). La letalidad se ha concentrado principalmente en la ciudad de Lima, la capital del Perú, dada la alta densidad poblacional (1). La ciudad de Lima se encuentra rodeada por nueve provincias, y todas ellas en conjunto conforman el departamento de Lima. Lamentablemente, a nuestro conocimiento, a pesar de la gran conectividad entre la capital y las provincias, no se ha explorado la letalidad ni sus tendencias en habitantes de las provincias de Lima. En ese sentido, los resultados de este estudio brindan una inicial caracterización de esta población.

Nuestro estudio exploró diversos factores predictivos de muerte por COVID-19 en las nueve provincias del departamento de Lima, e identificó una letalidad de 7.5% correspondiente a la primera ola pandémica. Esta elevada letalidad es producto de la limitada y poca capacidad de respuesta del sistema de salud, así como también producto de la falta de unidades de cuidados intensivos, dispositivos médicos y personal especializado en los hospitales provinciales (13).

Modelo predictivo para la letalidad por COVID-19

Considerando que la pandemia por COVID-19 ha traído un desproporcionado uso de recursos sanitarios, ha sido necesario estructurar sistemas con el objetivo de priorizar poblaciones en estado de vulnerabilidad para prevenir la ocurrencia de desenlaces que afecten la salud de las personas. Nosotros construimos cuatro modelos de predicción considerando diferentes estrategias de selección de variables

para la predicción de muerte por COVID-19. En base a los criterios utilizados en este estudio para elegir los mejores modelos, se optó por elegir los modelos de las estrategias 1 y 4. No obstante, la utilización de los otros dos modelos es posible.

El modelo de la estrategia 1 incluyó variables con base epidemiológica discutida en estudios previos que, a nuestro conocimiento, son de los pocos que incluyeron características clínicas de la enfermedad (7, 8, 31). Dichos estudios utilizaron información poblacional de Estados Unidos y describieron modelos con perfomances notablemente buenos (AUC=0.80) (7, 8, 31) . Si bien se hipotetiza que el uso de modelos en poblaciones diferentes a las que fueron creados podría no ser adecuado dada las características diferentes entre poblaciones, sistema de salud, o epidemiología de la COVID-19, nosotros observamos que, el uso de las variables previamente descritas logró alcanzar un buen performance para la predicción de muerte en la población evaluada en este estudio. Incluso, observamos un AUC mayor de 0.8, el cual fue consistente en la validación por subgrupos poblacionales y en diferentes escenarios. El performance observado en este estudio es comparable al reportado por otros que usaron información poblacional (7, 8, 31) y que inclusive incorporaron un mayor número de variables clínicas (7, 31) y de laboratorio (8). Sin embargo, si bien las variables clínicas de nuestro modelo fueron obtenidas al enrolamiento, es probable que otros síntomas previos a la atención o síntomas que aparecieron durante la evolución jueguen un rol importante en la predicción. Por otro lado, nosotros consideramos algunas variables como proxys de variables no medidas. Por ejemplo, la variable fiebre e hipertensión arterial fueron consideradas como proxis de temperatura y presión arterial al enrolamiento, siendo estas últimas las que realmente fueron incluidas en los estudios previos (7, 8, 31). A pesar de

esto, consideramos que esta limitación no afecta de forma significativa el performance del modelo, debido a la estrecha relación de los proxis utilizados con las variables no medidas. Por otro lado, debido a que las comorbilidades fueron obtenidas mediante autoreporte, pudiendo incluso haber un sub-diagnóstico de algunas de estas, es probable que se subestime su efecto y aporte en el modelo creado, y consecuentemente se subestime el performance del modelo construido bajo la estrategia 1.

El modelo construido en base a la estrategia 4 únicamente incluyó tres variables, las cuales estuvieron consistentemente incluidas en las otras tres estrategias de selección. Si bien el modelo de la estrategia 4 es el más parsimónico y tiene un performance comparable al modelo de la estrategia 1, este fue construido en base a lo observado en la construcción de los otros modelos (post-hoc). Por tanto, la estrategia 4 podría estar afectada por múltiples limitaciones previamente descritas (53). No obstante, al ser consistente con modelos e información previamente descrita (54-56), tener un performance comparable a otros modelos más complejos (7, 8, 31), y ser consistente y robusto ante los diferentes escenarios planteados en el presente estudio, nosotros consideramos que puede ser utilizado con cautela.

A pesar de que todos los modelos descritos en este estudio son potencialmente útiles para predecir muerte, nosotros sugerimos el uso de los modelos generados por la estrategia 1 o 4. Para la elección de uno de estos, o de cualquier otro, proponemos que se realice una evaluación previa de la calidad de los datos. Por ejemplo, de no tener confianza en la forma en cómo se midió o recolectó información de las comorbilidades, se podría usar el modelo de la estrategia 4 para predecir muerte.

Por el contrario, en caso se tenga certeza de la calidad de los datos relacionados a comorbilidades, probablemente el modelo 1 sea el ideal para predecir muerte, considerando que podría alcanzarse un potencial ligero y mejor performance de predicción.

Los modelos seleccionados identificaron algunas variables que se asociaron significativamente con el desenlace, los cuales no variaron en el análisis de sensibilidad cuando se estimó los riesgos relativos de las mismas. Estos fueron la dificultad respiratoria, edad, el sexo masculino, la diabetes, la enfermedad renal, y la obesidad. A continuación, discutimos estos hallazgos.

La dificultad respiratoria fue de las características que incrementó más el odds de muerte, aproximadamente 7 veces más. Modelamientos de inteligencia artificial han identificado que la dificultad respiratoria es de las más predictivas para muerte por COVID-19 (7). Esto es debido a que los estadios severos de COVID-19 se caracterizan por presentar hipoxia, que a menudo conducen a una insuficiencia respiratoria, la cual resulta en disnea. Estudios previos han reportado una mayor probabilidad de muerte conforme decrece la saturación de oxígeno (57, 58), lo cual también se ha corroborado en la población peruana (14, 15, 59).

Identificamos que conforme la edad incrementa, el odds de muerte aumenta. Esta relación ha sido previamente descrita (60-62), observándose inclusive que la letalidad se incrementa en casi el 50% en grupos de edad avanzada (62). Esto probablemente se debe a la carga de morbilidad que trae el incremento de la edad, pues es conocido que, a mayor edad, hay mayor probabilidad de tener comorbilidades, polifarmacia, y fragilidad, por ende, un riesgo mayor de desenlaces

negativos para la salud. Es importante considerar que la relación entre edad y muerte puede ser diferente en escenarios con alta cobertura de vacunación en personas de edad avanzada y más altas tasas de infección en población adulta joven (62). Por tanto, la extrapolación de este hallazgo a contextos con estrategias de vacunación implementadas debe realizarse con cautela.

El ser varón incrementó aproximadamente dos veces el odds de muerte por COVID-19. Se hipotetiza que las hormonas sexuales influyen en la respuesta de la enfermedad. Previamente se ha sugerido que la mayor expresión de genes relacionados a la inmunidad y ubicados en los cromosomas X de las mujeres conllevan a producir una mayor respuesta de anticuerpos y una mejor protección contra las infecciones (63), como la de COVID-19. Recientemente, se ha reportado que las mujeres mantienen una alta reactividad inmunológica post-infecciones virales y generan títulos de anticuerpos más altos por las vacunas en comparación a los varones (64). Por tanto, si consideramos la extrapolación de lo observado a un escenario sin intervenciones basadas en el uso de vacunas a uno con uso de vacunas, inclusive, se esperaría que esta mayor posibilidad de muerte se mantenga elevada en varones.

Comorbilidades como la diabetes, la enfermedad renal, y la obesidad incrementaron el odds de muerte por COVID-19. Si bien se conoce que la COVID-19 produce mielopoyesis, desregulación de células T y de natural killers, y una producción descontrolada de citoquinas; aún se desconoce hasta qué punto una comorbilidad subyacente influye en la respuesta inmunitaria frente a la infección por SARS-CoV-2 (65). Sin embargo, estudios previos han reportado que la obesidad y la diabetes

conllevan a una desregulación de las células inmunes y hormonas, lo que conduce a una disminución de las defensas y una mayor probabilidad de hiperinflamación (65-67), que consecuentemente se traduce en una potenciación de la infección y en el desarrollo de múltiples eventos adversos para la salud. Por otro lado, en cuanto a la enfermedad renal crónica, se conoce que la función renal normal contribuye a la homeostasis inmunitaria, pues filtra las citoquinas circulantes y los componentes patógenos inmunogénicos y, por lo tanto, limita la inflamación (68). En este caso, la función renal disminuida conduce a una mayor activación de las células inmunes innatas y mayor producción de citoquinas (69), y posterior tormenta de citoquinas, generando así una exponenciación de la gravedad de la enfermedad (65). Si bien estudios previos han descrito la relación entre otras comorbilidades y muerte (65, 70-72), la limitación concerniente a la medición de nuestras variables puede tomar un rol importante en la no asociación encontrada.

Implicancias para la práctica clínica

En la actualidad, diferentes guías de práctica clínica sobre COVID-19, como las de Perú, incorporan modelos predictivos de mortalidad o empeoramiento de la enfermedad para la priorización de grupos vulnerables para el inicio de tratamiento, hospitalizaciones, etc (73); sin embargo, estas suelen incluir variables de laboratorio y basarse en estimaciones reportadas en otros países. Nuestro modelo, al incorporar únicamente variables clínicas y sociodemográficas puede ser utilizado en la práctica clínica en contextos donde las pruebas de laboratorio son escasas o no están disponibles, de esta forma, podría constituirse como una herramienta de aplicación en centros de atención de primer nivel. De este modo, es posible aportar con la identificación de grupos priorizados para la transferencia a centros de mayor

complejidad o realización de monitoreo de la enfermedad de forma continua. Así mismo, nuestro modelo, al ser validado en una población peruana, permite una mayor certeza de la evidencia para su extrapolación en la toma de decisiones.

Limitaciones y fortalezas

Este estudio está afectado por ciertas limitaciones. Primero, debido a que únicamente se analizó la información registrada en el sistema de vigilancia, la letalidad por COVID-19 podría encontrarse subestimada dado que no se incluyó información de: casos no notificados, casos que no acudieron a algún centro de salud, casos identificados pero no registrados en el sistema por los notificadores, o de casos que no fueron confirmados por falta de pruebas diagnósticas. Por lo que la extrapolación de los modelos sólo aplica para personas sintomáticas con pruebas confirmatorias para COVID-19. Segundo, dado que en el Perú las pruebas serológicas que detectan anticuerpos fueron implementadas como diagnósticas de COVID-19, nuestros resultados podrían estar afectados por mala clasificación asociada al uso de una prueba de limitado valor diagnóstico. Sin embargo, para lidiar con esta limitación, se realizó un análisis de sensibilidad que incluyó únicamente casos diagnosticados por RT-PCR. Esta evaluación sugirió que, a pesar de existir mala clasificación, el performance de los modelos fue robusto en el subgrupo poblacional analizado. Tercero, los modelos construidos en este estudio únicamente toman en cuenta la primera infección de los individuos para evitar las múltiples mediciones de la infección, excluyendo la posibilidad de analizar el número de reinfecciones como variable predictora. Futuros estudios deberán corroborar si esta variable influiría sustancialmente en el performance de predicción de letalidad. Cuarto, para la construcción y estimación de los modelos utilizamos

regresiones logística simples, por ser el tipo más conocido y de fácil interpretación. Sin embargo, es importante tomar en cuenta que para outcomes binarios existen otros métodos para construir y estimar el performance de modelos de predicción, como el modelo de regresión logística aditiva generalizada, redes neuronales, splines de regresión aditiva multivariante, y otras extensión de la regresión logística para datos correlacionados (74). Así mismo, es posible usar análisis de supervivencia o tiempo-evento para predecir este tipo de desenlace; sin embargo, son modelos menos flexibles que la regresión logística que ameritan asumpsiones estrictas (74). Quinto, debido a que el sistema de vigilancia no contaba con información de laboratorio, como leucocitos, grupo sanguíneo, plaquetas, dímero D, entre otros, no fue posible su evaluación dentro de los modelos de predicción. Si bien modelos de predicción de letalidad de COVID-19 han pretendido incluir estas variables, estas no han aportado sustancialmente en la mejora del performance, obteniéndose estimaciones comparables a los modelos donde se incluye únicamente variables clínicas (20). Por lo que su aporte en un contexto de escasos recursos sanitarios, como el peruano (y más aún provincia), podría no ser beneficioso por temas de factibilidad, pues estas sólo estarían disponibles después de haber accedido a una atención de emergencia, hospitalaria o ambulatoria. Así mismo, variables relacionadas con los recusos sanitarios (por ejemplo número de camas hospitalarias disponibles, etc.) no fueron consideradas dentro de los modelos a pesar que podrían relacionarse con el desenlace de letalidad. Si bien dicho dato podría haber sido accesible para el periodo en el que esta población fue enrolada, considerando que los datos de los recursos sanitarios disponibles durante la pandemia eran de libre acceso; en la actualidad, dicho reporte no se encuentra

actualizado e inclusive su obtención podría no ser factible. Por lo que, su inclusión, complejizaría un modelo que prentede ser aplicado en diferentes contextos en la práctica clínica. Así mismo, considerando que el performance de los modelos contruidos en este estudio reportan áreas bajo la curva que oscilan entre el 80%, consideramos que el aporte de dicha variable no sería sustancialmente importante. Sexto, la búsqueda de un mejor sistema de salud, como el de la capital del departamento de Lima, pudo haber motivado el desplazamiento de casos hacia otras localidades diferentes a las analizadas. Por lo cual, estos casos podrían no haber sido notificados a la DIRESA-LIPRO, ni haber sido incluidos en los análisis. Aun así, dada las restricciones de movilidad implementadas, particularmente durante la primera ola pandémica, es posible que, de haber existido movilización, esta haya sido mínima y no afecte nuestros resultados de forma significativa. Por último, considerando que los datos fueron recolectados durante los primeros periodos de la pandemia de COVID-19, la extrapolación de los resultados a un contexto actual (diferentes variantes virales y vacunación) debe ser cautelosa. Sin embargo, si bien esperariamos que en estos contexto el performance del modelo sea afectado, no esperaríamos una diferencia en las variables incluidas en este. Inclusive, estudios previos en contextos con predominancia de otras variantes virales, la evidencia actual refiere un cambio en las tendencias de letalidad según el tipo de variante (75), pero no un cambio en los factores predictores conocidos desde el inicio de la pandemia (76, 77).

A pesar de las limitaciones descritas, a nuestro conocimiento, este estudio es uno de los pocos que ha explorado múltiples estrategias de selección de variables para la creación de modelos predictivos para una población latinoamericana, y también

uno de los pocos que ha realizado una validación de modelos previamente creados en otros contextos a nivel poblacional. Así mismo, este estudio analiza datos derivados de un sistema específico que incluye una población de provincias aledañas a la capital del país, por lo cual, poseen características diferentes a la capital que se asemejan a otros departamentos del Perú. En general, el estudio aporta al conocimiento actual, incrementando su certeza de la evidencia principalmente en términos de precisión y evidencia directa para ser aplicado en la toma de decisiones en el contexto peruano. Por último, presentamos un ejercicio que puede ser utilizado para otros escenarios similares a la pandemia por COVID-19, permitiendo tener modelos locales de rápida validacion.

VII. CONCLUSIONES

En este estudio se describen cuatro modelos con diferentes estrategias de selección de variables para la predicción de muerte por COVID-19. Los resultados sugieren que, independientemente de la estrategia, los modelos tuvieron performances comparables. Sin embargo, dos modelo de predicción; cuyas área bajo la curva fueron óptimas tanto en la validación y en diferentes escenarios construidos en los análisis de sensibilidad, mostraron superioridad debido a su menor número de variables incluidas y ligero mayor performance para la predicción de la letalidad en COVID-19. El primer modelo estuvo constituido por 13 variables incluyendo; dos variables sociodemográficas, dos variables de síntomas del cuadro clínico, y nueve variables sobre las comorbilidades. Por otro lado, el segundo modelo seleccionado es un resumen del primer modelo, el cual únicamente incluye tres variables (dos sociodemográficas y un síntomas del cuadro clínico). Futuros estudios deberán corroborar el performance y validar la utilidad de los modelos descritos bajo condiciones reales y actuales; inclusive valorar su uso en otras pandemias virales con características similares.

VIII. RECOMENDACIONES

- Se sugiere realizar validaciones externas de los modelos seleccionados en poblaciones de otros departamentos del Perú.
- Se sugiere realizar validaciones externas en el contexto actual de COVID-19, que se caracteriza por la incorporación de vacunas y nuevas variantes en la población.
- Corroborar la utilidad de los modelos seleccionados basándose en el uso de datos objetivos sobre las comorbilidades mediante estudios poblacionales.

IX. REFERENCIAS BIBLIOGRÁFICAS

1. Johns Hopkins University Center for Systems Science and Engineering [Internet]. USA: CSSE; 2021 [citado 17 jul 2021] COVID-19 Data Repository. [Disponible en: https://www.arcgis.com/apps/dashboards/index.html#/bda7594740fd40299423467b48e9ecf6].

2. Wu JT, Leung K, Bushman M, Kishore N, Niehus R, de Salazar PM, et al. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. Nature Medicine. 2020;26(4):506-10.

3. Coronavirus Resource Center [Internet]. USA: JHU; 2023 [citado 26 julio 2023] Mortality analyses [Disponible en: https://coronavirus.jhu.edu/data/mortality].

4. Siddiqui SH, Sarfraz A, Rizvi A, Shaheen F, Yousafzai MT, Ali SA. Global variation of COVID-19 mortality rates in the initial phase. Osong Public Health Res Perspect. 2021;12(2):64-72.

5. World Health Organization PAHO. Why predictive modeling is critical in the fight against COVID-19? In: Department of evidence and intelligence for action health, editor. Suiza: Pan American Health Organization; 2020.

6. Velasco-Rodríguez D, Alonso-Dominguez J-M, Vidal Laso R, Lainez-González D, García-Raso A, Martín-Herrero S, et al. Development and validation of a predictive model of in-hospital mortality in COVID-19 patients. PLOS ONE. 2021;16(3):e0247676.

7. Yadaw AS, Li Y-c, Bose S, Iyengar R, Bunyavanich S, Pandey G. Clinical features of COVID-19 mortality: development and validation of a clinical prediction model. The Lancet Digital Health. 2020;2(10):e516-e25.

8. Akama-Garren EH, Li JX. Unbiased identification of clinical characteristics predictive of COVID-19 severity. Clin Exp Med. 2021:1-13.

9. Ramspek CL, Jager KJ, Dekker FW, Zoccali C, van Diepen M. External validation of prognostic models: what, why, how, when and where? Clin Kidney J. 2021;14(1):49-58.

10. Vasquez-Apestegui BV, Parras-Garrido E, Tapia V, Paz-Aparicio VM, Rojas JP, Sanchez-Ccoyllo OR, et al. Association between air pollution in Lima and the high incidence of COVID-19: findings from a post hoc analysis. BMC Public Health. 2021;21(1):1161.

11. Levin AT, Owusu-Boaitey N, Pugh S, Fosdick BK, Zwi AB, Malani A, et al. Assessing the burden of COVID-19 in developing countries: systematic review, meta-analysis and public policy implications. BMJ Glob Health. 2022;7(5).

12. Garcia PJ, Alarcón A, Bayer A, Buss P, Guerra G, Ribeiro H, et al. COVID-19 Response in Latin America. Am J Trop Med Hyg. 2020;103(5):1765-72.

13. Ramírez-Soto MC, Ortega-Cáceres G. Analysis of Excess All-Cause Mortality and COVID-19 Mortality in Peru: Observational Study. Trop Med Infect Dis. 2022;7(3).

14. Yupari-Azabache I, Bardales-Aguirre L, Rodriguez-Azabache J, Barros-Sevillano JS, Rodríguez-Diaz Á. Factores de riesgo de mortalidad por COVID-19 en pacientes hospitalizados: Un modelo de regresión logística. Revista de la Facultad de Medicina Humana. 2021;21:19-27.

15. Hueda-Zavaleta M, Copaja-Corzo C, Bardales-Silva F, Flores-Palacios R, Barreto-Rocchetti L, Benites-Zapata VA. Factores asociados a la muerte por COVID-19 en pacientes admitidos en un hospital público en Tacna, Perú. Revista Peruana de Medicina Experimental y Salud Pública. 2021;38(2).

16. Murrugarra-Suarez S, Lora-Loza M, Cabrejo-Paredes J, Mucha-Hospinal L, Fernandez-Cosavalente H. Factores asociados a mortalidad en pacientes Covid- 19 en un Hospital del norte de Perú. Revista del Cuerpo Médico Hospital Nacional Almanzor Aguinaga Asenjo. 2020;13:378-85.

17. Rodríguez-Zúñiga MJM, Quintana-Aquehua A, Díaz-Lajo VH, Charaja-Coata KS, Becerra-Bonilla WS, Cueva-Tovar K, et al. Factores de riesgo asociados a mortalidad en pacientes adultos con neumonía por SARS- CoV-2 en un hospital público de Lima, Perú. Acta Médica Peruana. 2020;37:437-46.

18. Ahmad N, Hasan MG, Barbhuiya RK. Identification and prioritization of strategies to tackle COVID-19 outbreak: A group-BWM based MCDM approach. Appl Soft Comput. 2021;111:107642.

19. Russo AG, Decarli A, Valsecchi MG. Strategy to identify priority groups for COVID-19 vaccination: A population based cohort study. Vaccine. 2021;39(18):2517-25.

20. de Jong VMT, Rousset RZ, Antonio-Villa NE, Buenen AG, Van Calster B, Bello-Chavolla OY, et al. Clinical prediction models for mortality in patients with covid-19: external validation and individual participant data meta-analysis. BMJ. 2022;378:e069881.

21. Soto A. Barreras para una atención eficaz en los hospitales de referencia del Ministerio de Salud del Perú: atendiendo pacientes en el siglo XXI con recursos del siglo XX. Revista peruana de medicina experimental y salud pública. 2019;36:304-11.

22. Pathak EB, Menard JM, Garcia RB, Salemi JL. Joint Effects of Socioeconomic Position, Race/Ethnicity, and Gender on COVID-19 Mortality among Working-Age Adults in the United States. Int J Environ Res Public Health. 2022;19(9).

23. Ford JD, Zavaleta-Cortijo C, Ainembabazi T, Anza-Ramirez C, Arotoma-Rojas I, Bezerra J, et al. Interactions between climate and COVID-19. The Lancet Planetary Health. 2022;6(10):e825-e33.

24. Stephens KE, Chernyavskiy P, Bruns DR. Impact of altitude on COVID-19 infection and death in the United States: A modeling and observational study. PLOS ONE. 2021;16(1):e0245055.

25. Chowdhury SR, Chandra Das D, Sunna TC, Beyene J, Hossain A. Global and regional prevalence of multimorbidity in the adult population in community settings: a systematic review and meta-analysis. eClinicalMedicine. 2023;57.

26. Yamamoto N, Yamamoto R, Ariumi Y, Mizokami M, Shimotohno K, Yoshikura H. Does Genetic Predisposition Contribute to the Exacerbation of COVID-19 Symptoms in Individuals with Comorbidities and Explain the Huge Mortality Disparity between the East and the West? Int J Mol Sci. 2021;22(9).

27. López MGF, Tarazona AS, Cruz-Vargas JADL. Distribución regional de mortalidad por Covid-19 en Perú. Revista de la Facultad de Medicina Humana. 2021;21:326-34.

28. Montenegro-Idrogo JJ, Chiappe González AJ. Ejecución presupuestal descentralizada y letalidad por COVID-19 en Perú. Revista peruana de medicina experimental y salud pública. 2020;37(4):781-2.

29. Congreso de la república. Avance de la ejecución del gasto público destinado a la lucha contra el COVID-19. In: 51/2020-2021 RtNo, editor. Perú: Congreso de la república del Perú,; 2020.

30. Zegarra Zamalloa CO, Contreras PJ, Orellana LR, Riega Lopez PA, Prasad S, Cuba Fuentes MS. Social vulnerability during the COVID-19 pandemic in Peru. PLOS Global Public Health. 2022;2(12):e0001330.

31. Banoei MM, Dinparastisaleh R, Zadeh AV, Mirsaeidi M. Machine-learning-based COVID-19 mortality prediction model and identification of patients at low and high risk of dying. Crit Care. 2021;25(1):328.

32. National Library of Medicine [Internet]. USA: NIH; 2021 [citado 10 december 2022] SARS-CoV-2 [Disponible en: https://www.ncbi.nlm.nih.gov/mesh/2052180].

33. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet. 2020;395(10224):565-74.

34. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020;579(7798):270-3.

35. Chan JF, Kok KH, Zhu Z, Chu H, To KK, Yuan S, et al. Genomic

characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. Emerg Microbes Infect. 2020;9(1):221-36.

36. Li J, Huang DQ, Zou B, Yang H, Hui WZ, Rui F, et al. Epidemiology of COVID-19: A systematic review and meta-analysis of clinical characteristics, risk factors, and outcomes. J Med Virol. 2021;93(3):1449-58.

37. Grant MC, Geoghegan L, Arbyn M, Mohammed Z, McGuinness L, Clarke EL, et al. The prevalence of symptoms in 24,410 adults infected by the novel coronavirus (SARS-CoV-2; COVID-19): A systematic review and meta-analysis of 148 studies from 9 countries. PLOS ONE. 2020;15(6):e0234765.

38. Cohen PA, Hall LE, John JN, Rapoport AB. The Early Natural History of SARS-CoV-2 Infection: Clinical Observations From an Urban, Ambulatory COVID-19 Clinic. Mayo Clin Proc. 2020;95(6):1124-6.

39. Tostmann A, Bradley J, Bousema T, Yiek WK, Holwerda M, Bleeker-Rovers C, et al. Strong associations and moderate predictive value of early symptoms for SARS-CoV-2 test positivity among healthcare workers, the Netherlands, March 2020. Euro Surveill. 2020;25(16).

40. Infectouse Disease Society of America. COVID-19 Prioritization of Diagnostic Testin USA: IDSA; 2020 [

41. World Health Organization. Recommendations for national SARS-CoV-2 testing strategies and diagnostic capacities. Suiza: WHO; 2021.

42. Johns Hopkins University [Internet]. USA: JHU; 2022 [citado 16 dic 2022] COVID-19 Data Repository [Disponible en: https://github.com/CSSEGISandData/COVID-19].

43. Ministerio de Salud del Perú. Directiva sanitaria para la vigilancia epidemiológica de la enfermedad por coronavirus en el Perú. In: CDC, editor. Perú: MINSA; 2020.

44. Instituto Nacional de Estadística e Informática. Características de la Población. In: INEI, editor. Peru: INEI; 2017.

45. González-García N, Castilla-Peón MF, Solórzano Santos F, Jiménez-Juárez RN, Martínez Bustamante ME, Minero Hibert MA, et al. Covid-19 Incidence and Mortality by Age Strata and Comorbidities in Mexico City: A Focus in the Pediatric Population. Frontiers in Public Health. 2021;9.

46. Riley RD, Ensor J, Snell KIE, Harrell FE, Martin GP, Reitsma JB, et al. Calculating the sample size required for developing a clinical prediction model. BMJ. 2020;368:m441.

47. Thakur B, Dubey P, Benitez J, Torres JP, Reddy S, Shokar N, et al. A systematic review and meta-analysis of geographic differences in

comorbidities and associated severity and mortality among individuals with COVID-19. Scientific Reports. 2021;11(1):8562.

48. Ministerio de Salud del Perú [Internet]. Peru: MINSA; 2022 [citado 20 nov 2022] SINADEF: Sistema Informático Nacional de Defunciones [Disponible en: https://www.minsa.gob.pe/defunciones/].

49. Hasanin T, Khoshgoftaar TM, Leevy JL, Seliya N. Examining characteristics of predictive models with imbalanced big data. Journal of Big Data. 2019;6(1):69.

50. Anand A, Pugalenthi G, Fogel GB, Suganthan PN. An approach for classification of highly imbalanced data using weighting and undersampling. Amino Acids. 2010;39(5):1385-91.

51. Mitchell Gail JMS, B. Singer,. Clinical prediction model: A Practical Approach to Development,

Validation, and Updating. Statistics for Biology and Health, editor. Switzerland: Springer Nature 2019.

52. Johns Hopkins [Internet]. JHU: USA; 2022 [citado 18 agosto 2022] Mortality Analyses [Disponible en: https://coronavirus.jhu.edu/data/mortality].

53. Curran-Everett D, Milgrom H. Post-hoc data analysis: benefits and limitations. Curr Opin Allergy Clin Immunol. 2013;13(3):223-4.

54. Geldsetzer P, Mukama T, Jawad NK, Riffe T, Rogers A, Sudharsanan N. Sex differences in the mortality rate for coronavirus disease 2019 compared to other causes of death: an analysis of population-wide data from 63 countries. Eur J Epidemiol. 2022;37(8):797-806.

55. Bonanad C, García-Blas S, Tarazona-Santabalbina F, Sanchis J, Bertomeu-González V, Fácila L, et al. The Effect of Age on Mortality in Patients With COVID-19: A Meta-Analysis With 611,583 Subjects. J Am Med Dir Assoc. 2020;21(7):915-8.

56. Mas-Ubillus G, Ortiz PJ, Huaringa-Marcelo J, Sarzo-Miranda P, Muñoz-Aguirre P, Diaz-Ramos A, et al. High mortality among hospitalized adult patients with COVID-19 pneumonia in Peru: A single centre retrospective cohort study. PLOS ONE. 2022;17(3):e0265089.

57. Chen T, Wu D, Chen H, Yan W, Yang D, Chen G, et al. Clinical characteristics of 113 deceased patients with coronavirus disease 2019: retrospective study. BMJ. 2020;368:m1091.

58. Grasselli G, Zangrillo A, Zanella A, Antonelli M, Cabrini L, Castelli A, et al. Baseline Characteristics and Outcomes of 1591 Patients Infected With SARS-CoV-2 Admitted to ICUs of the Lombardy Region, Italy. JAMA. 2020;323(16):1574-81.

59. Tapia Cruz M. Características clinico-epidemiológicas y factores asociados a mortalidad en pacientes adultos menores de 60 años con neumonía por COVID-19. 2020.

60. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. Jama. 2020;323(13):1239-42.

61. Mehra MR, Desai SS, Kuy S, Henry TD, Patel AN. Cardiovascular Disease, Drug Therapy, and Mortality in Covid-19. New England Journal of Medicine. 2020;382(25):e102.

62. Elo IT, Luck A, Stokes AC, Hempstead K, Xie W, Preston SH. Evaluation of Age Patterns of COVID-19 Mortality by Race and Ethnicity From March 2020 to October 2021 in the US. JAMA Network Open. 2022;5(5):e2212686-e.

63. Taneja V. Sex Hormones Determine Immune Response. Front Immunol. 2018;9:1931.

64. Klein SL, Marriott I, Fish EN. Sex-based differences in immune function and responses to vaccination. Trans R Soc Trop Med Hyg. 2015;109(1):9-15.

65. Kreutmair S, Kauffmann M, Unger S, Ingelfinger F, Núñez NG, Alberti C, et al. Preexisting comorbidities shape the immune response associated with severe COVID-19. Journal of Allergy and Clinical Immunology. 2022;150(2):312-24.

66. Cai S-H, Liao W, Chen S-W, Liu L-L, Liu S-Y, Zheng Z-D. Association between obesity and clinical prognosis in patients infected with SARS-CoV-2. Infectious Diseases of Poverty. 2020;9(1):80.

67. Guo W, Li M, Dong Y. Diabetes is a risk factor for the progression and prognosis of COVID-19 [published online March 31, 2020]. Diabetes Metab Res Rev.

68. Tecklenborg J, Clayton D, Siebert S, Coley SM. The role of the immune system in kidney disease. Clin Exp Immunol. 2018;192(2):142-50.

69. Girndt M, Sester M, Sester U, Kaul H, Köhler H. Defective expression of B7-2 (CD86) on monocytes of dialysis patients correlates to the uremia-associated immune defect. Kidney Int. 2001;59(4):1382-9.

70. Rabbani G, Shariful Islam SM, Rahman MA, Amin N, Marzan B, Robin RC, et al. Pre-existing COPD is associated with an increased risk of mortality and severity in COVID-19: a rapid systematic review and meta-analysis. Expert Review of Respiratory Medicine. 2021;15(5):705-16.

71. Liang X, Shi L, Wang Y, Xiao W, Duan G, Yang H, et al. The association of hypertension with the severity and mortality of COVID-19 patients: Evidence

based on adjusted effect estimates. J Infect. 2020;81(3):e44-e7.

72. Desai A, Sachdeva S, Parekh T, Desai R. COVID-19 and Cancer: Lessons From a Pooled Meta-Analysis. JCO Glob Oncol. 2020;6:557-9.

73. EsSalud. Guía de práctica clínica: Manejo de COVID-19. In: Instituto de Evaluación de Tecnologías en Salud e Investigación, editor. Perú: IETSI; 2021.

74. Ewout W. Steyerberg. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. Mitchell Gail JMS, B. Singer,, editor. USA: Springer; 2019.

75. Stepanova M, Lam B, Younossi E, Felix S, Ziayee M, Price J, et al. The impact of variants and vaccination on the mortality and resource utilization of hospitalized patients with COVID-19. BMC Infectious Diseases. 2022;22(1):702.

76. Sharma J, Rajput R, Bhatia M, Arora P, Sood V. Clinical Predictors of COVID-19 Severity and Mortality: A Perspective. Front Cell Infect Microbiol. 2021;11:674277.

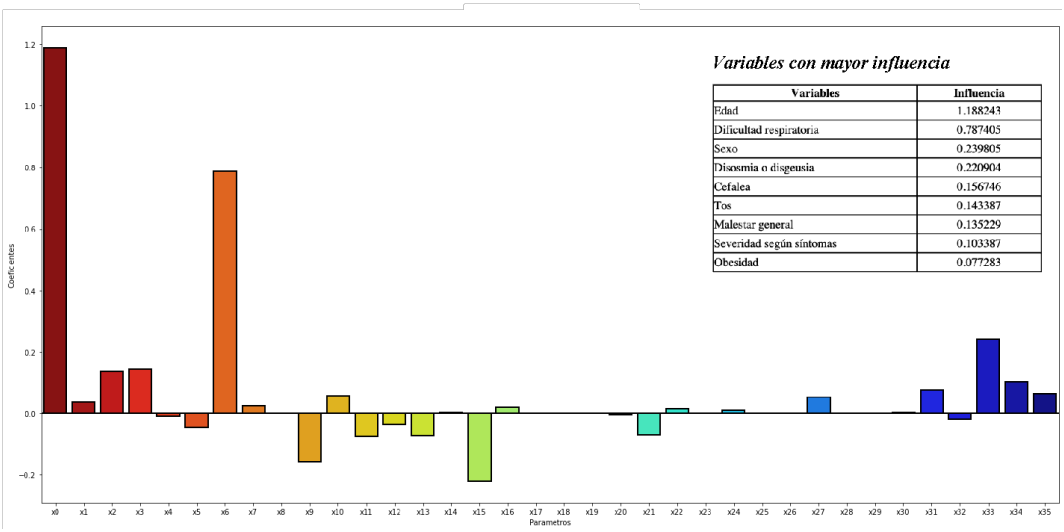
77. Hippisley-Cox J, Khunti K, Sheikh A, Nguyen-Van-Tam JS, Coupland CAC. Risk prediction of covid-19 related death or hospital admission in adults testing positive for SARS-CoV-2 infection during the omicron wave in England (QCOVID4): cohort study. BMJ. 2023;381:e072976.

XIII. ANEXOS

Anexo 1. Distribución de la población total de las provincias de Lima en base a la población total del Perú

**Los datos de la distribución del porcentaje de población de las provincias de Lima fue obtenido según la última información reportada por el Centro Nacional de Epidemiología, Prevención y Control de Enfermedades en el 2016*

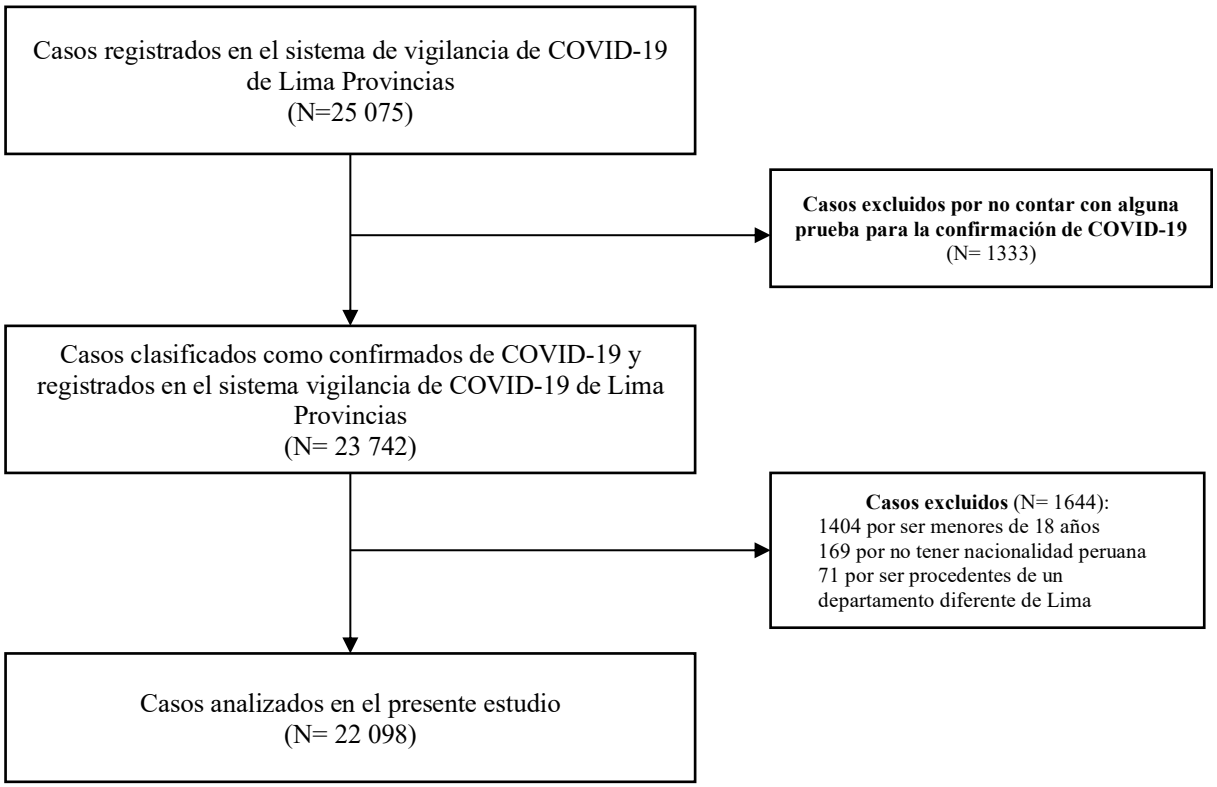
Anexo 2. Desarrollo de la selección de variables mediante técnica de Lasso



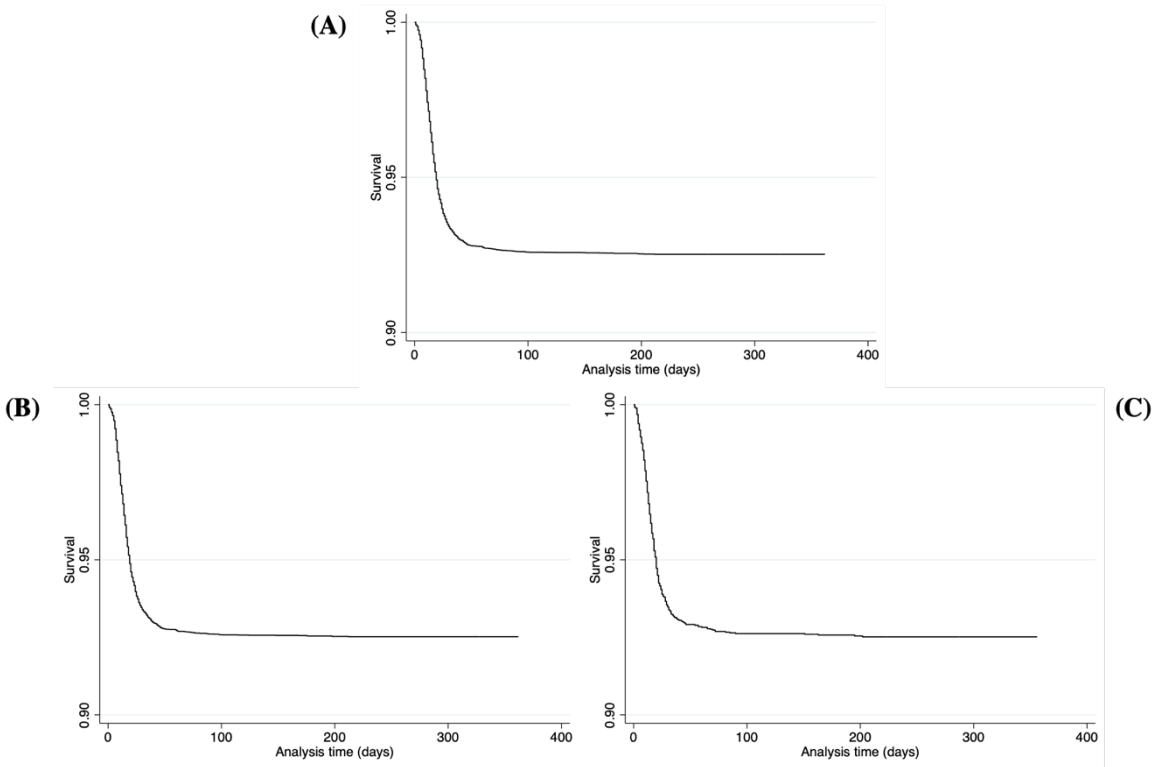
Variables	Influencia
Edad	1.188243
Dificultad respiratoria	0.787405
Sexo	0.239805
Disosmia o disgeusia	0.220904
Cefalea	0.156746
Tos	0.143387
Malestar general	0.135229
Severidad según síntomas	0.103387
Obesidad	0.077283

***Leyenda:** Edad (x0); Fiebre (x1); Malestar general (x2); Tos (x3); Dolor de garganta (x4); Congestión nasal (x5); Dificultad respiratoria (x6); Diarrea (x7); Vómitos (x8); Cefalea (x9); Confusión (x10); Dolor muscular (x11); Dolor abdominal (x12); Dolor tórax (x13); Dolor articular (x14); Disosmia o disgeusia (x15); Dolor de oído (x16); Embarazo (x17); Aborto (x18); Enfermedad cardiovascular (x19); Hipertensión (x20); Dislipidemia (x21); Diabetes (x22); Tiroidopatia (x23); Hepatopatia (x24); Enfermedad neurológica (x25); Inmunodeficiencia (x26); Enfermedad renal (x27); Enfermedad pulmonar (x28); Asma (x29); Cáncer (x30); Obesidad (x31); Tuberculosis (x32); Sexo (x33); Severidad según síntomas (x34); Número de comorbilidades (x35)*

Anexo 3. Flujograma de selección de casos



Anexo 4. Supervivencia de los casos diagnosticados con COVID-19 en el departamento de Lima, Perú



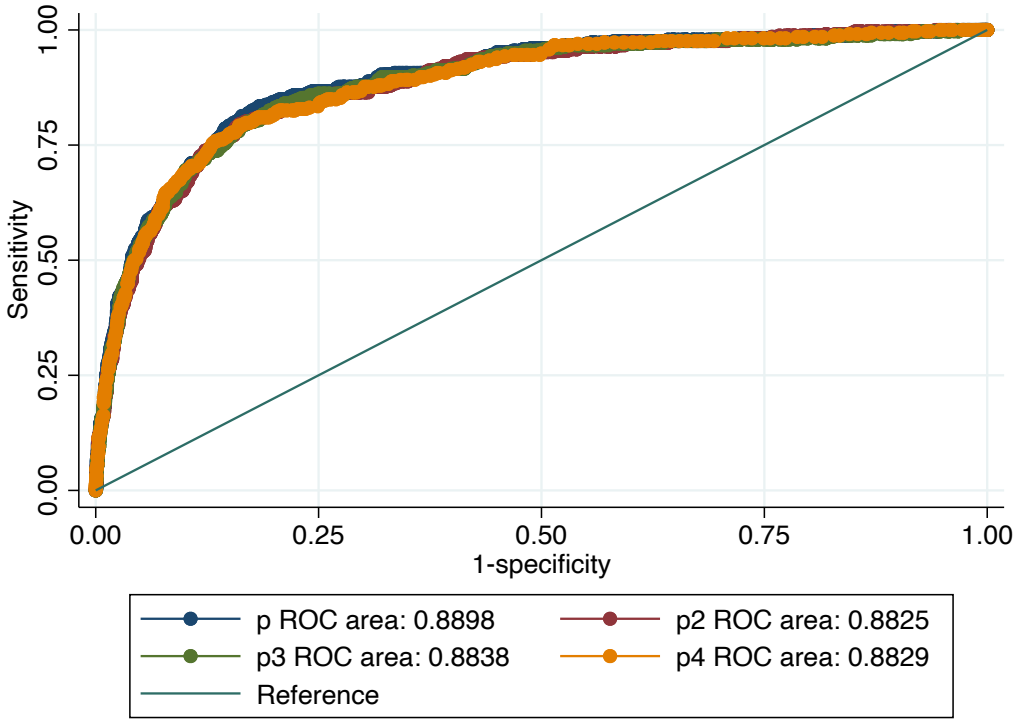
(A) Supervivencia global para la población general (B) Supervivencia global en el dataset para creación de modelos (C) Supervivencia global en el dataset para validación de modelo

Anexo 5. Análisis de sensibilidad de variables predictoras para la letalidad en pacientes con COVID-19 en análisis de regresión de poisson (N= 17 678)

Variables	Modelo de la estrategia 1	Modelo de la estrategia 2	Modelo de la estrategia 3	Modelo de la estrategia 4
	RR (IC95%)	**RR (IC95%)**	**RR (IC95%)**	**RR (IC95%)**
Edad	**1.06 (1.05 - 1.06)**	**1.06 (1.05 –1.06)**	**1.06 (1.05 – 1.06)**	**1.06 (1.05 – 1.06)**
Sexo				
Femenino	Ref.	Ref.	Ref.	Ref.
Masculino	**1.47 (1.33 - 1.63)**	**1.43 (1.30 – 1.58)**	**1.46 (1.32 – 1.61)**	**1.47 (1.33 – 1.62)**
Cuadro clínico				
Fiebre	1.09 (0.99 – 1.20)	–	1.09 (0.99 – 1.20)	–
Malestar general	–	**1.24 (1.12 – 1.36)**	**1.22 (1.11 – 1.35)**	–
Tos	–	**1.34 (1.18 – 1.51)**	**1.35 (1.19 – 1.53)**	–
Dolor de garganta	–	–	0.97 (0.88 – 1.07)	–
Congestión nasal	–	–	0.90 (0.80 – 1.01)	–
Sensación de dificultad respiratoria	**4.65 (4.12 – 5.24)**	**4.48 (3.97 – 5.06)**	**4.41 (3.90 – 4.98)**	**4.81 (4.28 – 5.40)**
Diarrea	–	–	–	–
Náuseas y vómitos	–	–	–	–
Cefalea	–	**0.75 (0.68 – 0.84)**	0.75 (0.67 – 0.83)	–
Confusión	–	–	**1.33 (1.00 – 1.75)**	–
Dolor muscular	–	–	–	–
Dolor abdominal	–	–	–	–
Dolor de tórax	–	–	–	–
Dolor articular	–	–	1.02 (0.80 – 1.29)	–
Disosmia y disgeusia	–	**0.23 (0.11 – 0.49)**	–	–
Dolor de oído	–	–	–	–
Severidad según síntomas				
Sin síntomas de severidad		Ref.	Ref.	
Con síntomas de severidad	–	1.10 (0.98 – 1.25)	1.12 (0.99 – 1.27)	–
Comorbilidades				
Enfermedad cardiovascular	1.03 (0.77 – 1.37)	–	0.93 (0.78 – 1.12)	–
Hipertensión arterial	0.93 (0.73 – 1.19)	–	0.92 (0.71 – 1.20)	–
Dislipidemia	–	–	–	–
Diabetes	**1.24 (1.06 – 1.45)**	**–**	1.12 (0.92 – 1.36)	–
Tiroidopatía	–	–	–	–
Hepatopatía	–	–	0.87 (0.56 – 1.34)	–
Enfermedad neurológica	–	–	0.96 (0.65 – 1.42)	–
Inmunodeficiencia	1.27 (0.25 – 6.57)	–	–	–

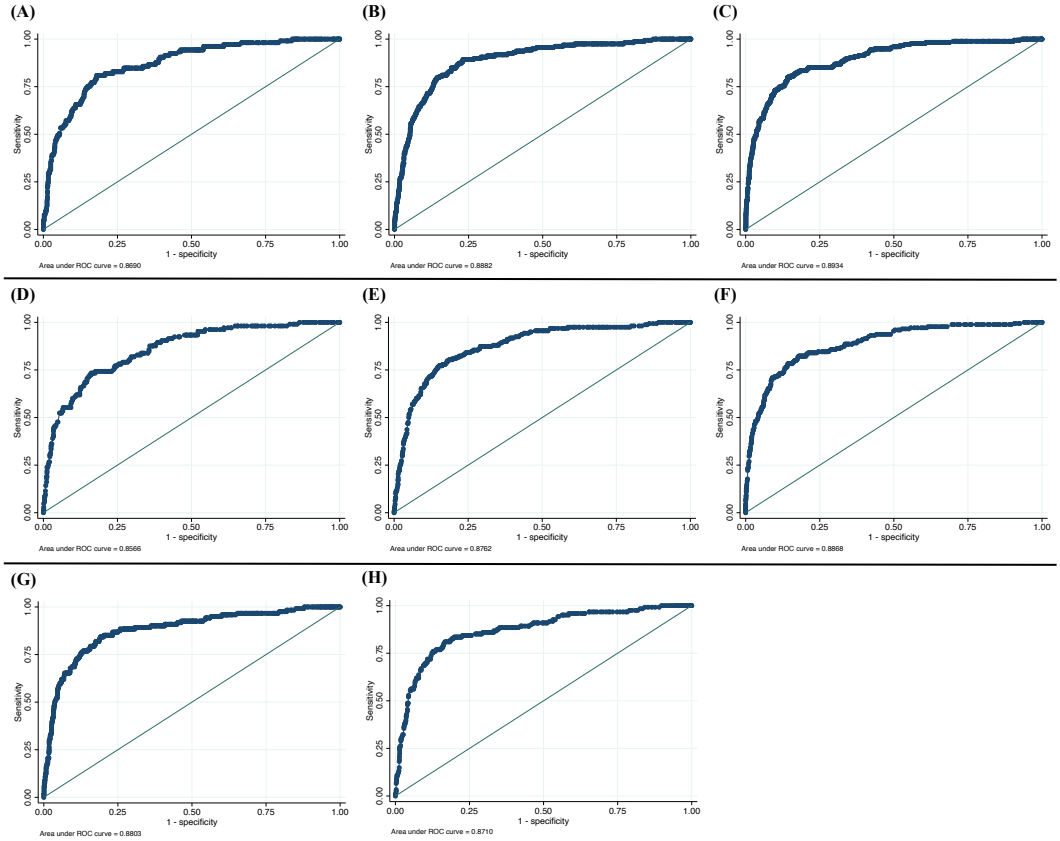
Enfermedad renal	**1.44 (1.05 – 2.01)**	**–**	**1.44 (1.01 – 2.05)**	–
Enfermedad pulmonar	1.03 (0.77 – 1.37)	–	0.93 (0.68 – 1.26)	–
Asma	1.03 (0.69 – 1.54)	–	–	–
Cáncer	1.20 (0.80 – 1.81)	–	1.03 (0.69 – 1.56)	–
Obesidad	**1.65 (1.33 – 2.05)**	**1.58 (1.26 – 1.96)**	–	–
Tuberculosis	–	–	–	–
Número de comorbilidades				
Sin ninguna comorbilidad			Ref.	
Con una o dos comorbilidades	–	–	**1.23 (1.03 – 1.46)**	–
Más de tres comorbilidades	–	–	1.17 (0.70 – 1.96)	–

Anexo 6. Curvas ROC del performance de predicción entre los cuatros modelos construidos



**Los valores de áreas bajo la curva especificados, corresponden a los siguientes modelos: 1) "p ROC area": Modelo de la estrategia 1; 2) "p2 ROC area": Modelo de la estrategia 2; 3) "p3 ROC area": Modelo de la estrategia 3; 4) "p4 ROC area": Modelo de la estrategia 4*

Anexo 7. Análisis de sensibilidad de performance de modelo seleccionado en dataset para validación



1. Validación del modelo de la estrategia 1 (A) en casos confirmados por RT-PCR, y en casos confirmados que fueron notificados durante (B) el primer periodo y el (C) segundo periodo de la pandemia; 2. Validación del modelo de la estrategia 4 (D) en casos confirmados por RT-PCR, y en casos confirmados que fueron notificados durante (E) el primer periodo y el (F) segundo periodo de la pandemia; 3. Validación del modelo de la estrategia 1 en el periodo entre picos de incidencia de COVID-19 (G) y (H) validación del modelo de la estrategia 4 en el periodo entre picos de incidencia de COVID-19