

**VALIDACIÓN DE ESCALAS DE PREDICCIÓN DE MORTALIDAD EN  
PACIENTES CON ENFERMEDAD POR  
CORONAVIRUS 2019 EN EL REGISTRO DE PACIENTES RESCATE COVID19**

**Custodio Alberto Ruiz Bedoya, MD**  
Médico Especialista en Medicina Interna

**Universidad Autónoma de Bucaramanga  
Facultad de Salud  
Escuela de Medicina  
Maestría en Métodos para la Producción y Aplicación del Conocimiento  
Científico en Salud  
Bucaramanga  
2021**

**VALIDACIÓN DE ESCALAS DE PREDICCIÓN DE MORTALIDAD EN  
PACIENTES CON ENFERMEDAD POR  
CORONAVIRUS 2019 EN EL REGISTRO DE PACIENTES RESCATE COVID19**

**Custodio Alberto Ruiz Bedoya, MD.**  
Estudiante de Maestría

**Trabajo de Investigación para optar a Título de Máster en  
producción y aplicación del conocimiento científico en salud  
Énfasis en investigación**

Director Proyecto de Grado  
**Dr. Juan Carlos Villar Centeno, MD, MSc, PhD.**  
Médico especialista en Medicina Interna. Máster en ciencias clínicas de la salud  
y metodología de investigación. Doctor en metodología de la investigación en  
salud.

Asesora temática y metodológica  
**Eliana Yineth Vaquiro Herrera, RN, MSN**  
Enfermera especialista en Cuidado Crítico. Máster en Epidemiología Clínica

Universidad Autónoma de Bucaramanga  
Maestría en Métodos para Producción y Aplicación del Conocimiento  
Científico en Salud  
Bucaramanga Santander  
2021

## INFORMACIÓN GENERAL DEL PROYECTO

<b>Título:</b> Validación de escalas de predicción de mortalidad en pacientes con enfermedad por coronavirus 2019 en el registro de pacientes RESCATE COVID-19
<b>Investigador principal:</b> Custodio Alberto Ruiz Bedoya, Estudiante de Maestría en métodos para la PRACCIS, UNAB FOSCAL. <b>Director del Proyecto:</b> Juan Carlos Villar Centeno. Médico Internista. MSc. PhD. <b>Directora Metodológica:</b> Eliana Yineth Vaquiro Herrera. Enfermera especialista en Cuidado Crítico.
<b>Entidades:</b> Universidad Autónoma de Bucaramanga. Fundación Cardioinfanti, Instituto de Cardiología. Clínica Chicamocha. Clínica Chicamocha <b>Dirección:</b> Calle 157 # 14 - 55 Urbanización el Bosque <b>Teléfono:</b> 6436111 <b>Nit:</b> 890.200.499-9 <b>E-mail:</b> custodioruizbedoya@gmail.com; cruiz494 @unab.edu.co <b>Ciudad:</b> Floridablanca, Santander
<b>Tipo de entida</b> Universidad Pública: Universidad Privada: <input checked="" type="checkbox"/> Entidad Pública: <input type="checkbox"/> ONG: <input type="checkbox"/> <b>Tipo de contribuyente:</b> Entidad de Derecho Público: <input type="checkbox"/> Entidad de Economía Mixta: <input type="checkbox"/>
<b>Lugar de Ejecución del Proyecto:</b> Fundación Cardioinfantil. Clínica Chicamocha. Universidad Autónoma de Bucaramanga <b>Ciudad:</b> Floridablanca <b>Departamento:</b> Santander
<b>Duración del proyecto:</b> 6 meses
<b>Tipo de Proyecto:</b> Investigación Básica: <input type="checkbox"/> Investigación Aplicada: <input checked="" type="checkbox"/> Desarrollo Tecnológico o Experimental: <input type="checkbox"/>
<b>Descriptores/Palabras claves:</b> <b>Programa Nacional de Ciencia y Tecnología sugerido:</b> Ciencia y Tecnología de la salud: <input checked="" type="checkbox"/> Ciencias Básicas: <input type="checkbox"/> Ciencia y Tecnología del Mar: <input type="checkbox"/>



## **AGRADECIMIENTOS**

Al Dr. Juan Carlos Villar por su mentoría a lo largo de mi carrera profesional y por sus enormes enseñanzas en lo personal.

A Miguel Ochoa y a Eliana Vaquiro por no dejarme desfallecer en los momentos difíciles y darme herramientas para finalizar este proyecto.

A mi padre, mis hermanos, mi novia y mi hijo porque le dan sentido todos los esfuerzos.

Al Dr. Víctor Mauricio Herrera por su labor docente, su paciencia y su disposición para hacer posible este estudio.

A los investigadores y colaboradores de los diferentes centros que han participado en la creación de la cohorte RESCATE COVID-19 a quienes les debo el reconocimiento de ser parte fundamental de este estudio.

*Dedicado a mi madre. Mi más grande inspiración y mi mejor compañera entre las letras ...*

# CONTENIDO

<b>1. PLANTEAMIENTO DEL PROBLEMA Y JUSTIFICACIÓN.....</b>	<b>18</b>
<b>2. MARCO TEÓRICO Y ESTADO DEL ARTE.....</b>	<b>22</b>
2.1. Breve recuento histórico .....	22
2.2. Situación actual de la epidemia por COVID-19 en Colombia .....	23
2.3. Diagnóstico, manifestaciones clínicas y curso de la COVID-19 .....	24
2.4. Dificultades para el diagnóstico temprano de pacientes con COVID-19 en Colombia: .....	27
2.5. Escalas de predicción de mortalidad o complicaciones por la COVID-19. ....	27
<b>3. OBJETIVOS .....</b>	<b>29</b>
3.1. Objetivo General.....	29
3.2. Objetivos específicos.....	29
<b>4. DESARROLLO METODOLÓGICO POR OBJETIVOS.....</b>	<b>30</b>
4.1. Diseño general: .....	30
4.2. Para el desarrollo del primer objetivo: Proceso de selección estructurada de las escalas de predicción disponibles .....	30
4.3. Para el desarrollo del segundo objetivo: Caracterización de las escalas potencialmente aplicables .....	33
4.4. Para el desarrollo del tercer objetivo: Validación externa de las escalas..	39
4.4.1. Metodología de la Cohorte RESCATE COVID-19:.....	39
4.4.2. Calidad de los datos:.....	40
4.4.3. Variables y unidades de medida:.....	41
4.4.4. Manejo de los datos faltantes en la cohorte de validación: .....	44
4.4.5. Metodología para la evaluación del desempeño de las escalas en términos de su capacidad de discriminación y calibración en la cohorte local:.....	46
4.4.6. Metodología para la evaluación del desempeño de las escalas en términos de su capacidad de discriminación y calibración en los pacientes sospechosos de la cohorte local: .....	49
<b>5. CONSIDERACIONES ÉTICAS .....</b>	<b>50</b>
<b>6. RESULTADOS.....</b>	<b>52</b>
<b>6.1. Proceso de evaluación del desempeño de la escala Zhang-China al aplicarlo a la cohorte RESCATE COVID-19:.....</b>	<b>52</b>
6.1.1. Pasos 1-3 del proceso de evaluación del desempeño del modelo: Análisis univariado – bivariado y modelo de regresión logística múltiple:.....	52
6.1.2. Paso 4 de la evaluación del desempeño del modelo: Capacidad de discriminación. ....	55
6.1.3. Paso 5 de la evaluación del desempeño del modelo: Calibración. ....	56

6.1.4.	Evaluación del desempeño de la escala Zhang-China según probabilidades calculadas en plataforma digital: .....	57
6.1.5.	Evaluación del desempeño de la escala Zhang-China según los niveles de riesgo obtenidos a partir de los puntajes: .....	59
6.1.6.	Conclusiones del análisis del desempeño de la escala Zhang-China al aplicarla a la cohorte RESCATE COVID-19:.....	62
<b>6.2.</b>	<b>Proceso de evaluación del desempeño de la escala Xie-China al aplicarlo a la cohorte RESCATE COVID-19: .....</b>	<b>62</b>
6.2.1.	Pasos 1-3 del proceso de evaluación del desempeño del modelo: Análisis univariado – bivariado y modelo de regresión logística múltiple:.....	62
6.2.2.	Paso 4 de la evaluación del desempeño del modelo: Capacidad de discriminación. 66	
6.2.3.	Paso 5 de la evaluación del desempeño del modelo: Calibración. ....	67
6.2.4.	Evaluación del desempeño de la escala Xie-China según puntajes calculados: ...	68
6.2.5.	Conclusiones del análisis del desempeño de la escala Xie-China al aplicarla a la cohorte RESCATE COVID-19:.....	68
<b>6.3.</b>	<b>Proceso de evaluación del desempeño de la escala Bello-México al aplicarlo a la cohorte RESCATE COVID-19 .....</b>	<b>69</b>
6.3.1.	Pasos 1-3 del proceso de evaluación del desempeño del modelo: Análisis univariado – bivariado y modelo de regresión logística múltiple:.....	69
6.3.2.	Paso 4 de la evaluación del desempeño del modelo: Capacidad de discriminación. 73	
6.3.3.	Paso 5 de la evaluación del desempeño del modelo: Calibración. ....	74
6.3.4.	Evaluación del desempeño de la escala Bello-México según los puntajes calculados a cada paciente de la cohorte RESCATE COVID-19: .....	75
6.3.5.	Evaluación del desempeño de la escala Bello-México según los niveles de riesgo obtenidos a partir de los puntajes: .....	77
	Tabla 8. Comparativo de áreas bajo la curva ROC con sus respectivos intervalos de confianza del 95% correspondientes a cada uno de los modelos, sus escalas y los niveles de riesgo. .	80
6.3.6.	Conclusiones del análisis del desempeño de la escala Bello-México al aplicarla a la cohorte RESCATE COVID-19:.....	81
<b>6.4.</b>	<b>Proceso de evaluación de los modelos en la población de pacientes con RT-PCR para SARS-COV2 negativa de la cohorte rescate COVID-19. ....</b>	<b>84</b>
6.4.1.	Evaluación del desempeño de la escala Zhang-China en pacientes con RT- PCR negativa para SARS-COV2 de la cohorte RESCATE COVID-19 .....	84
6.4.2.	Evaluación del desempeño de la escala Xie-China en pacientes con RT- PCR negativa para SARS-COV2 de la cohorte RESCATE COVID-19. ....	86
6.4.3.	Evaluación del desempeño de la escala Bello-México en pacientes con PCR negativa. 87	
<b>7.</b>	<b>DISCUSIÓN.....</b>	<b>90</b>
<b>8.</b>	<b>Conclusiones .....</b>	<b>93</b>
<b>9.</b>	<b>CLASIFICACIÓN DE LOS RESULTADOS DEL ESTUDIO SEGÚN COLCIENCIAS: .....</b>	<b>94</b>
	<b>REFERENCIAS BIBLIOGRÁFICAS .....</b>	<b>95</b>
	<b>ANEXOS.....</b>	<b>102</b>

## LISTADO DE TABLAS

Tabla 1. Caracterización de los estudios con los modelos de predicción de mortalidad seleccionados para su validación externa.....	36
Tabla 2. Caracterización de las variables independientes incluidas en cada uno de los modelos.....	42
Tabla 3. Caracterización de los predictores propuestos por el modelo Zhang-China y evaluación de su asociación con mortalidad intrahospitalaria en la cohorte RESCATE COVID-19: .....	54
Tabla 4. Modelo de regresión logística múltiple (variables indicadoras) entre niveles de riesgo obtenidos a partir de los puntajes calculados por la escala Zhang-China y mortalidad intrahospitalaria en la cohorte RESCATE-COVID-19. .	60
Tabla 5. Caracterización de los predictores propuestos por el modelo Xie-China y evaluación de su asociación con mortalidad intrahospitalaria en la cohorte RESCATE COVID-19 .....	64
Tabla 6. Caracterización de los predictores propuestos por el modelo Bello-México y evaluación de su asociación con mortalidad intrahospitalaria en la cohorte RESCATE COVID-19. ....	71
Tabla 7. Modelo de regresión logística múltiple (variables indicadoras) entre niveles de riesgo obtenidos a partir de los puntajes calculados por la escala Bello-México y mortalidad intrahospitalaria en la cohorte RESCATE-COVID-19.....	78
Tabla 8. Evaluación de los puntos de corte de la escala Bello-México para su posible aplicación clínica en escenarios de primer contacto médico. ....	82
Tabla 9. Asociación entre predictores del modelo Bello-México y mortalidad intrahospitalaria en el grupo de paciente con RT-PCR para SARS-COV2 negativa de la cohorte RESCATE COVID-19.....	88



## LISTADO DE FIGURAS

Figura 1. Proceso de selección de las escalas de predicción de mortalidad con potencial para validación local en pacientes de la cohorte RESCATE COVID-19:	32
Figura 2. Datos faltantes de cada una de las variables de las escalas evaluadas:	45
Figura 3. Proceso para la evaluación del desempeño de las escalas seleccionadas según su capacidad de discriminación y su calibración.....	46
Figura 4. Gráfico de bosque del análisis multivariado de la escala Zhang-China en la cohorte RESCATE COVID-19.....	55
Figura 5. Curva ROC del modelo Zhang-China para predicción de mortalidad intrahospitalaria en la cohorte de validación externa: .....	56
Figura 6. Gráfico de calibración entre probabilidades esperadas por el modelo Zhang-China y las observadas en la cohorte de validación según deciles de probabilidad esperada: .....	57
Figura 7. Curva ROC del modelo Zhang-China para predicción de mortalidad intrahospitalaria en la cohorte de validación externa según los puntajes de la escala: .....	58
Figura 8. Gráfico de calibración entre probabilidades esperadas por los puntajes de la escala de Zhang-China y la probabilidad observada en la cohorte de validación:.....	59
Figura 9. Curva ROC del modelo Zhang-China para predicción de mortalidad intrahospitalaria en la cohorte de validación externa según los niveles de riesgo:	61
Figura 10. Gráfico de bosque del análisis multivariado de la escala Xie-China en la cohorte RESCATE COVID-19.....	65

Figura 11. Curva ROC del modelo Xie-China para predicción de mortalidad intrahospitalaria en la cohorte de validación externa: .....	66
Figura 12. Gráfico de calibración entre probabilidades esperadas por el modelo Xie-China y las .....	67
Figura 13. Gráfico de bosque del análisis multivariado de la escala Bello-México en la cohorte RESCATE COVID-19 .....	72
Figura 14. Curva ROC del modelo Bello-México para predicción de mortalidad intrahospitalaria en la cohorte de validación externa .....	73
Figura 15. Gráfico de calibración entre probabilidades esperadas por el modelo Bello-México y las observadas en la cohorte de validación según deciles de probabilidad esperada .....	74
Figura 16. Curva ROC del modelo Bello-México para predicción de mortalidad intrahospitalaria en la cohorte de validación externa según los puntajes de la escala: .....	76
Figura 17. Gráfico de calibración entre probabilidades esperadas por los puntajes de la escala de .....	77
Figura 18. Curva ROC de la escala Bello-México para predicción de mortalidad intrahospitalaria en la cohorte de validación externa según los niveles de riesgo: .....	79
Figura 19. Gráfico de calibración entre probabilidades esperadas por los niveles de riesgo puntajes de la escala de Bello-México y la probabilidad observada en la cohorte de validación .....	79
Figura 20. Curva ROC del modelo Zhang-CHina para predicción de mortalidad intrahospitalaria en la cohorte de validación externa de pacientes con RT-PCR para SARS-COV2 negativa de la cohorte RESCATE COVID-19.....	85

Figura 21. Curva ROC del modelo Xie-China para predicción de mortalidad intrahospitalaria en la cohorte de validación externa de pacientes con RT-PCR negativa de la cohorte RESCATE COVID-19. ....	86
Figura 22. Gráfico de calibración entre probabilidades esperadas por el modelo Xie-China y las observadas en la cohorte de pacientes con RT-PCR negativa de la cohorte RESCATE COVID-19.....	87
Figura 23. Curva ROC del modelo Bello-México para predicción de mortalidad intrahospitalaria en la cohorte de validación externa de pacientes con RT-PCR negativa de la cohorte RESCATE COVID-19. ....	89
Figura 24. Gráfico de calibración entre probabilidades esperadas por el modelo Bello-México y las observadas en la cohorte de pacientes con RT-PCR negativa de la cohorte RESCATE COVID-19.....	89

## LISTADO DE ANEXOS

Anexo 1: Carta de aval de la Clínica Chicamocha para proyecto RESCATE-COVID-19. ....	102
Anexo 2. Proceso de selección de los modelos de predicción documentados por la revisión sistemática realizada por Wynants .....	104
Anexo 3: Proceso de caracterización variables .....	147
Anexo 4: Protocolo Proyecto RESCATE COVID-19: Respuesta de Conocimiento para la Atenuación de la Epidemia por COVID-19.....	149
Anexo 5: Declaración de confidencialidad firmada por Investigador principal de la Clínica Chicamocha. ....	150
Anexo 6: Certificado de buenas prácticas clínicas de investigador principal Clínica Chicamocha .....	152
Anexo 7: Caracterización de las variables continuas incluidas en los modelos:..	153
Anexo 8: Boletín 9. REGISTRO DE CASOS PROBABLES O CONFIRMADOS DE COVID-19. ESTUDIO RESCATE COVID-1 .....	161
Anexo 9:...Análisis de desempeño del modelo Zhang-China al retirar la variable PCR y puntos de corte para la toma de decisiones clínicas. ....	163
Anexo 10: Modelos de regresión logística múltiple obtenidos en STATA 14 .....	167
Anexo 11: Evaluación del desempeño de los modelos chinos en el grupo de pacientes con RT-PCTR para SARS-COV2 negativa de la cohorte RESCATE COVID 19. ....	179

## RESÚMEN

**Título:** Validación de escalas de predicción de mortalidad en pacientes con enfermedad por coronavirus 2019 en el registro de pacientes RESCATE COVID-19.

**Introducción:** La pandemia causada por la enfermedad por coronavirus 2019 (COVID-19) ha revelado la necesidad de optimizar los recursos en salud. Las herramientas que permiten diferenciar entre poblaciones con diferentes niveles de riesgo como las escalas o modelos de predicción, adquieren gran valor en este nuevo escenario. A la fecha se encuentran disponibles más de 50 escalas de pronóstico en COVID-19 pero la gran mayoría no han tenido procesos de validación externa y aún tienen un bajo potencial de generalización.

**Objetivo:** Evaluar el desempeño de al menos tres escalas de predicción de mortalidad, en términos de su capacidad de discriminación, calibración y potencial generalización a la práctica clínica, al aplicarlas en una cohorte nacional de pacientes con infección por SARS-COV2 (RESCATE COVID-19).

**Metodología:** A partir de una revisión sistemática viva de la literatura que contiene todas las escalas publicadas hasta mayo de 2020 y siguiendo un proceso estructurado, se eligieron las tres con mayor potencial de aplicación local. En cada una se realizó un proceso de caracterización de sus predictores (distribución, análisis bivariado y multivariado con regresión logística múltiple), evaluación de su capacidad de discriminación (área bajo las curvas ROC) y calibración (de manera gráfica y estadística: prueba de bondad de ajuste de Hosmer-Lemeshow, calibración en promedio "CITL", pendiente de calibración). Además, se analizó el desempeño de los modelos utilizando sus puntajes o niveles de riesgo como variables independientes.

**Resultados:** Tras el proceso de selección resultaron dos estudios orientales (Zhang-China, Xie-China) y uno latinoamericano (Bello-México) como candidatos para su validación local. Los predictores propuestos por los modelos son en su mayoría variables sociodemográficas, antecedentes de comorbilidades, laboratorios generales y algunas derivadas de la atención clínica. Los modelos fueron evaluados en 1342 pacientes con diagnóstico confirmado de COVID-19 que hacen parte de la cohorte RESCATE COVID-19 (mortalidad 12.7%). Esta incluye pacientes atendidos en seis hospitales ubicados 3 diferentes ciudades de Colombia.

De los 7 predictores por la escala de Zhang-China solo 2 resultaron con OR significativo tras ser ajustados en un modelo de regresión logística multivariada (edad: OR 1.03; IC95% 1.01-1.06 y proteína C reactiva: OR: 1.07; IC95% 1.03 – 1.10). Su capacidad de discriminación resultó apenas aceptable (AUC: 0.78) y no demostró estar calibrada estadísticamente (HL  $p=0.002$ ) ni en su evaluación gráfica. Al utilizar sus puntajes (calculados con una herramienta disponible en línea) y niveles de riesgo (según los puntos de corte propuestos) como variables predictoras, la escala tiene menor capacidad de discriminación (AUC: 0.73 y 0.67), persiste sin calibración estadística (HL  $p<0.05$ ) y su evaluación gráfica empeora. Por su parte todos los predictores propuestos por la escala de Xie-China mostraron una asociación significativa con la mortalidad intrahospitalaria. Su capacidad de discriminación fue buena (AUC: 0.81) y sus predicciones se ajustaron a las observaciones de la cohorte (HL  $p=0.22$ ). Sin embargo, no fue posible realizar un análisis de la escala de puntos basada en el nomograma que proponen sus autores. Se encontró que se trata de un sistema complejo de puntuación, con alta vulnerabilidad a errores de medición, demandante de tiempo y con poca probabilidad de aplicación a la práctica clínica diaria local. Cuatro de los 9 predictores de la escala Bello-México tuvieron medidas de asociación ajustadas significativas. Su capacidad de discriminación fue buena (AUC:0.80) y demostró calibración estadística (HL  $p=0.32$ ) y gráfica. El sistema de puntuación como escala y los niveles de riesgo resultaron con mejor desempeño que el modelo (AUC: 0.87) y con muy buena calibración (HL  $p>0.05$ ). Tras la evaluación de los puntos de corte para su aplicación clínica se eligió el umbral de 4 puntos en la escala para dividir entre alto y bajo riesgo (VPN: 98%, Sensibilidad del 90%)

### **Conclusiones:**

La escala de Bello-México resultó tener muy buen desempeño al aplicarse a los pacientes con diagnóstico de COVID-19 de la cohorte RESCATE-COVID19. Su sistema de puntuación sencillo, con buena capacidad de discriminación entre pacientes de alto y bajo riesgo, ofrece predicciones calibradas de las probabilidades de mortalidad intrahospitalaria. Teniendo en cuenta la naturaleza de sus predictores se considera útil en escenarios de atención temprana de pacientes con diagnóstico o sospecha de la enfermedad. Con un punto de corte de 4 puntos la escala resulta útil como estrategia de tamización del pronóstico ya que puntajes menores tienen muy baja tasa de falsos negativos. Su uso puede informar mejor (aunque no reemplazar) la decisión clínica y optimizar la utilización de recursos. Es necesario evaluar su desempeño en otras poblaciones antes de generalizar su uso.

La escala de Zhang-China no logró un buen desempeño al aplicarse a la cohorte. El modelo de Xie-China, aunque prometedor por su desempeño, presenta una

herramienta para cálculo de puntajes compleja, vulnerable de error de medición y demandante de recursos, que no la hace apta para su aplicación clínica local.

**Palabras Clave:** Validación. Modelos de predicción. Escalas de predicción. Pronóstico. Validez externa. Discriminación. Calibración. Enfermedad por coronavirus 2019. COVID-19

## SUMMARY

**Title:** Validation of mortality prediction scales for patients with COVID-19 in the *RESCATE COVID-19* cohort.

**Introduction:** The pandemic caused by COVID-19 has exposed the need to optimize healthcare resources. Tools that facilitate the distinction between populations with varying risk levels, such as risk assessment scales or predictive models are therefore of vital importance. To date, there are more than 50 prognostic scales available for COVID-19 but the vast majority of them lack proper external validation and overall potential to be generalizable.

**Objective:** Evaluate the performance of at least three mortality prediction scales, their discriminatory ability, calibration and clinical generalizability by applying them to a Colombian national cohort of patients infected with SARS-CoV-2 (Project RESCATE COVID-19)

**Methods:** After a systematic and comprehensive literature review of all mortality risk scales published before May 2020, followed by a structured selection process, the top three scales with the highest potential for applicability at the local level were chosen. For each one, all predictors were characterized by analyzing their distribution and performing bivariate and multivariate analysis, with multiple logistic regressions. The discriminating capacity of all models was evaluated by estimating area under ROC curves and their calibration was assessed both graphically and statistically using the Hosmer-Lemeshow test for goodness-of-fit, the calibration slope and the calibration in the large (CITL). Moreover, analyses of model performance were conducted for all three models using their risk levels as independent variables.

**Results:** The selection process resulted in two Asian (Zhang-China, Xie-China) and one Latinoamerican (Bello-México) studies as the top candidates for evaluation at the local level. The vast majority of predictors proposed by these models were sociodemographic variables, history of comorbidities, general laboratory results and some variables derived from clinical care. All three models were evaluated in 1342 patients with confirmed diagnostic of COVID-19 from the patient cohort of the Colombian project RESCATE COVID-19, which has an overall mortality rate of 12.7%. This cohort includes patients distributed across six hospitals located in 3 different cities in Colombia. From the 7 predictors in the Zhang-China scale, only 2 resulted with a significant OR value after being adjusted by a multivariate logistic



regression model (age with OR = 1.03; CI 95% 1.01-1.06 and C Reactive Protein with OR = 1.07; IC 95% 1.03 – 1.10). The discriminatory capacity of this scale proved to be only acceptable (AUC = 0.78), it showed insufficient calibration, both statistically (HL p=0.002) and graphically. When using scores calculated by an online tool and risk levels based on the proposed thresholds as predictive variables, the discriminatory capacity of this scale increases (AUC = 0.73 and 0.67), but it maintains poor statistical calibration (HL p < 0.05), and the graphical assessment of calibration worsens. On the other hand, all predictors proposed by the Xie-China scale proved to have a significant association with intrahospital mortality. Their discriminatory capacity proved to be high (AUC = 0.81) and its predictions match the empirical data from our cohort (HL p = 0.22). However, it was not possible to perform an analysis of this scale based on the scoring system proposed by the authors. The scoring system showed to be highly complex, extremely sensitive to small measuring errors, highly time-consuming and with very low likelihood of proving useful in a daily clinical setting. Finally, 4 out of 9 predictors in the Bello-México scale showed significant adjusted association metrics. Their discriminatory capacity was high (AUC = 0.80) and they showed proper statistical (HL p = 0.32) and graphical calibration assessments. The scoring system and the risk levels showed even better performance than the model (AUC = 0.87) and resulted in good calibration (HL p>0.05). After the evaluation of the thresholds for its clinical applicability, a threshold of 4 points in the scale was chosen to properly discriminate low and high-risk patients (VPN: 98%, 90% sensitivity)

**Conclusions:** From the top three mortality risk scales, the Bello-México showed the highest potential, with high performance when applied to patients diagnosed with COVID-19 in the RESCATE COVID-19 cohort and a simple scoring system with good discriminatory capacity for high and low-risk patients, offering calibrated predictions of the intrahospital mortality probability. Considering the nature of its predictors, this scale is deemed useful for early assessment in patients with a diagnosis for COVID-19 or suspicion of the disease. When using a threshold of 4 points, this scale is considered to be useful for prognostics given that lower scores tend to be associated with very low rate of false negatives. Applying this scale could better inform, but not replace, the clinical decision-making process to optimize resources. The scale Zhang-China, on the other hand, did not show good performance when applied to the patient cohort of the project RESCATE COVID-19. Finally, the Xie-China model, however promising in its performance, showed a very complex scoring system that is not only demanding in resources but would also make this risk scale highly vulnerable to human error, making it overall less likely to be applicable at local hospitals.

**Key words:** Validation. Mortality. Predictive models. Predictive scales. Pronostics. External validation. Discrimination. Calibration. COVID-19

## 1. PLANTEAMIENTO DEL PROBLEMA Y JUSTIFICACIÓN

Iniciando el mes de Abril del 2021 en Colombia se han reportado casi dos millones y medio de casos confirmados de infección por SARS-CoV2 y más de 30,300 defunciones a causa de alguna de las complicaciones de la COVID-19 (1). Según el centro de recursos para coronavirus de la universidad de John Hopkins actualmente nuestro país ocupa el decimoprimer lugar en reporte de casos positivos y defunciones a nivel mundial (2). En su último boletín (corte el 24 de Marzo) la Organización Panamericana de la Salud (OPS) una tasa de incidencia de 4.671 casos por 100.000 habitantes y una de mortalidad de 1.238 muertes por millón de habitantes(3).

A pesar de la inversión de recursos dirigidos a mejorar la capacidad instalada para el diagnóstico y manejo de los casos, la pandemia por COVID-19 sigue siendo hoy una amenaza para el sistema de salud nacional. Durante el pico epidémico entre julio y agosto del año pasado la ocupación de camas de unidades de cuidados intensivos (UCI) permaneció cercana al 100% pese al incremento relativo en la oferta de un 91% (de 6,000 a 10,000 aproximadamente). Hoy en Colombia se ha logrado disponer de 11,000 camas de UCI de las cuales se ha reportado una ocupación superior aproximada del 60% durante el último mes (4). Teniendo en cuenta que entre el 15- 22% de los casos ha requerido manejo intrahospitalario y entre el 1-3% de una UCI, la posibilidad de un segundo pico epidémico sigue poniendo en riesgo de colapso al sistema.

A la fecha se han procesado más de nueve millones de pruebas diagnósticas tipo RT-PCR (del inglés: *reverse transcription polymerase chain reaction*) y más de dos millones de pruebas de antígeno viral para SARS-CoV2 en Colombia (1). Si bien con el tiempo transcurrido de pandemia ha aumentado la disponibilidad y cantidad de pruebas procesadas, la demora en los tiempos de reporte de resultados es aún constante. El tiempo de demora promedio en entre notificación y diagnóstico en el último mes ha sido de 4.07 días pero en algunas poblaciones como en indígenas y afrodescendientes es de 11 días (5). Esta dificultad para una rápida clasificación de los pacientes con casos sospechosos genera un aumento en la utilización de servicios de salud. No es infrecuente que los pacientes los casos sospechosos requieran manejo intrahospitalario y su estancia transcurra sin conocer el resultado de la prueba.

Un componente adicional en la problemática de la atención nacional a la pandemia es la disponibilidad y utilización de los medicamentos con potencial efecto en el

manejo de la COVID-19. La ausencia antivirales con posible efectividad disponibles en nuestra región y el limitado arsenal de tratamientos con posibles beneficios como la dexametasona (6), colchicina (7), o los anticoagulantes (8) hacen necesario el uso racional de los mismos.

En este complejo escenario de una enfermedad con muy altas tasas de incidencia y demandas de recursos de salud (pruebas diagnósticas, servicios hospitalarios y medicamentos) en un país con escasa posibilidad de aumento en la oferta de servicios de salud es necesario generar estrategias que permitan una optimización en la utilización de los recursos. La caracterización clínica de los pacientes con mayor riesgo de desarrollar complicaciones derivadas de la enfermedad en estadios tempranos es una de las estrategias que puede ayudar a cumplir este objetivo. Motivados por la misma necesidad los países que primero se vieron enfrentados a la epidemia ya han desarrollado múltiples escalas de predicción de pronóstico. En Colombia no se conocen hasta el momento desarrollo de alguno de estos modelos y existen pocas publicaciones de algunos desarrollados en latino américa (9).

En una revisión sistemática viva de la literatura se documentaron cincuenta modelos de pronóstico de mortalidad y de deterioro clínico publicados hasta el momento en el mundo (10). Los autores manifiestan su preocupación por la potencial generalización de los modelos a pesar de que la mayoría mostró una muy buena capacidad de predicción de mortalidad o curso clínico desfavorable. Advierten que la evaluación de calidad fue en términos generales baja, el riesgo de sesgos alto y validación externa muy infrecuente. Muy pocos de los estudios documentados incluyeron información de población latinoamericana.

La mayoría de los modelos utilizan variables sociodemográficas, clínicas, de imágenes y de resultados de laboratorio documentados al inicio de la atención médica para predecir complicaciones de la enfermedad o mortalidad. Se presentan a manera de nomograma, árboles de decisión, sistema de puntaje o herramientas en línea como calculadoras digitales. Otros han sido desarrollados con herramientas sofisticadas de aprendizaje por computadoras o “deep learning” como estrategia de modelamiento y se presentan como herramientas digitales disponibles en línea.

A nivel nacional se han realizado esfuerzos por documentar modelos de predicción que tengan potencial aplicabilidad local. La unidad de evidencia de la Universidad de Antioquia realizó una revisión sistemática de la literatura en la que se encontraron 33 modelos publicados (11). Sus autores catalogaron los estudios en su mayoría como de baja calidad, publicados sin previa revisión de pares y con hallazgos

similares a los publicados por Wynants y colaboradores. En el documento se da una recomendación en contra de la utilización de las escalas a nivel nacional ante la falta de validación externa.

Actualmente se está llevando a cabo un proyecto llamado RESCATE COVID-19 (Respuesta de Conocimiento para la Atenuación de la Epidemia por COVID-19) en seis centros hospitalarios de tercer nivel de complejidad de las ciudades de Bogotá, Medellín y Bucaramanga. Dentro de sus objetivos se encuentra el desarrollar, como contribución a la mitigación del impacto de la pandemia, una plataforma dinámica de información asistencial que involucre resultados clínicos de los casos sospechosos de COVID-19. Para esto se ha ensamblado una cohorte prospectiva de casos sospechosos con las características de base sociodemográficas, de atención clínica y los resultados de la atención los centros asistenciales vinculados. El último boletín con la caracterización sociodemográfica, clínica y resultados de la atención de los primeros 1342 pacientes se encuentra disponible en línea en la página de la Fundación Cardioinfantil:

<https://cardioinfantil.org/investigaciones/proyectos/rescate/>

La cohorte cuenta con información de la mayoría de los factores que en otros estudios han mostrado una posible asociación con un peor pronóstico de complicaciones graves y mortalidad por COVID-19. Se dispone de información sociodemográfica, de comorbilidades (enfermedad cardiovascular, obesidad, falla renal), alteraciones bioquímicas (LDH, PCR, Dímero D, Ferritina, Linfocitos), radiológicas (número de opacidades, presencia de vidrio esmerilado, derrame pleural) y gasométricas (PaFiO<sub>2</sub>). De igual manera se dispone de desenlaces como mortalidad y complicaciones de la enfermedad. Por su diseño longitudinal, la inclusión de pacientes con cierta heterogeneidad (atendidos en cinco centros hospitalarios de tres departamentos), la constitución de sus variables y la posibilidad de evaluar incidencia de los desenlaces se considera un escenario para la realización de estudios de validación local de los modelos pronósticos disponibles.

Como estudiante de maestría en métodos para la PRACCIS de la Universidad Autónoma de Bucaramanga y médico internista en ejercicio clínico durante la pandemia he participado en la vinculación de la Clínica Chicamocha como centro asistencial participante del proyecto RESCATE COVID-19. Tras presentar el proyecto a los comités de investigación y de ética de la clínica se me otorgó el aval para realizar el proyecto como investigador principal de la sede junto con un grupo de seis coinvestigadores (Ver Anexo 1). Además de las responsabilidades que se

me asignaron también se me otorgó la posibilidad de utilizar la información de la cohorte como soporte para la generación de nuevos estudios.

Dada la necesidad de generar herramientas que ayuden en la adecuada disposición y utilización de los servicios de salud en el contexto de la pandemia, se propone el desarrollo de un estudio de validación de escalas de pronóstico a nivel local. Para esto se realizó proceso de selección de las escalas disponibles en la literatura que pudieran aplicarse en la cohorte RESCATE COVID-19 y posteriormente se evaluó su desempeño. La intención fue documentar al menos una escala que cuente con buena capacidad para discriminar entre pacientes con bajo y alto riesgo de fallecer por la enfermedad y que pueda aplicarse de manera sencilla en la práctica clínica diaria.

## 2. MARCO TEÓRICO Y ESTADO DEL ARTE

### 2.1. Breve recuento histórico

El primer brote de casos de neumonía de causa desconocida ocurrió en cinco pacientes que consultaron a diferentes hospitales de la provincia de Hubei en Wuhan (China) entre el 19 y el 29 de diciembre de 2019 (12). Cuatro de ellos requirieron hospitalización por síndrome de dificultad respiratoria del adulto (SDRA) y uno de ellos falleció por falla ventilatoria (13). El análisis inicial de los casos sugirió al mercado de alimentos de la provincia como un lugar común entre los pacientes por lo que se consideró potencialmente relacionado con el contagio. El mercado fue clausurado por las autoridades a principios de enero de 2020 para iniciar las investigaciones sobre los posibles orígenes del patógeno (14). Se identificó un nuevo coronavirus como el posible agente causal de los casos al que inicialmente se le dio el nombre de 2019-nCoV.

El 12 de enero del año 2020 se habían reportado 41 casos de pacientes hospitalizados por neumonía y SDRA con aislamiento del virus en mucosas respiratorias y solamente uno de ellos con alta carga de comorbilidades había fallecido (15). A las manifestaciones clínicas de la infección se le dio el nombre de COVID-19 (por sus siglas en inglés: *coronavirus disease 2019*). En ese momento se consideró que el riesgo de propagación era bajo por su aparente relación con el consumo de carnes de animales procedente del mercado y la falta de evidencia de contagio entre personas ni hacia el personal de salud que atendió los primeros casos (16). Una semana después 51 provincias de China reportaron 571 casos y en Hubei había 95 pacientes en estado crítico y 17 fallecidos por causa de la nueva infección.

Desde entonces el crecimiento de la epidemia fue exponencial logrando expandirse a toda china y a 19 países donde se reportaban para finales de enero 9,692 y 90 casos respectivamente (17). En ese momento la Organización Mundial de la Salud (OMS) declaró un estado de emergencia en salud pública de importancia internacional. El primer caso en el nuestro continente correspondió a un paciente de 35 años que consultó en Washington (Estado Unidos) con síntomas respiratorios tras un viaje familiar a Wuhan y fue confirmado con COVID-19 en Enero 19 (18).

A pesar de diferentes estrategias gubernamentales instauradas en los primeros países afectados que incluyeron cuarentenas y cierres de aeropuertos, para el 11 de Marzo del año 2020 la OMS decretó el estado de pandemia (14). En el mismo documento se designa al agente causante de la pandemia como SARS-CoV2 (del inglés: *Severe Acute Respiratory Syndrome – coronavirus type 2*) como se denominaría desde entonces. Para ese momento 114 países habían reportado un total de 118,000 casos y 4,221 fallecidos. Colombia hacía ya parte de los países con reportes de casos ya que el primer reporte fue realizado el 6 de marzo de ese año, en una paciente de 19 años de edad (19).

En octubre del 2020, el centro de recursos para coronavirus de la Universidad Johns Hopkins registró más de 45 millones de casos reportados en el mundo y más de 1,200,000 muertes a causa de la COVID-19 (2). Estados Unidos, India y Brasil fueron para esa fecha los países con mayor número de casos confirmados y de muertes (relación de casos/muertes: >9 millones/>229.000, >8 millones/>150.000 y >5 /120,000 respectivamente). En la región de centro y sur américa Brasil, Argentina (1,143,000) y Colombia (1,048,000) eran los países con mayor número de casos reportados. El comportamiento de la epidemia en el tiempo ha sido variable en los países. Durante los últimos días los 10 países con más número de casos confirmados son Estados Unidos, Brasil, India, Polonia, Turquía, Italia, Alemania, Ucrania, Francia y República Checa (20).

## **2.2. Situación actual de la epidemia por COVID-19 en Colombia**

El Instituto Nacional de Salud (INS) ha generado una plataforma dinámica con actualización diaria del análisis de los datos sobre el comportamiento de la epidemia en Colombia (1). Para el momento de la redacción de este documento se registran 2,428,048 casos confirmados (2,300,887 recuperados y 55,062 activos) y 63,777 personas fallecidas desde el inicio de la pandemia. Los departamentos con mayor reporte de casos han sido Bogotá (689,038), Antioquia (377,291) y Valle del Cauca (209,106). La distribución de los casos por edad y sexo de los casos activos tiene incrementos en los grupos de mayor edad en comparación a la totalidad de los casos reportados. Actualmente se reporta una proporción de 48.48% en hombres, en personas de 60 años y más de 22% y el grupo con mayor proporción de casos activos es el de 20 a 39 años con el 37,2% (3,21).

Se ha documentado una letalidad en los casos positivos del 3,1% de los cuales el 63,7% son de género masculino. El 76,1% de los fallecidos están en el grupo de edad de mayores de 60 años. Los grupos etarios con mayor número de muertes son los de 70-79 años con 16.942, seguido por los de 80-89 años en donde se encuentran las mayores tasas de mortalidad en ambos sexos. La tasa de mortalidad nacional es de 1,238.7 muertes por cada 1,000,000 de habitantes (20).

Respecto a la utilización de los servicios de salud, durante la epidemia se ha mantenido la proporción del 20-30% de requerimiento de una cama de hospitalización y 2-5% de una Unidad de Cuidados Intensivos (UCI). Entre el 70-80% de los casos han permanecido en casa. La clasificación de los casos según la severidad también ha mantenido una proporcionalidad estable en el tiempo con 92-96% de casos en estadios leves (20).

Respecto a la oportunidad en el diagnóstico y la atención al manejo de los casos se documenta que el promedio de tiempo transcurrido entre el inicio de los síntomas y la notificación es de 4.1 días. Según el último reporte de la OPS en nuestro país el 80% de los resultados de las pruebas moleculares se obtiene en los primeros 7 días después de su reporte.

Durante el último mes la proporción de casos activos ha mostrado una tendencia oscilante con aumento en los últimos días de aproximadamente 0.3%.

### **2.3. Diagnóstico, manifestaciones clínicas y curso de la COVID-19**

El método diagnóstico aceptado como estándar de referencia consiste en la detección de ARN (ácido ribonucleico) viral mediante la RT-PCR (del inglés: *reverse transcription polymerase chain reaction*) en una muestra del tracto respiratorio superior con una técnica de hisopado nasofaríngeo (22). La posible utilidad de las pruebas serológicas en sangre fue para escenarios en los que al menos hayan transcurrido dos semanas después del inicio de los síntomas. Inicialmente algunos estudios reportaron rendimientos de diagnóstico aceptables con una sensibilidad del 85% y una especificidad del 90%. Sin embargo con su generalización se ha demostrado la alta tasa de clasificaciones incorrectas que ha generado un aumento del consumo de recursos en salud, por lo que hoy poco se recomiendan (23). El único escenario de probable utilidad es al aplicarlo a pacientes con alta sospecha clínica y un resultado de RT-PCR negativo (24). El uso combinado de las pruebas puede ser de utilidad clínica dependiendo del momento de la infección (25). Las



pruebas de antígeno han demostrado utilidad para el diagnóstico de infección aguda o temprana con desempeño similar al de la PCR. Sin embargo, el escenario de mejor rendimiento es en pacientes sintomáticos después de 24-48 horas de inicio de los síntomas. Su capacidad diagnóstica tiene mejores valores de especificidad y de valores predictivos negativos por lo que un resultado negativo no descarta la infección y ante la sospecha clínica se hace necesario realizar RT-PCR (26).

El espectro de manifestaciones clínicas de la infección por SARS-CoV2 es amplio e incluye desde pacientes asintomáticos hasta pacientes con manifestaciones severas muy tempranas. Los síntomas más frecuentes de la COVID-19 son la fiebre, cefalea, tos, disnea y en ocasiones diarrea, náuseas y vómito (27). La neumonía es la complicación más grave de la enfermedad y la principal causa de hospitalización o ingreso a UCI (28).

Los hallazgos imagenológicos relacionados con la neumonía se han caracterizado con mayor claridad con imágenes tomográficas (TAC: tomografía axial computarizada de tórax) que además han demostrado una mayor capacidad de discriminación diagnóstica al compararse con las radiografías de tórax (29,30). Además de permitir caracterizar diferentes estadios de severidad de la enfermedad se ha propuesto como una herramienta complementaria de diagnóstico al agregar sensibilidad y especificidad al diagnóstico con pruebas moleculares (31). Sin embargo, aún no se recomienda como una estrategia diagnóstica única para confirmar infección por SARS-COV2. Las principales características evidenciadas en la TAC de pacientes con neumonía por SARS-CoV2 son opacidades en vidrio deslustrado, seguido por evidencia de consolidaciones generalmente periféricas, bilaterales y basales, encontrándose mayor índice de mortalidad en pacientes con dichos hallazgos imagenológicos (32).

El síndrome de dificultad respiratoria del adulto es una entidad clínica que requiere caracterización radiológica, clínica y gasométrica. Se considera una complicación grave de múltiples patologías respiratorias que amerita manejo en la UCI con ventilación mecánica invasiva, estrategias específicas de ventilación y relajación neuromuscular (33). Se considera una de las manifestaciones graves de la COVID-19 que consume alta cantidad de recursos de los servicios de salud ya genera una alta mortalidad (34). Los niveles de saturación de oxígeno han mostrado ser uno de los principales factores predictores de mortalidad, rangos entre 80 y 85 % de saturación basal se han relacionado en gran medida con la mortalidad en los pacientes hospitalizados (35).

Otras de las complicaciones descritas de la COVID-19 son la falla renal y la injuria miocárdica aguda (36,37). Las dos entidades parecen explicarse por una suma de factores del efecto citotóxico directo potencial del virus y de la inflamación desencadenada por una respuesta inflamatoria sistémica severa que algunos autores han descrito como “sepsis viral” (38). Se ha demostrado que ambas condiciones son un factor predictor de mortalidad en los pacientes con COVID-19 (39).

Información derivada de estudios observacionales ha permitido caracterizar algunas de las variables sociodemográficas y clínicas que se han asociado a un peor pronóstico de los pacientes con COVID-19. Los antecedentes de enfermedades cardiovasculares o respiratorias crónicas como asma y EPOC, los eventos coronarios o cerebrovasculares, la hipertensión arterial y las condiciones que generan inmunosupresión han sido las documentadas de manera consistentes en diferentes de estudios (40–43).

Los hallazgos de laboratorio que han demostrado una asociación con manifestaciones severas de la enfermedad son la elevación de la LDH, el dímero D, la procalcitonina, creatinina, PCR y la alteración de los tiempos de coagulación (42,44–46). En cuanto a los hallazgos del hemograma que han mostrado una relación con mayor riesgo de mortalidad se encuentran la linfocitopenia, trombocitopenia y la neutrofilia (11,41).

De igual manera se ha estudiado la relación entre factores demográficos (edad, sexo entre otros) y la probabilidad de presentación severa de la enfermedad. Información de estudios realizados en etapas tempranas de la pandemia evidenció una relación de proporcionalidad directa entre la edad y el riesgo de complicaciones severas y mortalidad (aumento del 32% por cada 10 años de diferencia en edad) (47).

Ha llamado la atención la documentación de un posible aumento en la frecuencia de manifestaciones severas de la enfermedad en varones respecto a mujeres. Sin embargo, es posible que existan factores de confusión que modifiquen esta posible relación como la mayor carga de factores de riesgo cardiovascular en los hombres (48). Un ejemplo de esto es la alta prevalencia de obesidad en este género, condición parece favorecer la presentación grave de la COVID-19 por la mayor expresión de enzima convertidora de angiotensina 2 en el tejido adiposo, que a su vez funciona como receptor del virus durante su replicación (49,50).

#### **2.4. Dificultades para el diagnóstico temprano de pacientes con COVID-19 en Colombia:**

Desde los inicios de la pandemia en Colombia se han presentado múltiples situaciones que han redundado en demoras en la oportunidad para el diagnóstico de la enfermedad. Al principio no se disponía de suficientes reactivos para la realización de las pruebas diagnósticas, tampoco de la infraestructura técnica en muchas de las ciudades secundarias y terciarias. Aun hoy el proceso diagnóstico se encuentra centralizado en las ciudades capitales y sesgado por las vinculaciones contractuales propias de nuestro modelo de salud (51). Informes de la OPS en los primeros meses de la pandemia y hasta finales del 2020 reportaban demoras promedio al diagnóstico de 10-15 días en la población general y mayores en poblaciones vulnerables (52). Actualmente el 80% de las pruebas se reportan en los 7 días siguientes a la toma de la muestra y el promedio nacional se encuentra en 4.1 días.

Como consecuencia de este retraso en el diagnóstico que se suma al error implícito de la prueba y al generado por tomas inadecuadas de muestras de hisopado se aumenta la utilización de servicios de salud. Además, el manejo de los pacientes con casos sospechosos durante los primeros estadios de la enfermedad (o incluso durante todo el curso de la misma) puede ser muy similar independientemente del resultado final de su prueba.

#### **2.5. Escalas de predicción de mortalidad o complicaciones por la COVID-19.**

Una de las estrategias que puede favorecer el buen uso de los recursos disponibles para la atención de la pandemia es la generación de herramientas que ayuden a discriminar entre los pacientes con alto y bajo riesgo de presentar complicaciones (53,54). La Dra. Wynants y colaboradores han sido líderes mundiales en la recopilación y análisis de los modelos de predicción y pronóstico generados durante la pandemia. Han generado una iniciativa de colaboración internacional (COVID PRECISE) para el acompañamiento en el desarrollo de este tipo de herramientas con la intención de mejorar su calidad y lograr su generalización (55).

Para la realización de este trabajo se utilizó el primero de sus estudios en el que se recolectaron (de manera sistemática) y analizaron todos los modelos publicados hasta mayo de 2020 (10). En el momento de redactar los resultados de este estudio ya cuentan con una nueva publicación actualizada con mayor información sobre el

análisis de la calidad de las escalas disponibles (56). En ambos la conclusión es similar: a pesar de que ya existen publicados más de 50 modelos de predicción de pronóstico de pacientes con enfermedad por coronavirus, ninguno cuenta con los criterios necesarios que permitan recomendar su uso generalizado. La mayoría ha generado estimados muy optimistas en su capacidad de discriminación (probablemente por errores en las estrategias de análisis) y no evalúan la calibración como medida de desempeño. Además, muy pocos cuentan con un proceso de validación externa.

Hasta donde se ha revisado la literatura aún no existen modelos de predicción de pronóstico desarrollados en Colombia y son pocos los que han surgido en América Latina. El desarrollo y validación de estas herramientas a nivel local es una necesidad para mejorar la atención y optimizar los recursos disponibles en salud para la atención de la COVID-19.

### 3. OBJETIVOS

#### 3.1. **Objetivo General**

- Realizar una validación externa de escalas de predicción de mortalidad al aplicarlas en la cohorte RESCATE COVID-19.

#### 3.2. **Objetivos específicos**

1. Seleccionar mediante un proceso estructurado las escalas de predicción de mortalidad disponibles que puedan ser aplicadas a la población de la cohorte.
2. Realizar una caracterización cualitativa y del desempeño de las escalas seleccionadas en sus respectivas cohortes de desarrollo y validación.
3. Evaluar el desempeño de cada una de las escalas seleccionadas en la cohorte RESCATE COVID-19 según su capacidad de discriminación calibración.
4. Evaluar el desempeño de las escalas seleccionadas en la población de pacientes sospechosos del registro.

## 4. DESARROLLO METODOLÓGICO POR OBJETIVOS

### 4.1. Diseño general:

Estudio de validación externa de escalas de pronóstico de mortalidad o deterioro clínico en pacientes de la cohorte RESCATE COVID-19. Se considera un estudio de evaluación de tecnologías disponibles en salud para su aplicación local. Sus primeros componentes tienen una naturaleza descriptiva ya que se pretende caracterizar de manera cualitativa a los modelos de predicción disponibles en la literatura. El componente analítico consiste en evaluar el desempeño de las escalas elegidas al aplicarlas a una población externa en términos de su capacidad de discriminación y calibración.

### 4.2. Para el desarrollo del primer objetivo: Proceso de selección estructurada de las escalas de predicción disponibles

Se utilizó como insumo una revisión sistemática viva de la literatura que incluye los modelos de predicción de diagnóstico y pronóstico para pacientes con COVID-19 que han sido publicados en PubMed, Embase, Arxiv, medRxiv, or bioRxiv hasta mayo de 2020 (10). Se realizó una selección por pasos de las escalas candidatas según el desenlace, las variables incluidas en el modelo y la posibilidad de reconstruirlos según la información publicada por los autores.

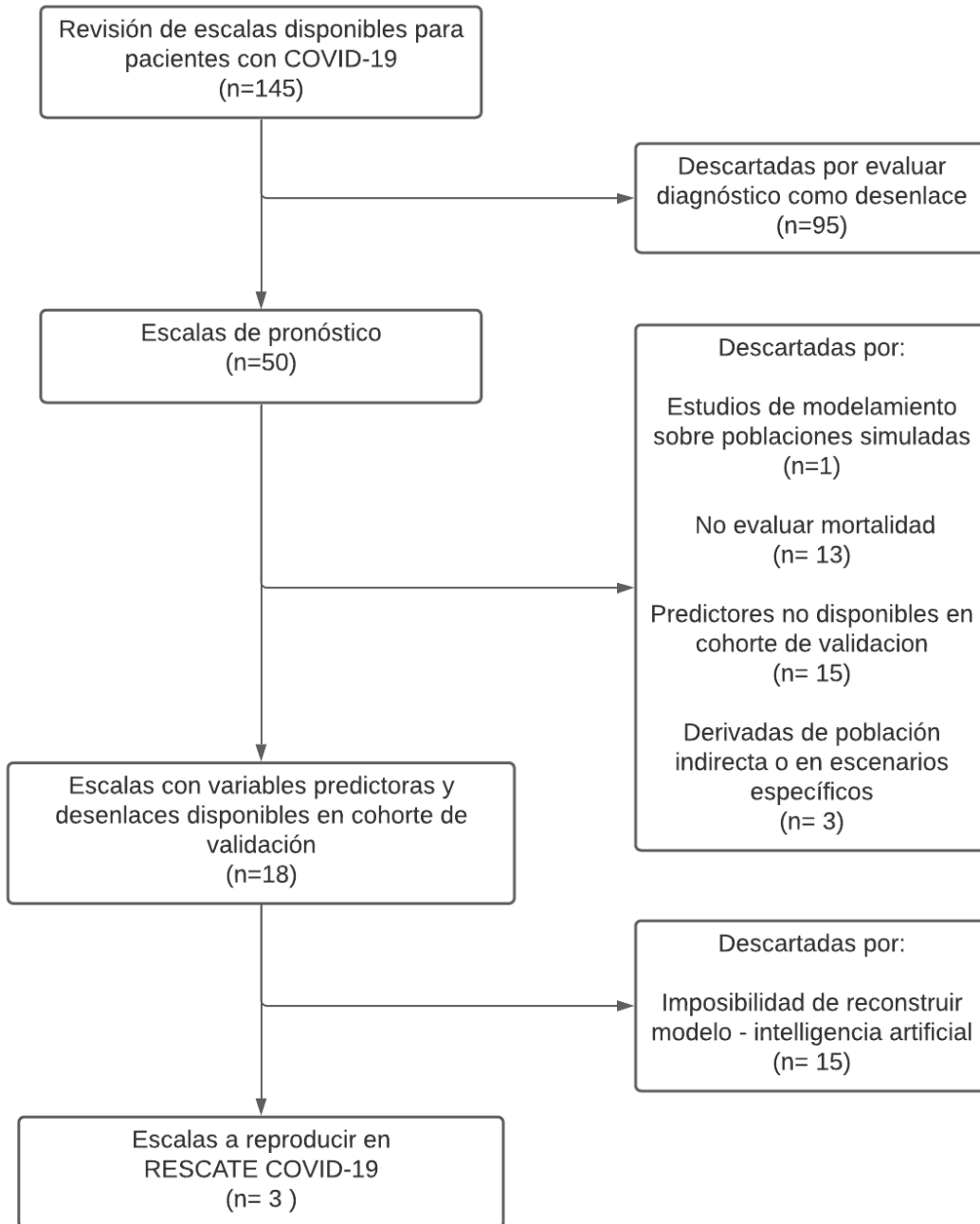
Se creó una tabla en la que se describen las características de cada uno de los estudios originales y que contiene el resultado del proceso de selección con las razones para su inclusión o descarte (Anexo 2).

- Paso 1: Se descartaron los estudios de modelamiento sobre poblaciones simuladas.
- Paso 2: Se descartaron los modelos que no evaluaron mortalidad como alguno de los desenlaces de interés.
- Paso 3: Se descartaron las escalas que no utilizaron como predictores a las variables sociodemográficas, clínicas, de laboratorio o de imágenes disponibles en la cohorte RESCATE COVID-19.

- Paso 4: Se descartaron los estudios que utilizaron como técnica de modelamiento el aprendizaje automatizado o "*machine learning*".
- Paso 5: Tras haber decantado a los estudios que evalúan mortalidad como desenlace y que utilizan variables predictoras disponibles en la cohorte de validación con técnicas de modelamiento basadas en regresión logística se evaluó su factibilidad para la replicación.

Teniendo en cuenta los recursos disponibles para la realización del estudio se decidió incluir un máximo de tres escalas con las características descritas para realizar el análisis y las comparaciones. El siguiente flujograma resume el proceso de selección:

**Figura 1. Proceso de selección de las escalas de predicción de mortalidad con potencial para validación local en pacientes de la cohorte RESCATE COVID-19:**





### **4.3. Para el desarrollo del segundo objetivo: Caracterización de las escalas potencialmente aplicables**

La caracterización detallada de cada uno de los modelos seleccionados se encuentra en la Tabla 1. Dos de los modelos seleccionados fueron derivados de cohortes retrospectivas de pacientes con diagnóstico confirmado de COVID-19 que consultaron a instituciones hospitalarias de China (57,58). En ambos casos los autores realizaron un proceso de validación externa de sus modelos en poblaciones con diferentes tasas de mortalidad, en distintos momentos de tiempo y con poblaciones demográficamente diversas. En uno de los casos la validación se realizó en una cohorte China y en el otro una cohorte de Londres.

El tercero de los estudios incluidos es el único dentro de la revisión sistemática que se realizó en población latinoamericana (59). Utiliza información de una base de datos nacional del ministerio de salud de México que contiene información sobre los pacientes a quienes se les había realizado una RT-PCR para SARS-COV2 en lo corrido de la pandemia. El modelo de predicción derivado de la cohorte se realizó únicamente con información de los pacientes positivos. Su proceso de validación fue interno, aplicando la escala a una submuestra del 20% seleccionada de manera aleatoria, pero no cuenta con una validación externa aún publicada.

A pesar de que algunos de los estudios generaron modelos para predicción de complicaciones de la enfermedad, para el interés de este estudio se utilizó únicamente información de los que usaron mortalidad como desenlace. Mientras que los dos estudios chinos evaluaron mortalidad intrahospitalaria el estudio mexicano evaluó mortalidad a 30 días después del resultado de la prueba.

Respecto a los predictores, los dos modelos orientales analizaron inicialmente de manera individual el efecto de variables sociodemográficas (edad y género), comorbilidades, síntomas y de laboratorio. Por su parte el diseño del modelo mexicano estuvo orientado a evaluar únicamente variables sociodemográficas, de antecedentes y algunas sobre diagnóstico clínico registrado en la base de datos nacional (v.g. diagnóstico de neumonía). Zhang y colaboradores agruparon las variables predictoras según su naturaleza (sociodemográficos, síntomas, laboratorios entre otros) y generaron modelos con las diferentes combinaciones.

El estudio con mayor tamaño de muestra es el de Bello-México (n=51.633), seguido por Zhang-China (n=775) y Xie-China (n=299). Las tasas de eventos en el estudio

de Xie fueron del 51% y del 47.6% (en las cohortes de derivación y validación respectivamente) siendo las más de los estudios seleccionados. La alta mortalidad es explicada por la población que compone estas cohortes: pacientes atendidos en los primeros hospitales designados por el gobierno chino para la atención del COVID-19 en las primeras etapas del brote epidémico local. Similar explicación tiene la incidencia de mortalidad en la cohorte de Londres en la que se validó el modelo de Zhang y colaboradores (34.1%). La incidencia de mortalidad en el estudio mexicano fue del 10.3%.

Los dos estudios orientales utilizaron estrategias de regresión logística múltiple para la generación de los modelos mientras que Bello y colaboradores utilizaron regresión de riesgo proporcional de Cox ya que contaron con más de una medición en el tiempo hasta el evento. Este último generó un sistema sencillo de puntos y niveles de riesgo como herramientas para la aplicación clínica del modelo. A su vez Zhang y colaboradores cuentan con una plataforma electrónica con acceso en línea para calcular las probabilidades del evento según las variables de modelo final (60). El modelo Xie-China utiliza un normograma para un abordaje semicuantitativo de puntuación con intención de traducción clínica, disponible en físico en el documento original (58).

Respecto a la evaluación del desempeño de los modelos el que resultó con mejor área bajo la curva ROC (*AUC por sus siglas en inglés*) en su cohorte de derivación fue el de Zhang seguido por Xie y Bello ( $AUC=0.91$ ,  $0.89$  y  $0.81$  respectivamente). Sin embargo, en la cohorte de validación la primera escala perdió capacidad de discriminación ( $AUC=0.74$ ) al aplicarse en una cohorte externa mientras que en la segunda y tercera su desempeño aumenta ( $AUC=0.98$  y  $0.83$ ). De esta manera en las cohortes de validación la escala de Xie y colaboradores ofrece una mejor discriminación entre pacientes con y sin riesgo de fallecer por COVID-19 seguida de la escala de Bello y Zhang. Como se explicó anteriormente, esta última evaluó diferentes modelos realizados con combinaciones de predictores según su naturaleza. La escala compuesta por una combinación de variables demográficas y de laboratorio (modelo "DL") resultó tener las mejores medidas de discriminación en y corresponden a las anteriormente reportadas.

El estudio de Bello no realizó un análisis de calibración de su modelo como una medida del desempeño. En los dos modelos chinos se observan niveles muy buenos de calibración en sus cohortes de derivación (calibración en promedio y pendientes de  $0.18$  y  $0.98$  en modelo Zhang;  $0$  y  $1$  en modelo Xie). Sin embargo, la calibración se ve afectada en las cohortes de validación de manera importante. En

el caso de Zhang la calibración en promedio se acercó a 0 (0.25) de manera que disminuyó la subestimación en general pero la pendiente se alejó de 1 (0.54) sugiriendo estimaciones extremas por el modelo. En el caso de Xie la calibración en promedio se tornó negativa y la pendiente sobrepasó el valor de 1 sugiriendo una sobreestimación sistemática del modelo y extrema en los niveles de mayor y menor riesgo.

Todos los estudios recibieron una calificación de alto riesgo de sesgos utilizando la herramienta PROBAST (61). En la tabla se expone la calificación según el riesgo de sesgos en cada componente de la herramienta. Los dos estudios orientales tienen buena calificación en cuanto a la descripción de la selección de participantes, predictores y desenlaces. Presentan baja calificación en el ítem de “análisis” especialmente por ausencia de información en el reporte sobre manejo de datos perdidos, estrategias para el análisis de las variables y en la conducción de los modelos. El estudio de México obtuvo calificación baja en la manera que reportan las características de su población y la selección de sus predictores además de las mismas dificultades en el análisis que los estudios chinos. La calificación fue consistente con la que obtuvieron todos los estudios en la revisión sistemática de Wynants y colaboradores (10). Se debe aclarar que las bajas puntuaciones en esta herramienta se pueden deber a deficiencias en los reportes y no necesariamente a falta de validez de los modelos. La validación externa que se pretende realizar es una de las estrategias que permiten complementar la evaluación sobre la utilidad y la potencial de generalización del uso de las escalas (62,63).

En conclusión, los tres modelos tienen fortalezas que los hacen merecedores de un proceso de validación externa. La naturaleza de sus predictores, las estrategias de modelamiento y la evaluación de un desenlace relevante y fácil de medir (“duro”) permite su replicación y evaluación. Los tres muestran capacidad de discriminación y calibración muy buena ( $AUC > 0.8$ , calibración en promedio cercanas a 0 y pendientes cercanas a 1) que se ven afectadas al ser evaluadas en cohortes externas como se ha descrito con la mayoría de modelos de predicción (64). El modelo mexicano a pesar de no contar con una validación externa muestra una buena capacidad de discriminación en su derivación, validación interna y proviene de una cohorte con una gran población de pacientes y número de eventos. Las escalas de Zhang y de Bello ofrecen sistemas de puntuación de fácil acceso (en línea en el primer caso) o cálculo (suma de puntos en el segundo caso). Por su parte la traducción clínica del modelo de Xie con un nomograma no permite una fácil adopción en la práctica clínica.

Tabla 1. Caracterización de los estudios con los modelos de predicción de mortalidad seleccionados para su validación externa.

Autor. País. Año. Revista.	Diseño	Desenlaces	VARIABLES en el modelo final	Tamaño muestra en Derivación/ Eventos. N/n. (%)	Tipo de Modelo	Formato disponible	Tipo de Validación	Tamaño muestra en Validación/ Eventos. N/n. (%)	Desempeño (AUC – S: E: VPP: VPN: R2:)	Riesgo de sesgos (PROBAST) Según 4 dominios
<b>Modelos que evalúan mortalidad como desenlace</b>										
<b>Zhang. China-UK. 2020. MedRxiv preprint</b>	Cohorte retrospectiva  Pacientes con RT-PCR positiva para SARS-COV2 que ingresan a servicio de urgencias  <u>Derivación:</u> Wuhan – China.  <u>Validación:</u> King's College Hospital - Londres	1. <u>Muerte intrahospitalaria</u>  2. Curso desfavorable:  - Ingreso a UCI . SDR - Intubación orotraqueal - Requerimiento de ECMO.	<b>Tres grupos:</b> <b>1. DCS:</b> Demográficos + Comorbilidades + Síntomas  <b>2. DCSL:</b> Demográficos + Comorbilidades + Síntomas + Laboratorios  <b>3. DL(Simplificado):</b> Demográficos + Laboratorios.  <b>Componentes:</b> <b>Demográficos (D):</b> - Edad (años) - Sexo masculino (%)  <b>Comorbilidades (C):</b> - EPOC (Si/No) - DM2 (Si/No) - Inmunocompromiso (Si/No) - Malignidad (Si/No) - HTA (Si/No) - Enfermedad Cardíaca (Si/No) - ERC (Si/No)  <b>Síntomas (S):</b> - Fiebre (Si/No) - Tos (Si/No) - Fatiga (Si/No) - Disnea (Si/No) - Diarrea (Si/No)	1. <u>Muerte:</u> 775/33 (4.3%)  2. Curso desfavorable ("Pron"): 772/75 (9.7%)	Regresión Logística múltiple.	Sistema de cálculo de probabilidades con <b>herramienta digital de acceso en línea.</b>  Clasificación en niveles de riesgo según puntos de corte de la probabilidad:  - <u>Muy Bajo riesgo</u> (Prob: <0.6%)  - <u>Riesgo Bajo</u> (Prob: 0.6 – 8.9%)  - <u>Alto riesgo</u> (Prob:8.9 - 58.5%)  - <u>Muy alto riesgo</u> (Prob> 58.5%)  Herramienta en línea disponible:  <a href="https://covid_datah.elps.life/prediction/">https://covid_datah.elps.life/prediction/</a>  <a href="https://covid_datah.elps.life/wuhan01-model/">https://covid_datah.elps.life/wuhan01-model/</a>	<b>Interna:</b> Bootstrapping (1000 repeticiones) de todo el proceso de construcción del modelo, incluida la selección paso a paso hacia atrás (método forward) de todos los predictores potenciales.  <b>Externa:</b> - En cohorte de Londres (King's College Hospital)	1. <u>Muerte:</u> 226/77 (34.1%)  2. Curso desfavorable: 226/97 (42.9%)	<b>1. Capacidad de Discriminación:</b>  <b>Cohorte de Derivación:</b>  <b>Estadístico C:</b> - DCS-Muerte: :0.79. - DCSL-Muerte: 0.89 - <b>DL-Muerte: 0.91</b> - DCS-Pron: 0.80 - DSCL-Pron: 0.88 - DL-Pron: 0.88  <b>Cohorte de Validación:</b>  <b>Estadístico C:</b> - <b>DL-Muerte: 0.74</b>  - DL-Pron: 0.72  <b>2. Calibración:</b>  <b>Cohorte de derivación:</b> - DCS-Muerte: * Pendiente =0.287	<b>Según 4 dominios:</b>  - Participantes: Bajo.  - Predictores: Bajo  - Desenlaces: Bajo  - Análisis: Alto

			<b>Laboratorios (L):</b> - Rec. Neutrófilos (10 <sup>9</sup> /L) - Rec. Linfocitos (10 <sup>9</sup> /L) - Rec. Plaquetas(10 <sup>9</sup> /L) - PCR (mg/ml) - Creatinina (μmol/L)					* Calibración en promedio = 0.384  - DCSL-Muerte: * Pendiente =0.460 * Calibración en promedio = 0.316  - DL-Muerte: * Pendiente =0.981 * Calibración en promedio (CITL) = 0.189  - DCS-Pron: * Pendiente =0.855 * Calibración en promedio = 0.132  - DSCL-Pron: * Pendiente =0.0.59 * Calibración en promedio = 0.894  <b>Cohorte de validación:</b>  - DL-Muerte: * Pendiente =0.543 * Calibración en promedio(CITL) = 0.251		
<b>Xie. China. 2020.</b>  <b>MedRxi v preprint</b>	Cohorte retrospectiva  Pacientes que ingresan y se confirma diagnóstico de COVID-19 por PCR  >18 años.	<u>Mortalidad intrahospitalaria</u>	- Edad - Lactato deshidrogenasa - Recuento de linfocitos - Saturación de oxígeno por pulsioximetría	<u>Mortalidad intrahospitalaria:</u> 299/155 (51.8%)	Regresión logística múltiple.  Proceso de selección paso a paso hacia atrás con mejora de bondad de ajuste evaluada por reducción en el Criterio de	<b>Nomograma</b>  Sistema semicuantitativo.	<b>Interna:</b> Método de remuestreo de Bootstrapping  1000 repeticiones de todo el proceso de construcción del modelo.  <b>Externa:</b> En diferentes centros	145/69 (47.6%)  Solo 127 de los 145 tenían los datos completos de variable**	<b>1. Capacidad de Discriminación:</b>  <b>Cohorte de derivación:</b> - Estadístico C: 0.89  <b>Cohorte de validación:</b> - Estadístico C: 0.98	<b>Según dominios: 4</b>  - Participantes: Bajo.  - Predictores: Bajo  - Desenlaces: Bajo  - Análisis: Alto

	<p><u>Derivación:</u> Hospital Wuhan Jinyintan</p> <p><u>Validación:</u> Hospital Wuhan Tongji</p>				Información de Akaike (AIC)		<p>hospitalarios de China.</p> <p>* <u>Diferencias en temporalidad:</u></p> <p>- Derivación: más temprana en el curso de la pandemia.</p> <p>- Diferencias demográficas.</p>	<p><b>2. Calibración:</b></p> <p><b>Cohorte de derivación:</b></p> <p>- Pendiente =1 - Calibración en promedio (CITL) = 0.00</p> <p><b>Cohorte de validación:</b></p> <p>- Pendiente= 2.227 - Calibración en promedio (CITL) = -0.192</p>		
<p><b>Bello-Chavolla. México. 2020.</b></p> <p><b>J Clin EndocrinMetab.</b></p>	<p>Base de datos nacional de Ministerio de Salud de México.</p> <p>Incluye casos PCR+, PCR- y sospechosos.</p> <p>Pacientes de población general que consultó por síntomas de COVID-19.</p> <p>Hospitalización: 38%</p>	<p><u>Mortalidad a 30 días.</u></p> <p><b>**El modelo de predicción se realizó sólo con casos positivos (RT- PCR).</b></p>	<p>- Edad &lt;40 años (SI/No)</p> <p>- Edad &gt;65 años (SI/No)</p> <p>- Diabetes (SI/No)</p> <p>- Diabetes &lt;40 años (SI/No)</p> <p>- Obesidad (SI/No)</p> <p>- Neumonía (SI/No)</p> <p>- ERC (SI/No)</p> <p>- EPOC (SI/No)</p> <p>- Inmunosupresión (SI/No)</p>	<p><u>En una muestra aleatoria del 80% de los positivos:</u> (Derivación): 41.307/4.276.</p> <p>- <u>Mortalidad:</u> 10.3%</p> <p>Población total de positivos: 51.633/5.332</p> <p>Mortalidad: 10.3%</p>	<p>Modelo de regresión de riesgo proporcional de Cox.</p> <p>Tiempo desde inicio de síntomas hasta muerte o censura.</p>	<p><b>Escala de puntos:</b></p> <p>- <u>Neumonía:</u> 7</p> <p>- <u>DM &lt;40 años:</u> 5</p> <p>- <u>&gt;65 años:</u> 3</p> <p>- <u>ERC:</u> 3</p> <p>- <u>Inmunosup:</u> 1</p> <p>- <u>EPOC:</u> 1</p> <p>- <u>Obesidad:</u> 1</p> <p>- <u>DM:</u> 1</p> <p>- <u>Edad &lt;40:</u> (-)6</p> <p><b>Categorías de riesgo según puntaje total en la escala:</b></p> <p>- <u>Muy alto:</u> ≥12</p> <p>- <u>Alto:</u> 8-11</p> <p>- <u>Moderado:</u> 4-7</p> <p>- <u>Bajo:</u> 1-3</p> <p>- <u>Muy Bajo:</u> ≤0</p>	<p>Training test Split en 80% y validado posteriormente en 20%.</p> <p>Muestra para entrenamiento elegida de manera aleatoria estratificada por mortalidad (Paquete Caret R)</p>	<p>Muestra aleatoria del 20% de los positivos 10.326/1.056</p> <p><u>Mortalidad:</u> 10.3%</p>	<p><b>1. Capacidad de Discriminación:</b></p> <p><b>Cohorte de derivación:</b></p> <p>- Estadístico C: 0.817 (Usando las variables del modelo)</p> <p>- Estadístico C: 0.823 usando sistema de puntos</p> <p>- Estadístico C: 0.810. Usando categorías de riesgo-</p> <p><b>Cohorte de validación:</b></p> <p>- Estadístico C: 0.830. Usando los predictores del modelo.</p> <p>- Estadístico C: 0.821. Usando categorías de riesgo.</p> <p><b>2. Calibración: No evaluada*</b></p>	<p><b>Según dominios: 4</b></p> <p>- Participantes: Poco claro</p> <p>- Predictores: Poco Claro</p> <p>- Desenlaces: Bajo</p> <p>- Análisis: Alto</p>

#### **4.4. Para el desarrollo del tercer objetivo: Validación externa de las escalas.**

##### **4.4.1. Metodología de la Cohorte RESCATE COVID-19:**

La validación de las escalas seleccionadas se realizó en la población de la cohorte RESCATE COVID-19. Está conformada por pacientes (adultos y niños) que han consultado a servicios de urgencias (requiriendo o no hospitalización) y han sido catalogados como casos sospechosos por sus médicos tratantes según los criterios del ministerio de salud vigentes al momento de la atención. Se han incluido de manera consecutiva pacientes con prueba positiva de RT-PCR para SARS-CoV2 (en muestra de tracto respiratorio) y una muestra estratificada de pacientes con resultado negativo (según fecha de atención, grupo de edad y estado vital al egreso). Los pacientes que han acudido a urgencias sin cumplir criterios clínicos de caso sospechoso y a quienes se les ha realizado prueba molecular por razones diferentes a la sospecha clínica (solicitud laboral, estudios prequirúrgicos o tamización entre otros) no hacen parte de esta cohorte.

Hasta el momento seis hospitales alto nivel de complejidad ubicados en Bogotá, Medellín y Bucaramanga han participado en el proyecto. La Fundación Cardioinfantil - Instituto de Cardiología (FCI-IC) como centro generador y líder del estudio se ha encargado de coordinar el proceso de recolección de la información. La toma de datos del componente de atención hospitalaria ha sido realizada de manera retrospectiva mediante la revisión de historias clínicas y posteriormente registrada en la plataforma web *Research Electronic Data Capture* (REDCap®).

Los módulos del registro incluyen información sociodemográfica, antecedentes médicos patológicos, síntomas al ingreso, tratamientos recibidos, resultados de laboratorios y de imágenes diagnósticas (radiografías y tomografías computarizadas). De igual manera, se dispone de información sobre desenlaces clínicos como complicaciones, ingreso a la unidad de cuidado intensivo, requerimiento de ventilación mecánica, nivel de gravedad, estancia hospitalaria y mortalidad, entre otros. Los detalles del protocolo del proyecto RESCATE COVID-19 se encuentran publicados en línea y se anexan al final de este documento (Ver Anexo 3).

Los informes de avance del proyecto con la caracterización demográfica y clínica de la población incluida, así como los desenlaces de interés están disponibles en línea en formato de boletines de publicación periódica e infografías (<http://38.130.204.219/2020/fci/rescate/>). El último de los boletines de avance, con la caracterización de los pacientes incluidos hasta febrero de 2021 se anexa al final de este documento (Anexo 4). Así mismo el primero de los manuscritos que se pretenden generar a partir de este proyecto se encuentra en proceso de publicación. Para efectos de este trabajo se utilizó la información de los pacientes vinculados al estudio hasta el 31 de diciembre de 2020 como se planificó en el cronograma de actividades. Se contó con información de 1342 pacientes con PCR positiva (casos confirmados) y 171 eventos fatales (mortalidad de 12.74%).

#### **4.4.2. Calidad de los datos:**

El proceso detallado de recolección de la información en la cohorte RESCTE COVID-19 se encuentra disponible en el protocolo y manual de operaciones (Anexo 2). Cada uno de los centros participante cuenta con un grupo de médicos previamente entrenados, que se encargan de la recolección de información en las historias clínicas de los pacientes seleccionados. En el centro coordinador del estudio (FCI-IC) se encuentra un comité de administración de datos que está conformado por el líder del proyecto, la coordinadora de la plataforma REDCap®, la coordinadora de registro hospitalario y seguimiento telefónico, el médico líder del equipo de recolección de datos y el ingeniero de datos del proyecto. Se ha encargado de supervisar la información consignada diariamente por los centros en la plataforma REDCap® y de revisar la calidad de los registros. Han realizado auditoría de los datos, estrategias para disminuir el riesgo de ingresar datos erróneos al definir rangos en las variables continuas, definir preguntas “puerta” de obligatorio diligenciamiento y opciones de selección múltiple. Se han encargado de comunicarse con el investigador principal de cada centro cuando se han identificado inconsistencias con la intención de resolverlas en el menor tiempo posible. El comité ha realizado copias de seguridad semanales (cada martes) que han sido almacenadas como mínimo en dos sitios o dispositivos diferentes (al menos uno físico).



#### **4.4.3. Variables y unidades de medida:**

Las variables que se describen a continuación corresponden a aquellas propuestas por los modelos seleccionados como predictores que fueron evaluadas en la cohorte RESCATE-COVID19 (Ver Tabla 2). Como se explicó previamente uno de los criterios para la selección de las escalas fue el uso de variables disponibles en la cohorte local en cuanto a su naturaleza, su estrategia de recolección y análisis. En todos los casos la variable dependiente con la que se analizaron los modelos en nuestra cohorte fue la mortalidad intrahospitalaria. Es definida como el registro en la historia clínica de fallecimiento durante el proceso de atención clínica tras haber ingresado como caso sospechoso de COVID-19 según los criterios del ministerio de salud vigentes al momento de la atención.

Para el caso de la escala de Bello-México existen tres variables cuya definición operativa en nuestra cohorte no coincide con aquella del estudio original. La edad se utiliza con dos puntos de corte ( $<40$  y  $\geq 65$  años) mientras que en RESCATE se obtuvo en su naturaleza continua. Con la intención de evaluar el modelo con las variables propuestas se generaron estas categorías dentro de la variable “edad” y se generaron dos nuevas variables (“edad  $<40$ ” y “edad  $\geq 65$ ”) finalmente usadas para la evaluación del modelo. Si bien es posible que la transformación a una variable discreta le reste capacidad de informar, es importante reconocer que las categorías pueden contribuir a una mejor interpretación clínica.

Por otra parte, los autores del estudio mexicano construyeron su cohorte con información de una historia clínica unificada a nivel nacional donde se encuentran los datos de todas las atenciones médicas de un individuo a lo largo de su vida. De esta manera podían verificar si el diagnóstico de diabetes había sido registrado antes de los 40 años. Eligen esta condición como una potencial variable de riesgo teniendo en cuenta evidencia previa sobre su relación con mayor riesgo de complicaciones cardiovasculares tempranas (65). Para nuestro caso no se cuenta con información verificable en una historia clínica nacional sobre el momento de registro de antecedentes como diabetes. Para efectos de la aplicación del modelo se decidió generar una variable llamada “diabetes  $<40$ ” que incluye a los pacientes con menos de 40 años y diagnóstico de diabetes en la historia clínica. Los pacientes con diabetes y 40 años o más se declararon como datos perdidos para evitar el sesgo de información.

Tabla 2. Caracterización de las variables independientes incluidas en cada uno de los modelos.

Variable predictora	Clasificación	Unidad de medida	Definición operacional en RESCATE COVID-19	Coincide con cohorte de derivación
<b>Modelo Zhang-China</b>				
<b>Sexo</b>	Discreta politómica	Masculino/Femenino /Indeterminado	Registro en la historia clínica de condición sexual	Si
<b>Edad</b>	Continuas	Años	Tiempo transcurrido en años desde fecha de nacimiento hasta la fecha de ingreso a urgencias	
<b>Recuento de neutrófilos</b>		x10 <sup>3</sup> cel/uL	Recuento celular en el primer hemograma realizado al ingreso y registrado en la historia clínica	
<b>Recuento de linfocitos</b>				
<b>Recuento de plaquetas</b>				
<b>Creatinina</b>		mg/dL	Primer valor de creatinina obtenido al ingreso y reportado en la historia clínica	
<b>Proteína C reactiva</b>	Primer valor de PCR obtenido al ingreso y reportado en la historia clínica			
<b>Modelo Xie-China</b>				
<b>Edad</b>	Continuas	Años	Tiempo transcurrido en años desde fecha de nacimiento hasta la fecha de ingreso a urgencias	Si
<b>Saturación de oxígeno</b>		Porcentaje de saturación de oxígeno (%)	Nivel de saturación de oxígeno obtenido por oxímetro de pulso al ingreso	
<b>Recuento de linfocitos</b>		x10 <sup>3</sup> cel/uL	Recuento celular en el primer hemograma realizado al ingreso y registrado en la historia clínica	
<b>Deshidrogenasa Láctica</b>		Unidades/Litro (U/L)	Primer valor de LDH obtenido al ingreso y reportado en la historia clínica	
<b>Modelo Bello-México</b>				
<b>Edad menor a 40 años (&lt;40)</b>	Discretas dicotómicas	Si/No	Cumple con la condición de edad menor a 40 años al momento de la consulta.	No
<b>Edad mayor o igual a 65 años (≥65)</b>			Cumple con la condición de edad 65 años o mayor al momento de la consulta.	
<b>Diabetes en pacientes menores de 40 años</b>			Cumplir con las dos condiciones: Antecedente registrado en la historia clínica de alguna de las condiciones Y edad menor a 40 años al momento de la consulta.	

<b>Neumonía</b>			Diagnóstico de neumonía registrado en la historia clínica y confirmado por investigador de cada centro como cumplimiento de criterios según las categorías de riesgo de Asociación Colombiana de Infectología.	Si
<b>Diabetes</b>			Antecedente registrado en la historia clínica durante la atención médica.	
<b>Obesidad</b>				
<b>Enfermedad renal crónica (ERC)</b>				
<b>Enfermedad pulmonar obstructiva crónica (EPOC)</b>				
<b>Inmunosupresión</b>				

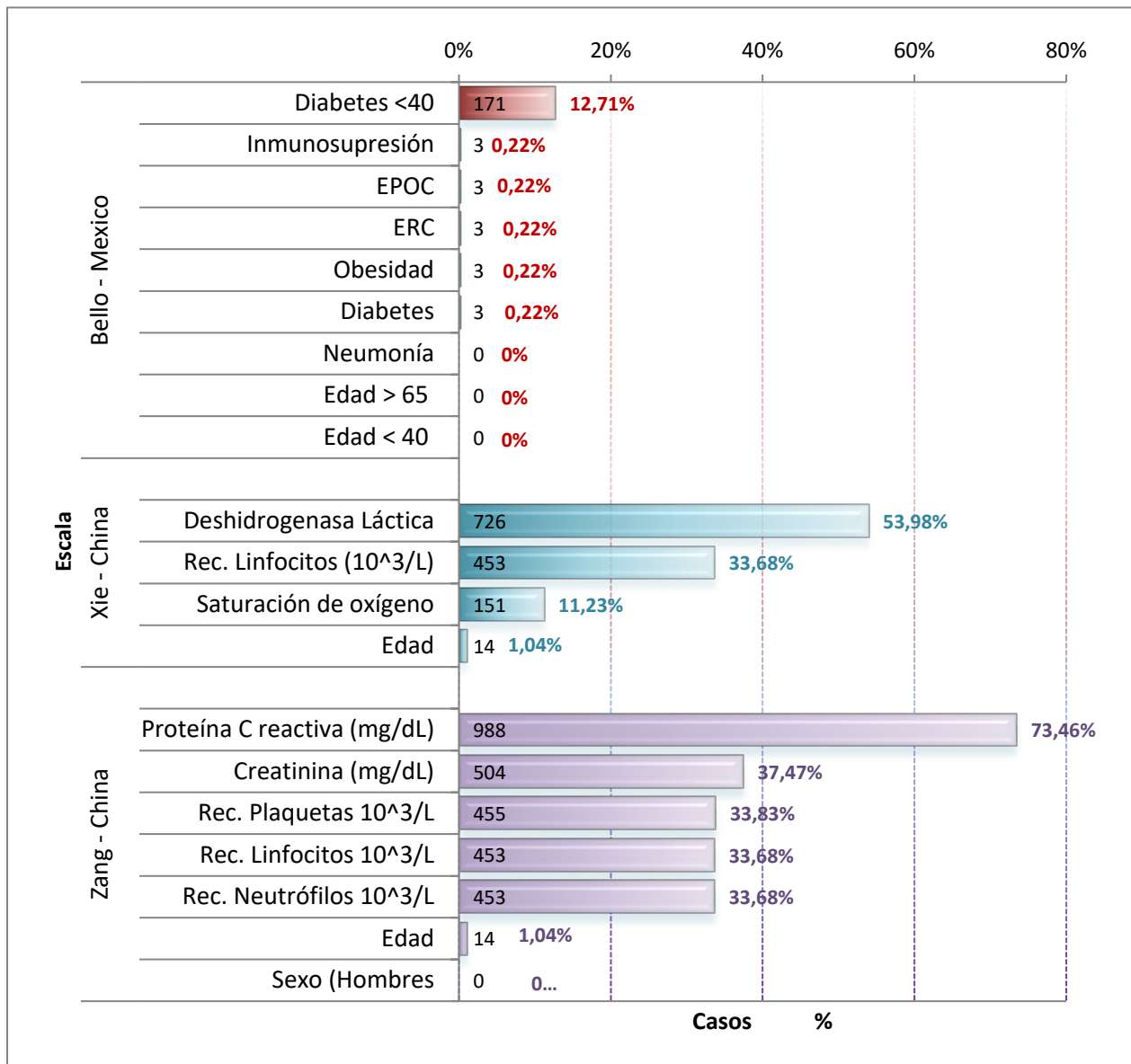
#### **4.4.4. Manejo de los datos faltantes en la cohorte de validación:**

El conteo y proporciones de los datos faltantes en cada una de las variables de las escalas evaluadas se encuentra detallado en la Figura 2. Las proporciones se calculan con base en los 1342 pacientes incluidos para el análisis. El modelo que se logró analizar en un mayor número de paciente fue el mexicano ya que la mayoría de sus predictores corresponden a características sociodemográficas y antecedentes que en la práctica clínica usual, por su relevancia, se suelen registrar en la historia clínica. La variable con mayor número de datos perdidos corresponde a “Diabetes <40 años” que como se explicó en la sección anterior corresponde a aquellos pacientes con diabetes y más de 40 años en los que no fue posible conocer el momento de su diagnóstico.

En cuanto a las variables de la escala de Xie, más de la mitad de los pacientes incluidos no contaban con información sobre la deshidrogenasa láctica al ingreso. De igual manera la proteína C reactiva fue el estudio de laboratorio con menos información disponible para la evaluación de la escala de Zhang. Lo anterior debido a que no corresponden a estudios solicitados como parte de la práctica clínica diaria local al ingreso de los pacientes en servicio de urgencias. Aproximadamente un tercio de la población con hemograma ni creatinina medidos a su ingreso.

Cada uno de los modelos fue evaluado en la población que contaba con la totalidad de la información de las variables que los componen. De esta manera cada uno de los modelos tiene un numero de observaciones diferentes que se reportan en cada una de las tablas y las figuras de la sección de resultados. La prevalencia de eventos en cada una de las poblaciones también se reporta en cada proceso de validación.

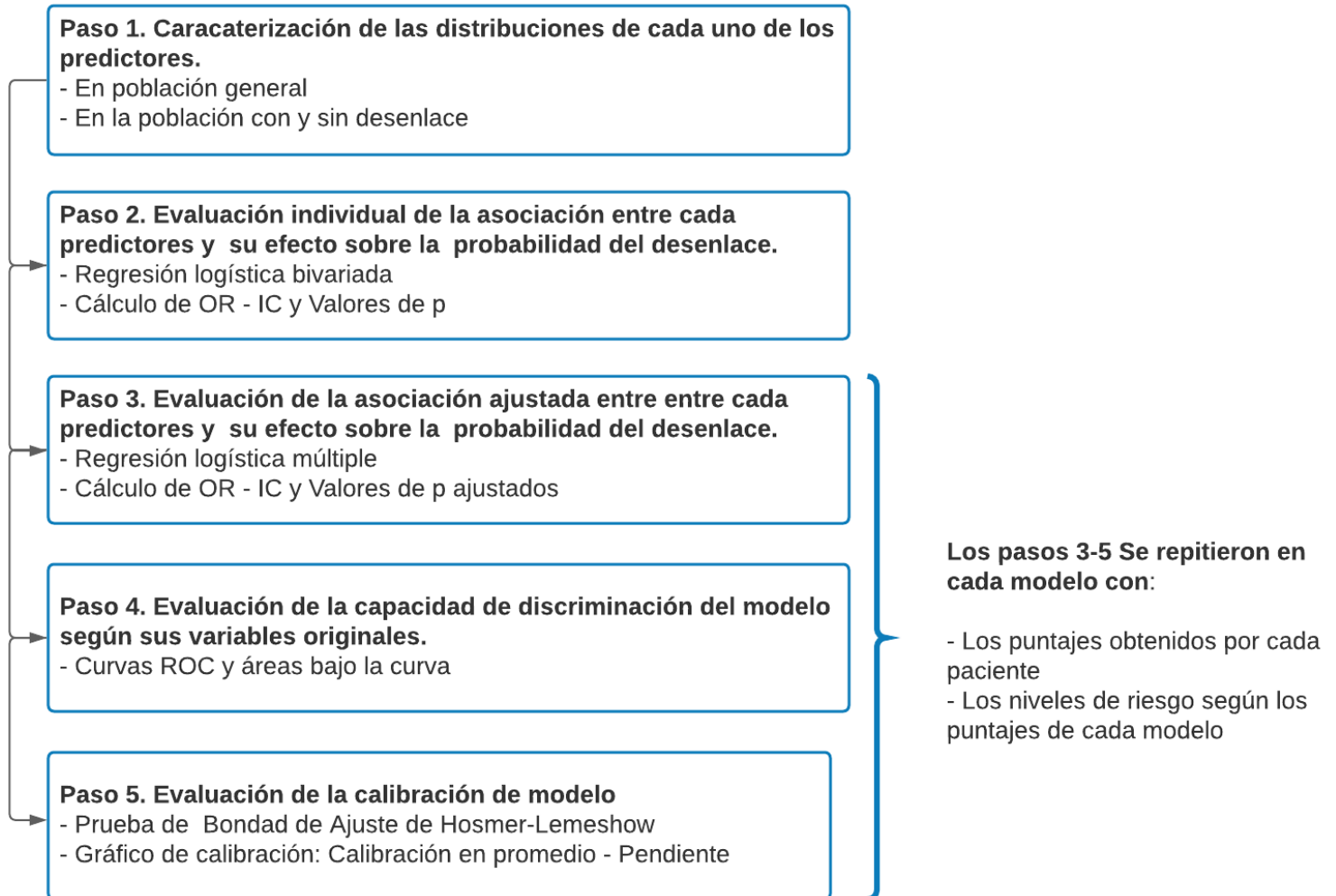
Figura 2. Datos faltantes de cada una de las variables de las escalas evaluadas:



Variable dependiente: Mortalidad intrahospitalaria: 171 eventos (12.7%)

#### 4.4.5. Metodología para la evaluación del desempeño de las escalas en términos de su capacidad de discriminación y calibración en la cohorte local:

Figura 3. Proceso para la evaluación del desempeño de las escalas seleccionadas según su capacidad de discriminación y su calibración.



OR: Odds Ratio. IC: Intervalos de confianza del 95% alrededor del OR. Valor para la prueba estadística. ROC: *receiver operating characteristic*.

El proceso realizado con cada uno de los modelos para su validación siguió un proceso estructurado de pasos que se describen en el la figura 2. El primer paso consistió en la evaluación de las características de cada una de las variables y su distribución en los grupos de pacientes con y sin el evento de interés (mortalidad intrahospitalaria). Para cada una se construyó un histograma, una gráfica de calibración y se realizó una prueba de estadística de Shapiro Wilk con la intención de evaluar su ajuste a una distribución normal (Anexo 3). Teniendo en cuenta las distribuciones de las variables continuas, se utilizaron sus medianas como medidas de tendencia central y sus rangos intercuartílicos (RIQ) como medidas de dispersión alrededor de la mediana. Para las variables discretas se utilizaron conteos y proporciones como medidas de resumen.

En el segundo paso se realizó una evaluación de la asociación de cada una de las variables predictoras y su efecto sobre la probabilidad del desenlace de manera individual. Se calcularon Odds Ratio (OR) como medidas de asociación con sus respectivos intervalos de confianza del 95% y sus valores de p. Teniendo en cuenta que se trata de un estudio de validación externa de escalas ya validadas, todas las variables fueron incluidas para el ajuste de sus medidas se asociación en el modelo multivariado. A pesar de esto se realizó el análisis bivariado con los valores de p con la intención de evaluar el cumplimiento del criterio estadístico para ingresar al modelo multivariado ( $p < 0.2$ ).

El tercer paso consistió en el ajuste de las medidas “crudas” de los OR entre los predictores y la probabilidad de mortalidad a través de la creación de un modelo de regresión logística multivariada. Para este caso se obtuvieron OR ajustados por el efecto de confusión de todas las variables incluidas en el modelo con sus respectivos intervalos de confianza del 95% y valores de p.

Una vez obtenidos las medidas de asociación ajustadas se procedió a evaluar la capacidad de discriminación del modelo (paso 4). Esta se refiere a su capacidad para diferenciar entre los pacientes que tienen alto y bajo riesgo de presentar el desenlace. Se evalúa mediante la construcción de curvas de características operativas del receptor (ROC del inglés: *Receiver Operating Characteristic*) y su respectiva área bajo la curva (AUC del inglés: *Area Under Curve*) o estadístico C con sus respectivos intervalos de confianza del 95%. Su interpretación depende de puntos de corte previamente establecidos que dan traducción calificación cualitativa del estadístico de la siguiente manera (66): Muy mala capacidad de discriminación (AUC 0.5-0.6), mala (AUC: 0.6-0.7), aceptable (AUC: 0.7-0.8), buena (AUC: 0.8-0.9) y excelente (AUC: 0.9-1).

El quinto paso consistió en la evaluación de la calibración de cada modelo. Se realizó con la intención de cuantificar el nivel de acuerdo que existe entre las probabilidades del evento predichas por cada modelo y las proporciones observadas en la cohorte de validación. Para esto se utilizó una prueba estadística de “bondad de ajuste” de Hosmer-Lemeshow que resultó en un valor de  $p$ . Su interpretación se realiza teniendo en cuenta su hipótesis nula que afirma que no existen diferencias entre los valores predichos y los observados. De esta manera los valores de  $p > 0.05$  permiten aceptar la  $H_0$  y se puede afirmar que el ajuste del modelo no se debe al azar y así valorar su calibración (67).

Adicionalmente se realizaron gráficos de calibración con la intención de obtener mayor información además de la aportada por la prueba estadística. Estos corresponden a diagramas de dispersión entre las probabilidades predichas por el modelo (en el eje de las X con valores de 0-1) y las observadas en la población en la Y (con las mismas unidades). Los valores se comparan con una línea de “perfecta calibración” o de referencia, que corresponde a una de 45 grados entre los ejes (64,68). Se agruparon los pacientes por deciles de predicción con la intención de generar una representación gráfica de la prueba de Hosmer-Lemeshow y así complementar su interpretación.

Para agregar información a la evaluación visual se utilizaron dos medidas adicionales: el de la “calibración en promedio” conocida en inglés como “*Calibration in the large (CITL)*” y la pendiente de calibración de una curva trazada sobre los puntos en el gráfico. La primera se obtuvo con la diferencia entre la tasa general de eventos en la población y el promedio de la probabilidad obtenida por el modelo. Tiene un valor objetivo de 0 (referente de buena calibración promedio) de manera que valores negativos sugieren una sobreestimación sistemática del modelo (tasa general de eventos < promedio de probabilidad según el modelo) y viceversa. La segunda corresponde a la pendiente de una curva trazada sobre los puntos basada en una estrategia de regresión llamada “LOWESS” del inglés: *Locally Weighted Scatterplot Smoothing* o diagrama de dispersión suavizado. De esta manera una pendiente <1 sugiere que las estimaciones son extremas (muy altas para las personas con alto riesgo y muy bajas para aquellas con bajo riesgo) mientras que una >1 sugiere estimaciones “moderadas” (muy altas en personas de bajo riesgo y muy bajas para los de alto riesgo) (64,69).

Posteriormente a cada uno de los pacientes que contaba con la información completa disponible, se le realizó el cálculo de los puntajes (o probabilidades) de



cada modelo. De igual manera se le asignó el nivel de riesgo que le corresponde según su puntaje o probabilidad teniendo en cuenta los puntos de corte propuestos por cada uno de los estudios originales. Con estos valores se realizaron nuevamente los pasos 3-5 para evaluar el desempeño de las escalas derivadas de los modelos, utilizando sus puntajes.

Finalmente se realizó una comparación de los diferentes componentes de la evaluación de desempeño y se eligió la escala que podría tener mejor aplicación en escenarios de práctica clínica local. Para este caso se calcularon los diferentes valores de sensibilidad, especificidad y valores predictivos en los diferentes puntos de corte de la escala. Según el potencial escenario de uso se seleccionó el umbral para la toma de decisiones clínicas que podría beneficiar a los usuarios de las escalas en la toma de decisiones.

#### **4.4.6. Metodología para la evaluación del desempeño de las escalas en términos de su capacidad de discriminación y calibración en los pacientes sospechosos de la cohorte local:**

Todo el proceso realizado para la evaluación del desempeño de cada uno de los modelos y sus respectivas escalas en la población de pacientes con diagnóstico confirmado de COVID-19 se replicó en la población de pacientes con resultado de RT-PCR negativa. De esta manera se realizó un análisis univariado, bivariado y multivariado de la asociación entre los predictores y la probabilidad del evento fatal. Así mismo se utilizaron curvas ROC y sus respectivas áreas bajo la curva para evaluar la capacidad de discriminación de los modelos y sus escalas. Se realizaron también análisis estadísticos y sus representaciones gráficas para la evaluación de la calibración de los modelos en esta población.

La validación de las dos bases de datos digitadas, el procesamiento de las variables y el análisis se efectuó en el software STATA/MP V 14.0 y en R en su versión 4.5.0. La codificación de la base de datos validada resultante, se muestra en el Anexo 4.

## 5. CONSIDERACIONES ÉTICAS

El proyecto RESCATE COVID-19 ha sido clasificado como “*investigación sin riesgo*” de acuerdo con la resolución 8430 de 1993 del Ministerio de Salud. Emplea principalmente técnicas y métodos documentales de recolección retrospectiva de datos y no contempla la realización de alguna intervención o modificación intencionada de variables sobre los pacientes que acepten participar. Únicamente se realizará una orientación basada en evidencia a los pacientes según los hallazgos derivados del contacto telefónico y la condición actual de salud que reporte.

La información obtenida por seguimiento telefónico no se utilizará para los propósitos de este estudio ya que requiere únicamente información obtenida de la revisión de historias clínicas de manera retrospectiva. Sin embargo se aclara que el proceso de contacto telefónico se ha realizado con la guía de un flujograma que sigue los lineamientos y recomendaciones actuales tanto nacionales (Ministerio de salud y de protección social) como internacionales (Organización Mundial de la Salud y los Centros para el Control y la Prevención de Enfermedades - CDC) para la atención por COVID-19 (70). La participación del paciente se condiciona a la aceptación de un consentimiento informado verbal que se realiza de manera telefónica siguiendo un protocolo descrito en el manual de operaciones del estudio.

El proyecto fue sometido a la aprobación del comité de investigaciones y de ética de cada una de las entidades hospitalarias que funcionan como centros del estudio. En el caso de la Clínica Chicamocha los dos comités dieron aval para la recolección de la información (Anexo 5) una vez todos los participantes del equipo de investigación firmaran una declaración de confidencialidad (Anexo 5) y presentaran las certificaciones de aprobación de un curso de buenas prácticas clínicas en investigación (Anexo 6).

Los datos recolectados en los componentes están siendo centralizados en la plataforma REDCap, a la cual tienen acceso con usuario y contraseña únicamente profesionales asignados del centro coordinador del proyecto (FCI-IC). Cada usuario puede observar los registros realizados en su centro, pero no podrá tener acceso a la información de otros centros. El almacenamiento, custodia y tratamiento de los datos obtenidos a partir de las herramientas diseñadas para el proyecto, está a cargo principalmente del centro coordinador (FCI-IC) y se maneja según la ley sobre datos personales (Ley 1581 de 2012). Al realizar la exportación de datos desde las plataformas, el comité de administración de datos garantizó que estos archivos

estuvieran anonimizados para garantizar la privacidad y confidencialidad de la información en todo momento. En ningún momento fueron revelados datos que puedan identificar a los participantes.

El proceso de validación de las escalas se considera un sub estudio anidado a la cohorte rescate en el que se utilizó información de los pacientes sin vincularlos a ningún identificador personal. La información que se derive de los productos de la investigación únicamente será utilizada con fines académicos y mantendrá la confidencialidad de los datos

## 6. RESULTADOS

### 6.1. Proceso de evaluación del desempeño de la escala Zhang-China al aplicarlo a la cohorte RESCATE COVID-19:

#### 6.1.1. Pasos 1-3 del proceso de evaluación del desempeño del modelo: Análisis univariado – bivariado y modelo de regresión logística múltiple:

El resultado de los primeros tres pasos de la evaluación de los modelos se resume en la tabla 3. En cuanto al componente descriptivo, la tabla permite realizar un contraste entre las diferencias de las medianas (o proporciones) de las variables predictoras según la presencia o no del evento. Al compararse con quienes no presentaron el evento, en el grupo de fallecidos hubo una mayor proporción de hombres (62-50%), con mayor edad (72-43 años), recuento de neutrófilos (8.1-5.5 celx10<sup>3</sup>/L), valores de creatinina (1.2-0.9 mg/dL) y de proteína C reactiva (14.7-4.2 mg/dL). Sus recuentos de linfocitos (0.86-1.2 celx10<sup>3</sup>/L) y de plaquetas (209-236 celx10<sup>3</sup>/L) fueron menores.

En el análisis bivariado todos los predictores mostraron asociación estadísticamente significativa con la probabilidad de mortalidad intrahospitalaria. Los hombres tuvieron un 60% más probabilidad de fallecer respecto a las mujeres. La probabilidad de fallecer en pacientes de 47 años fue de 7.1% que aumentó de manera significativa en un 6% por cada año de diferencia en la población. El aumento en el recuento de neutrófilos, creatinina y PCR tuvo una relación directamente proporcional con el la probabilidad del evento que aumentó en 2, 19 y 8% respectivamente por cada unidad de cambio en la variable. Mientras tanto el recuento de linfocitos y de plaquetas mostraron una relación inversa: por cada unidad de cambio en aumento de sus valores (10<sup>3</sup>/L), la probabilidad de fallecer disminuyó en 44% y 1% respectivamente.

El modelo multivariado se realizó con todos los predictores originales del estudio. Teniendo en cuenta que la PCR fue la variable con menor información disponible en la base de datos (73.46% de datos no disponibles) el modelo se realizó únicamente con 317 pacientes. Tras el ajuste por las demás variables solamente la edad y la PCR mantuvieron una asociación significativa con la probabilidad del evento. La dirección de esta asociación se mantuvo y la magnitud disminuyó respecto al valor crudo. Con este modelo se concluye que al aplicarse a la cohorte RESCATE COVID-19, la edad y la PCR son los factores que aumentan de manera

independiente la probabilidad de muerte intrahospitalaria. De esta manera por cada año de edad y por cada 1mg/dL de diferencia entre pacientes la probabilidad de fallecer durante la hospitalización aumenta en 3 y 7% respectivamente, independientemente del sexo, recuento de neutrófilos, linfocitos, plaquetas y de los valores de creatinina. Los modelos de regresión logística y su ecuación de la regresión "logit" se encuentran disponibles en el Anexo 10.

**Tabla 3. Caracterización de los predictores propuestos por el modelo Zhang-China y evaluación de su asociación con mortalidad intrahospitalaria en la cohorte RESCATE COVID-19:**

Predictor	Descripción		Análisis Bivariado (Regresión Logística Simple)			Análisis Multivariado ¶ (Regresión Logística Múltiple)				
	General n= 1342	Mortalidad		Prob. en la mediana (%)	OR crudo (IC 95%)	p	B	EE	OR ajustado (IC 95%)	p
		No n = 1145	Si n = 171							
Edad (años)	47 (31 - 63)	43 (30 - 59)	72 (60 - 83)	7.1	1.06 (1.05-1.07)	<0.01*	0.033749	0.0109636	1.03 (1.01 - 1.06)	<0.01*
Sexo † (% hombres)	694 (51.6)	587 (50.1)	106 (61.9)	10.0 ‡	1.62 (1.17 - 2.25)	<0.01*	-0.0480235	0.3788071	0.95 (0.45 - 2.00)	0.899
Rec. Neutrófilos (10 <sup>3</sup> /L)	5.7 (3.8 - 8.8)	5.5 (3.6 - 7.9)	8.1 (4.9 - 12.5)	17.2	1.02 (1.00 - 1.04)	0.01*	-0.0173325	0.0226546	0.98 (0.94 - 1.02)	0.444
Rec. Linfocitos (10 <sup>3</sup> /L)	1.1 (0.8 - 1.7)	1.2 (0.85 - 1.73)	0.86 (0.58 - 1.31)	18.8	0.56 (0.42-0.73)	<0.01*	0.0175533	0.0866529	1.01 (0.86 - 1.20)	0.839
Rec. Plaquetas (10 <sup>3</sup> /L)	232 (179 - 297)	236 (186 - 300)	209 (156 - 282)	18.3	0.99 (0.995- 0.999)	<0.01*	-0.0000377	0.0020441	0.99 (0.99 - 1.00)	0.985
Creatinina (mg/dL)	0.9 (0.7 - 1.2)	0.9 (0.7 - 1.1)	1.2 (0.87 - 1.8)	17.8	1.19 (1.09 - 1.31)	<0.01*	0.1794799	0.0952621	1.19 (0.99 - 1.44)	0.060
PCR (mg/dL)	4.9 (1.5 - 14.1)	4.2 (1.3 - 11.6)	14.7 (5.6 - 25.4)	8.1	1.08 (1.05 - 1.11)	<0.01*	0.0690812	0.0109636	1.07 (1.03 - 1.10)	<0.01*
<b>Bondad de ajuste Hosmer-Lemeshow: p = 0.0027</b>										

Los valores de laboratorio corresponden a los medidos al ingreso a urgencias.

Para todas las variables continuas se reporta mediana y el rango intercuartílico.

B: Coeficiente (B1), EE: Error Estándar, Prob: Probabilidad.

La descripción de cada predictor se calcula teniendo en cuenta sus datos perdidos

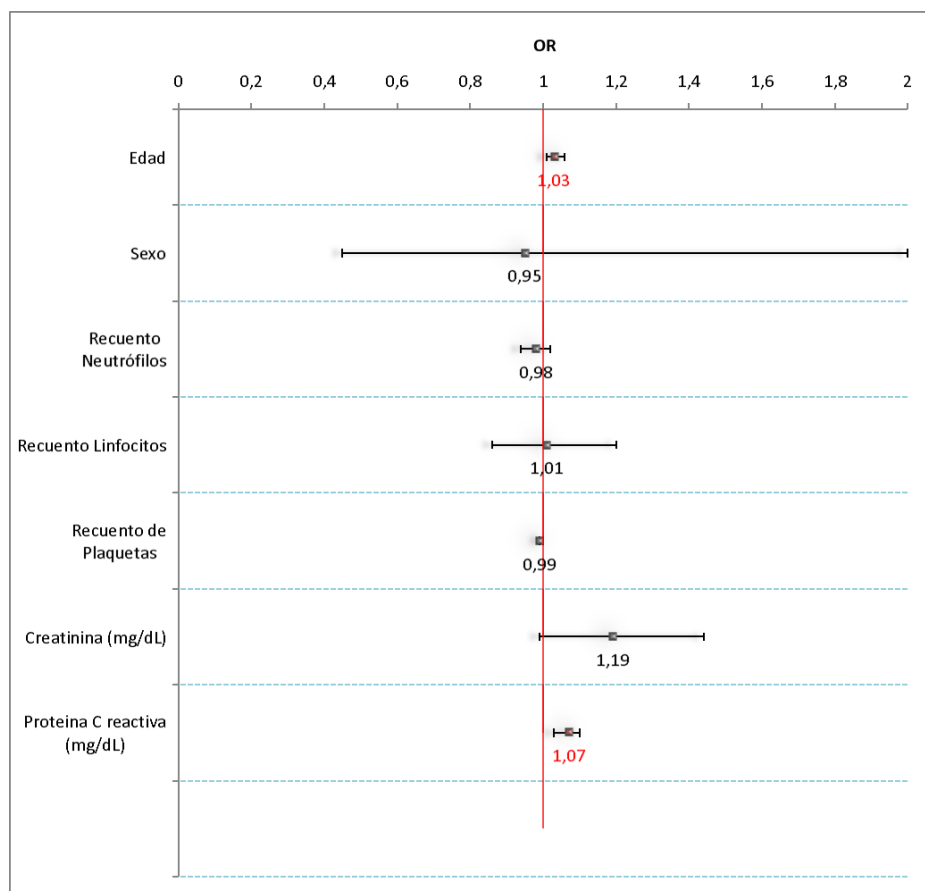
\* Asociación con significancia estadística p<0.05

† Para sexo se reporta conteo y proporción en porcentaje (%)

‡ Probabilidad en mujeres

¶ Número de observaciones en el modelo multivariado: 317

**Figura 4. Gráfico de bosque del análisis multivariado de la escala Zhang-China en la cohorte RESCATE COVID-19**



En rojo: Variables con asociación significativa tras ajuste por confusión.

En gris: Variables con asociación no significativa.

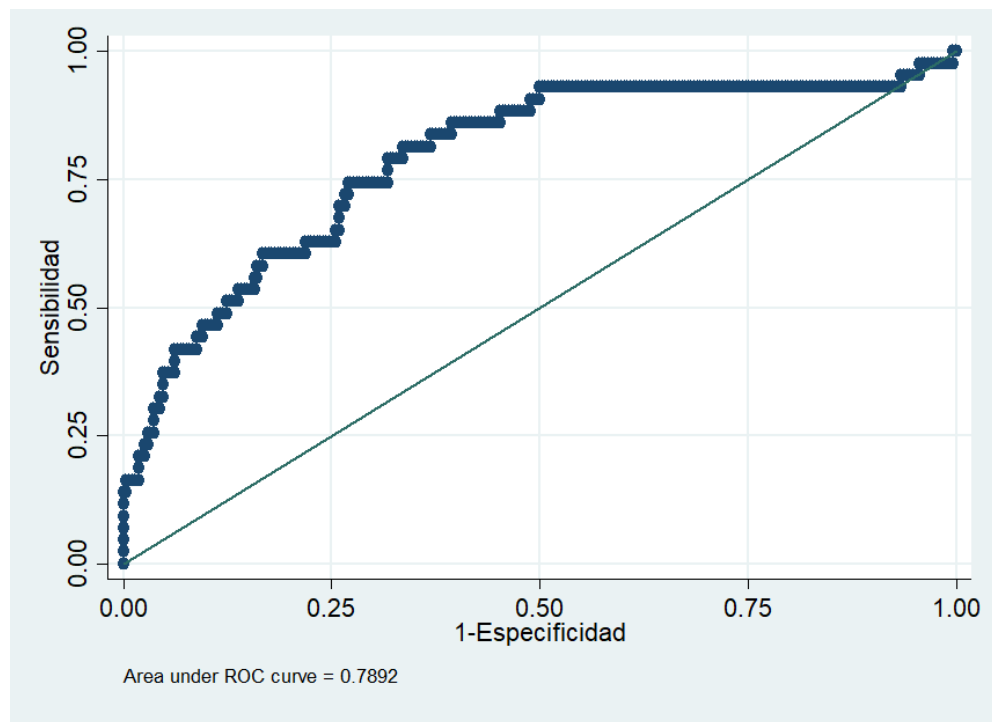
El tamaño de la figura central no representa ningún valor estadístico y se utiliza como referencia del estimado central de la asociación: OR

#### **6.1.2. Paso 4 de la evaluación del desempeño del modelo: Capacidad de discriminación.**

Se realizó una curva receptor operador (ROC) según las probabilidades predichas por el modelo y la observada en la cohorte en la que se calcula la capacidad de discriminación con diferentes puntos de corte (Figura 3). Se observa que en la proporción de individuos que fallecieron la probabilidad predicha por el modelo fue

mayor que la de aquellos que egresaron vivos el 78.9% de las veces (AUC: 0.789). En otras palabras, de una muestra de parejas tomadas al azar de un paciente fallecido y uno que egresó vivo, este modelo le dará una probabilidad mayor al primero que al segundo en 78 de cada 100 pares de elecciones. Su área corresponde a una capacidad de discriminación aceptable.

**Figura 5. Curva ROC del modelo Zhang-China para predicción de mortalidad intrahospitalaria en la cohorte de validación externa:**



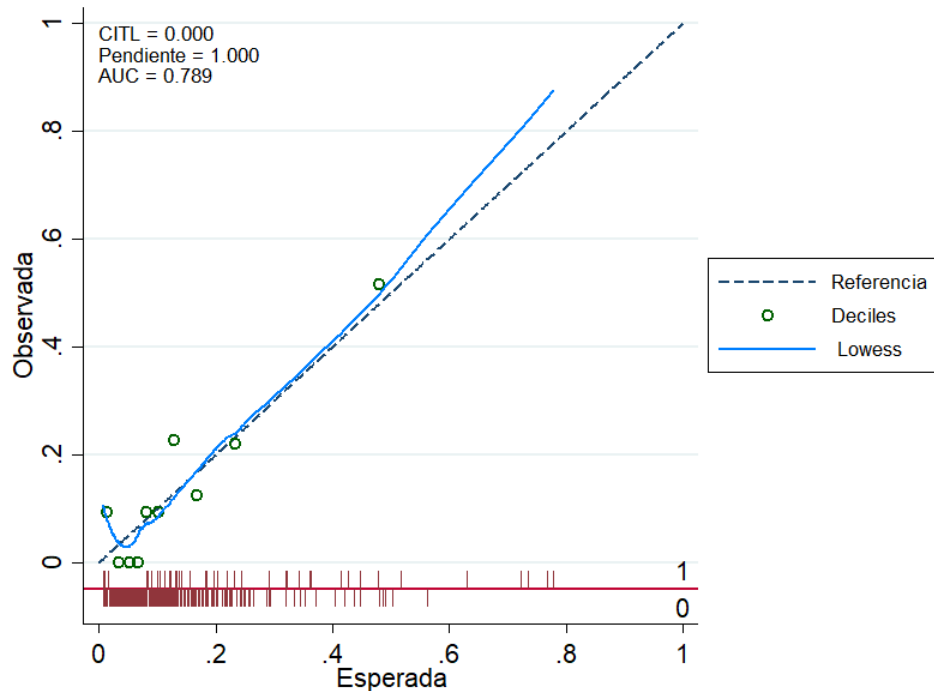
### **6.1.3. Paso 5 de la evaluación del desempeño del modelo: Calibración.**

Los estadísticos realizados corresponden con ausencia de calibración de los datos a las predicciones del modelo. La prueba de Bondad de ajuste Hosmer-Lemeshow resultó en un valor de  $p < 0.05$  ( $p = 0.0027$ ) que rechaza la hipótesis nula. A pesar de ausencia estadística de ajuste, la calibración en promedio (CITL) fue muy cercana a 0 y la pendiente de la curva muy cercana a 1. En la valoración gráfica se observa que en los deciles bajos de probabilidad el modelo tiende a sobre estimar mientras



que en los deciles más alto a subestimar. Teniendo en cuenta todos los criterios para la evaluación se considera mala calibración por estadísticos y análisis gráfico.

**Figura 6. Gráfico de calibración entre probabilidades esperadas por el modelo Zhang-China y las observadas en la cohorte de validación según deciles de probabilidad esperada:**



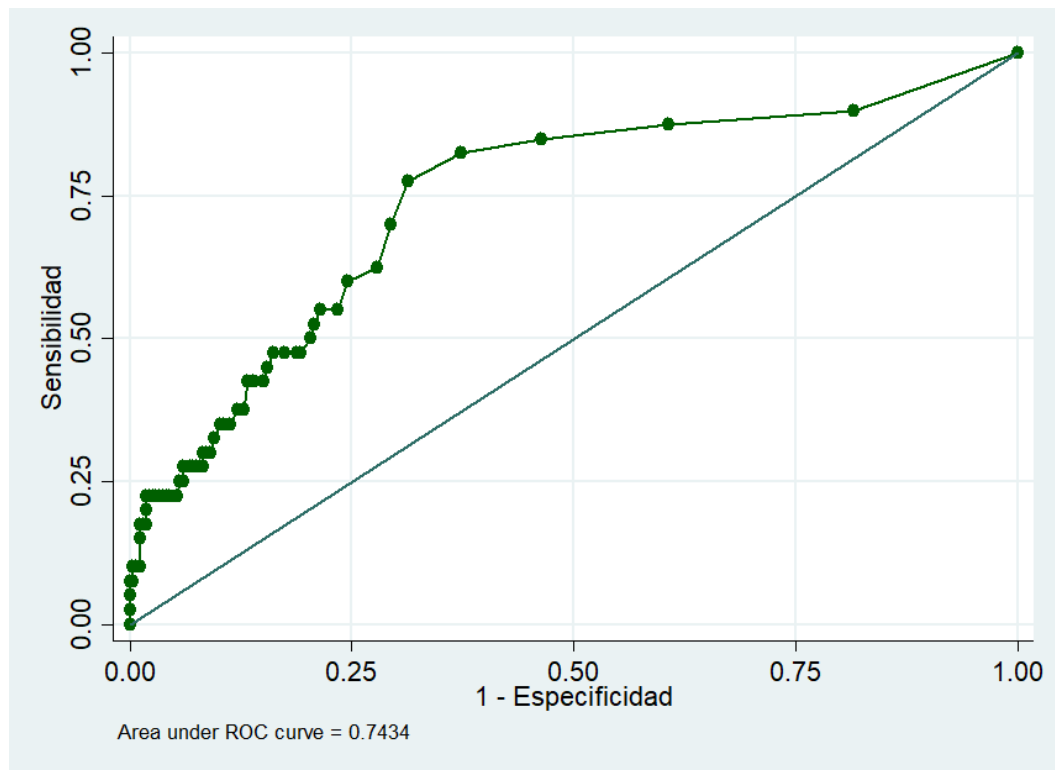
Prueba de Bondad de ajuste Hosmer-Lemeshow:  $p = 0.0027$

#### **6.1.4. Evaluación del desempeño de la escala Zhang-China según probabilidades calculadas en plataforma digital:**

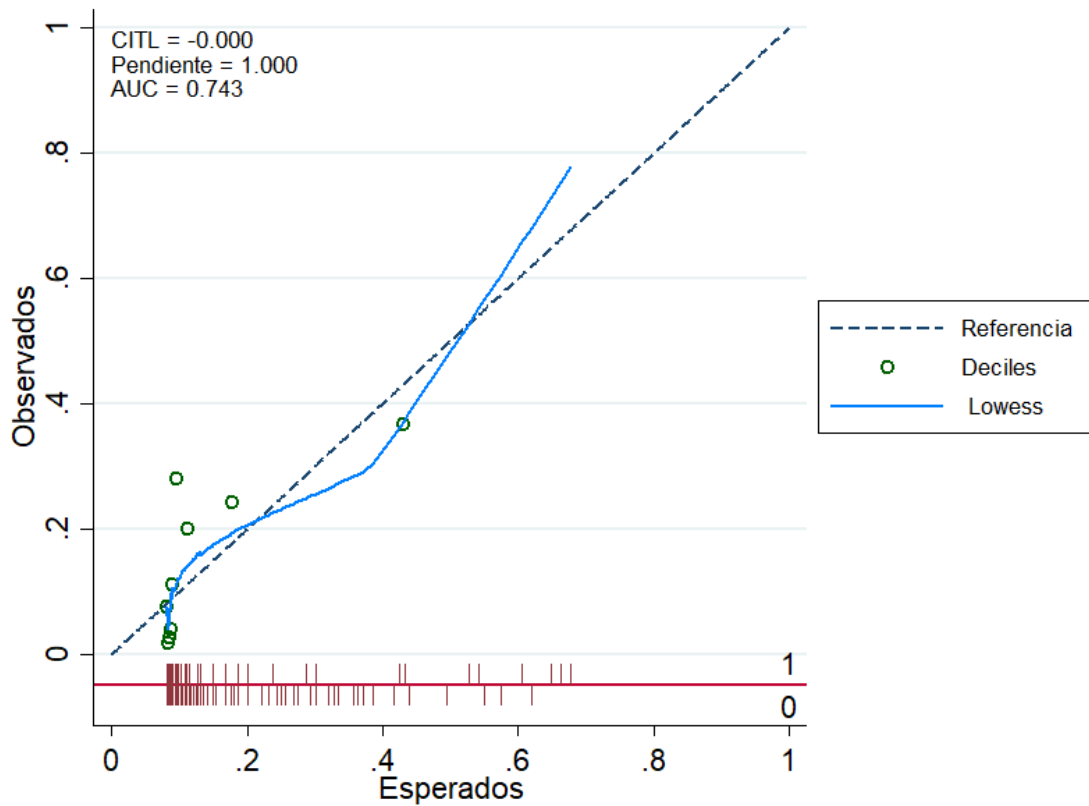
A cada uno de los 317 pacientes en los que se realizó la evaluación del modelo se le calculó la probabilidad del evento utilizando la plataforma digital que ofrecen los autores del estudio. Posteriormente se generó una variable que contiene este valor para cada paciente y se realizó un modelo de regresión logística en la que se evalúa su asociación con el desenlace. Se obtuvo un OR de 1.03 (IC95%: 1.01 – 1.04) con  $p < 0.01$  que se considera una asociación positiva significativa entre los puntajes calculados por la escala original y la probabilidad de fallecer en la cohorte de validación (Anexo 10). Sin embargo, el área bajo la curva disminuyó respecto a la

evaluación del modelo original (AUC: 0.74) logrando apenas una capacidad aceptable de discriminación (Gráfico 4). Este modelo generado con los puntajes de la escala tampoco muestra ajuste estadístico de los datos (prueba de HL:  $p=0.008$ ). La calibración se vio deteriorada al compararse con la evaluación del modelo. La CITL y la pendiente se mantuvieron cercanos a sus valores ideales, pero en la evaluación gráfica es fácil observar una alta dispersión de los deciles y una curva varía en sobre y sub estimación. Se considera que la utilización de la escala digital de Zhang y colaboradores tiene una pobre capacidad de discriminación y calibración al aplicarse a la cohorte local.

**Figura 7. Curva ROC del modelo Zhang-China para predicción de mortalidad intrahospitalaria en la cohorte de validación externa según los puntajes de la escala:**



**Figura 8. Gráfico de calibración entre probabilidades esperadas por los puntajes de la escala de Zhang-China y la probabilidad observada en la cohorte de validación:**



**Prueba de bondad de ajuste de Hosmer-Lemeshow:  $p=0.0081$**

### **6.1.5. Evaluación del desempeño de la escala Zhang-China según los niveles de riesgo obtenidos a partir de los puntajes:**

Se generó una variable categórica politómica con los niveles de riesgo según los puntajes de la escala propuestos por los autores. La variable tiene 4 niveles siendo la primera “muy bajo riesgo” (probabilidad <0.6%), seguida por “bajo riesgo” (0.6-8.9%), “alto riesgo” (8.9-58.5%) y “muy alto riesgo” (>58.5%). Se realizó un modelo de regresión logística múltiple utilizando los diferentes niveles como variables indicadoras. De esta manera se obtuvieron OR como medidas de asociación que compara la probabilidad del evento en cada nivel de riesgo con el nivel más bajo (“muy bajo riesgo”).

**Tabla 4. Modelo de regresión logística múltiple (variables indicadoras) entre niveles de riesgo obtenidos a partir de los puntajes calculados por la escala Zhang-China y mortalidad intrahospitalaria en la cohorte RESCATE-COVID-19.**

Descripción			Regresión Logística ‡			
Predictor	Vivos al egreso		B	EE	OR (IC 95%)	p
	Si n= 265 (%)	No n= 40 (10.5%)				
<b>Muy bajo riesgo (≤ 0.6%)</b>	49 (92.5)	4 (7.5)	-2.50552†	0.5200078	Ref.	Ref.
<b>Riesgo bajo (0.6 – 8.9%)</b>	154 (91.7)	14 (8.3)	0.1076307	0.5901951	1.11 (0.35 – 3.54)	0.855
<b>Riesgo Alto (8.9 – 58.5%)</b>	53 (80.3)	13 (19.7)	1.100183	0.6051439	3.00 (0.92 – 9.83)	0.069
<b>Muy alto riesgo (&gt; 58.5%)</b>	9 (50.0)	9 (50.0)	2.505526	0.7018763	12.25 (3.09 – 48.48)	<0.01*
<b>Total</b>	305		<b>Prueba de Hosmer-Lemeshow: p = 1</b>			

B: Coeficiente (B1), EE: Error Estándar.

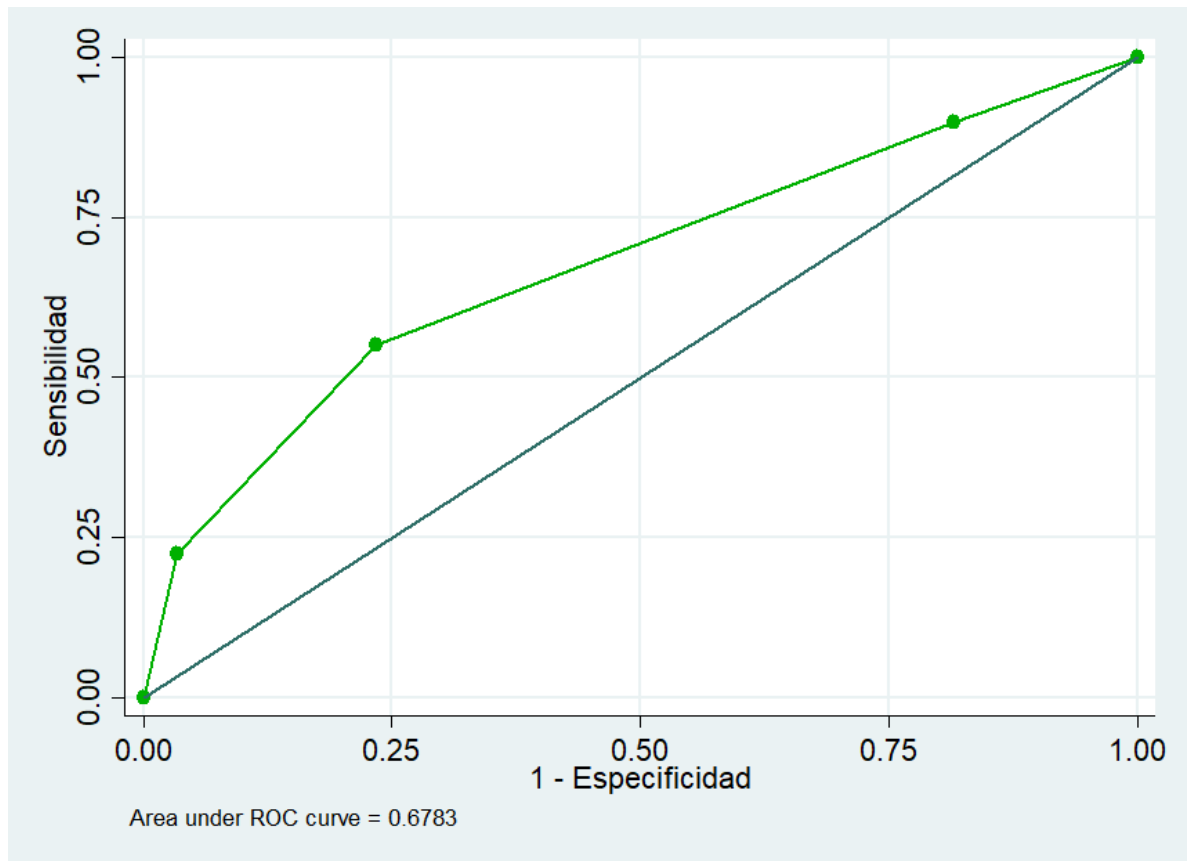
\* Asociación con significancia estadística  $p < 0.05$  tras ajustar efecto de covariables en el modelo.

† Probabilidad de morir en el nivel de muy bajo riesgo es de **16% (Probabilidad para el intercepto)**.

‡ Número de observaciones en el modelo: **305**

Los dos primeros niveles de riesgo no muestran asociaciones significativas con la probabilidad de fallecer al compararse con el nivel de muy bajo riesgo. En el último nivel se obtuvo una asociación positiva significativa con intervalos de confianza muy anchos. El área bajo la curva ROC resultante del modelo es de 0.67 considerándose como mala capacidad para discriminar entre pacientes con alto vs bajo riesgo (Figura 7). La prueba de bondad de ajuste dio un valor de  $p = 1$  que se interpreta como ajuste estadístico al modelo. Teniendo en cuenta su mala capacidad de discriminación y el mal desempeño del modelo se considera que la utilización de la escala digital de Zhang y colaboradores no tiene un buen desempeño en la cohorte local (al analizar puntajes y sus niveles de riesgo).

Figura 9. Curva ROC del modelo Zhang-China para predicción de mortalidad intrahospitalaria en la cohorte de validación externa según los niveles de riesgo:



Por último, se realizó un análisis de la capacidad de discriminación del modelo retirando la variable PCR teniendo en cuenta la elevada proporción de valores perdidos. Al retirar esta variable se logra realizar el análisis en 823 pacientes. A pesar de que el modelo de regresión logística múltiple resulta con más número de predictores con asociación estadísticamente significativa, su capacidad discriminación es prácticamente la misma (AUC: 0.78) al igual que sus medidas de calibración (Anexo 9)

### **6.1.6. Conclusiones del análisis del desempeño de la escala Zhang-China al aplicarla a la cohorte RESCATE COVID-19:**

Se considera que la evaluación global del modelo y la escala de Zhang tiene un mal desempeño al aplicarse en la cohorte RESCATE COVID-19. En el modelo únicamente dos variables mantienen una medida de asociación significativa tras ser ajustadas por el efecto de todas las variables. La capacidad de discriminación del modelo es aceptable, pero al utilizar la escala disponible en línea para calcular puntajes y establecer niveles de riesgo, esta empeora alcanzando un AUC <0.7. De igual forma no se cumplen criterios estadísticos ni visuales de buena calibración con el modelo. A pesar de que la calibración parece mejorar al evaluar los niveles de riesgo según los puntajes obtenidos en línea, la baja capacidad de discriminación y la baja prevalencia de desenlaces en las categorías de riesgo parecen indicar que dicho hallazgo carece de validez y utilidad clínica.

### **6.2. Proceso de evaluación del desempeño de la escala Xie-China al aplicarlo a la cohorte RESCATE COVID-19:**

#### **6.2.1. Pasos 1-3 del proceso de evaluación del desempeño del modelo: Análisis univariado – bivariado y modelo de regresión logística múltiple:**

El modelo propuesto por Xie y colaboradores contiene 4 variables, todas de naturaleza continua, como predictores de mortalidad. El resultado de los primeros tres pasos de la evaluación de los modelos se resume en la tabla 5. Dos de los predictores tuvieron una mediana superior en el grupo de pacientes con el desenlace respecto al grupo de pacientes que egresaron vivos (edad: 72-43 años y LDH: 511-353 U/L) mientras que los restantes se comportaron de manera inversa (linfocitos: 0.86-1.2  $10^3/L$  y oximetría al ingreso: 88-93%).

En el análisis bivariado todos los predictores mostraron asociación estadísticamente significativa con la probabilidad de mortalidad intrahospitalaria. Las medidas de asociación entre la edad y el recuento de linfocitos ya se conocían desde el modelo anterior. En cuanto a la LDH los pacientes con 384 U/L (mediana) tuvieron un 16% de probabilidad de presentar el desenlace. Se observó una asociación positiva en la que por cada unidad de diferencia en U/L se aumentó un 0.3% la probabilidad de

morir. Por su parte la asociación entre el porcentaje de oxigenación al ingreso y la mortalidad fue inversa. En personas con 93% de saturación la probabilidad de fallecer fue del 10.4% y disminuyó un 7% por cada punto porcentual de diferencia.

El modelo multivariado se realizó con todos los predictores originales del estudio. En este caso la variable LDH fue la de menor información disponible en la base de datos (53.9% de datos no disponibles) razón por la cual el modelo se realizó en 593 pacientes. Todos los OR ajustados obtenidos con el modelo de regresión logística resultaron estadísticamente significativos. Los estimados del efecto mantuvieron su dirección y su magnitud disminuyó respecto a sus valores crudos. La variable con mayor magnitud de asociación con el desenlace fue el recuento de linfocitos. Independientemente de la edad, el nivel de LDH o la oximetría al ingreso, la diferencia de cada 1.000 linfocitos/L en el hemograma de ingreso resultó en una disminución del 51% en la probabilidad de fallecer durante la hospitalización.

Los modelos de regresión logística y su ecuación de la regresión "logit" se encuentran disponibles en el Anexo 10.

**Tabla 5. Caracterización de los predictores propuestos por el modelo Xie-China y evaluación de su asociación con mortalidad intrahospitalaria en la cohorte RESCATE COVID-19**

Predictor	Descripción			Análisis Bivariado (Regresión Logística Simple)			Análisis Multivariado † (Regresión Logística Múltiple)			
	General n= 1342	Mortalidad		Probab. en la mediana (%)	OR crudo (IC 95%)	p	B	EE	OR ajustado (IC 95%)	p
		No n = 1145	Si n = 171							
<b>Edad (años)</b>	47 (31 – 63)	43 (30 – 59)	72 (60 – 83)	7.1	1.06 (1.05 – 1.07)	<0.01*	0.052641	0.0081355	1.05 (1.03 – 1.07)	<0.01*
<b>LDH (U/L)</b>	384 (283 – 517)	353 (271 – 466)	511 (359 – 700)	15.9	1.003 (1.002- 1.004)	<0.01*	0.0028183	0.0005149	1.002 (1.001 – 1.003)	<0.01*
<b>Rec. Linfocitos (10<sup>3</sup>/L)</b>	1.1 (0.8 – 1.7)	1.2 (0.85 – 1.73)	0.86 (0.58 – 1.31)	18.8	0.56 (0.43 – 0.73)	<0.01*	-0.6978789	0.2322192	0.49 (0.31 – 0.78)	<0.01*
<b>Oximetría (%)</b>	93 (88 – 95)	93 (89 – 95)	88 (80 – 93)	10.4	0.93 (0.92 – 0.95)	<0.01*	-0.0217269	0.0109015	0.98 (0.96 – 0.99)	0.04*
							<b>Bondad de ajuste de Hosmer-Lemeshow: p = 0.2294</b>			

Los valores de laboratorio y saturación de oxígeno corresponden a los medidos al ingreso a urgencias.

Para todas las variables continuas se reporta la mediana y los rangos intercuartílicos.

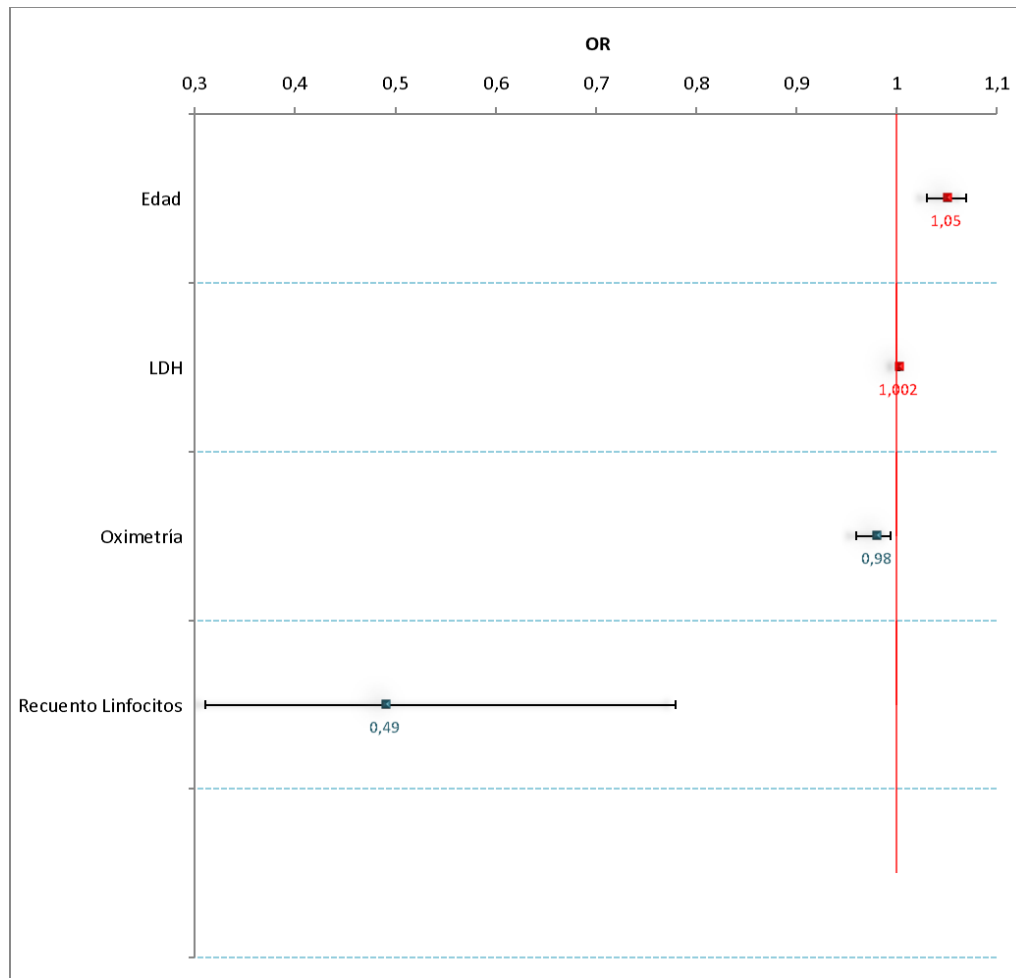
B: Coeficiente (B1), EE: Error Estándar, LDH: Lactato deshidrogenasa. Probab: Probabilidad.

\* Asociación con significancia estadística p<0.05

† Número de observaciones en el modelo multivariado: **593**



Figura 10. Gráfico de bosque del análisis multivariado de la escala Xie-China en la cohorte RESCATE COVID-19



En rojo: Variables con asociación positiva significativa tras ajuste por confusión.

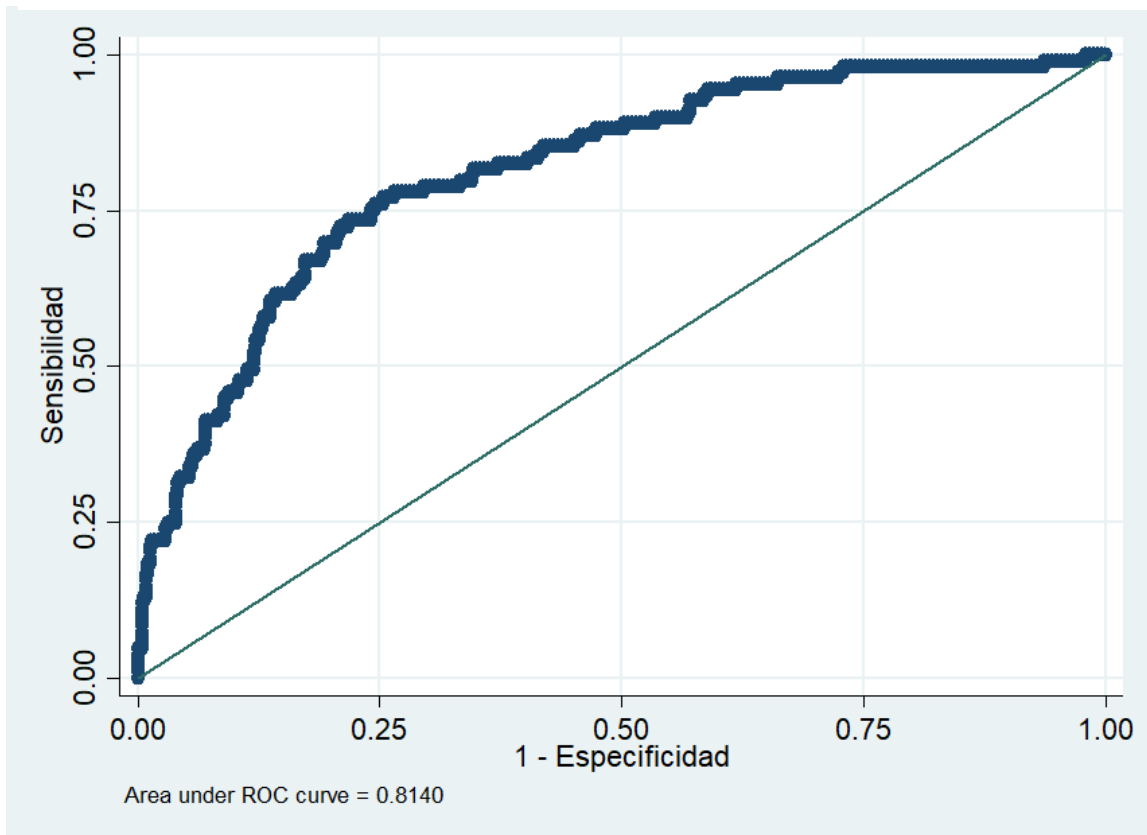
En azul: Variables con asociación negativa significativa tras ajuste por confusión.

El tamaño de la figura central no representa ningún valor estadístico y se utiliza como referencia del estimado central de la asociación: OR

### 6.2.2. Paso 4 de la evaluación del desempeño del modelo: Capacidad de discriminación.

Se realizó una curva receptor operador (ROC) según las probabilidades predichas por el modelo y la observada en la cohorte en la que se calcula la capacidad de discriminación con diferentes puntos de corte (Figura 4). Se observa que en la proporción de individuos que fallecieron la probabilidad predicha por el modelo fue mayor que la de aquellos que egresaron vivos el 81% de las veces (AUC: 0.814). En otras palabras, de una muestra de parejas tomadas al azar de un paciente fallecido y uno que egresó vivo, este modelo le dará una probabilidad mayor al primero que al segundo en 81 de cada 100 pares de elecciones. Su área corresponde a una capacidad de discriminación buena.

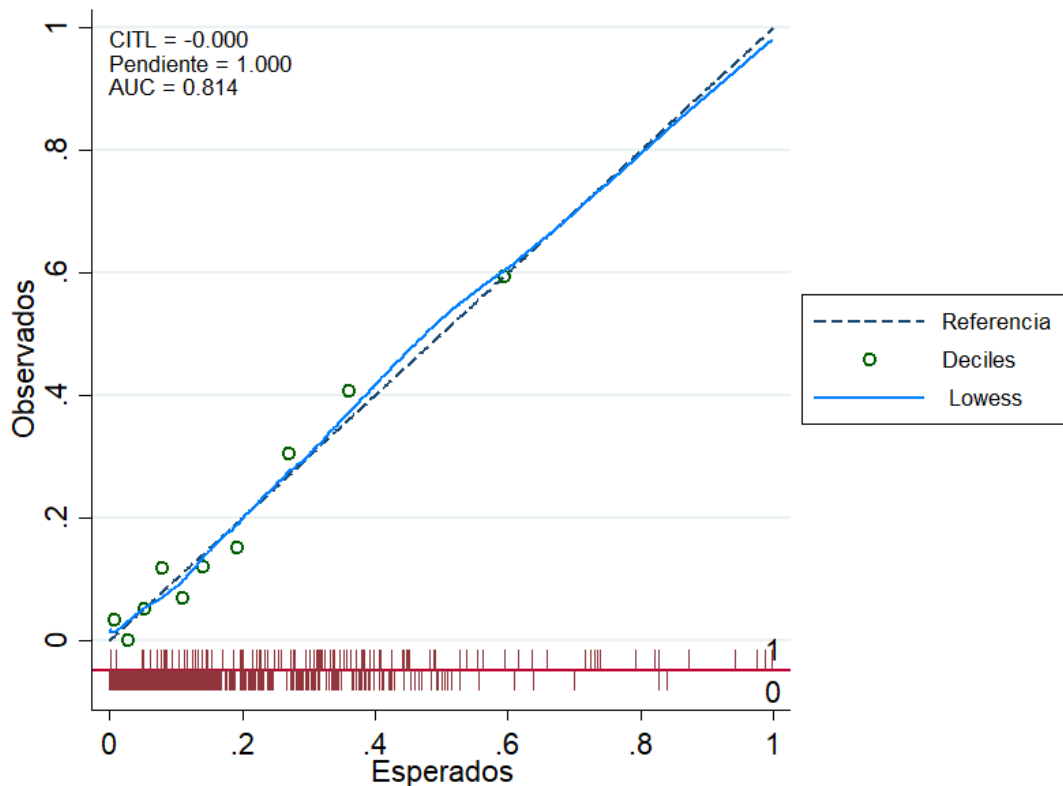
Figura 11. Curva ROC del modelo Xie-China para predicción de mortalidad intrahospitalaria en la cohorte de validación externa:



### 6.2.3. Paso 5 de la evaluación del desempeño del modelo: Calibración.

Los estadísticos realizados corresponden con una buena calibración del modelo. La prueba de Bondad de ajuste Hosmer-Lemeshow resultó en un valor de  $p > 0.05$  ( $p = 0.2294$ ) permitiendo afirmar que los datos evaluados se ajustan a las predicciones del modelo. De igual manera la calibración en promedio (CITL) fue muy cercana a 0 y la pendiente de la curva muy cercana a 1. En la valoración gráfica se observa que los diferentes deciles de probabilidad se acercan mucho a la línea de referencia y la línea de regresión que los une se parece mucho a la línea de referencia.

Figura 12. Gráfico de calibración entre probabilidades esperadas por el modelo Xie-China y las



Prueba de Bondad de ajuste Hosmer-Lemeshow:  $p = 0.2294$

#### **6.2.4. Evaluación del desempeño de la escala Xie-China según puntajes calculados:**

El sistema de puntuación propuesto por Xie y colaboradores consiste en un nomograma que refleja cada una de sus variables en dos escalas de puntos. Proponen que se utilice como herramienta en la práctica clínica para evaluar la probabilidad de muerte intrahospitalaria en pacientes con diagnóstico de COVID-19. Se intentó sin éxito aplicar la herramienta en pacientes de la cohorte RESCATE COVID-19 por diferentes factores. En primer lugar, se trata de una herramienta semicuantitativa que presenta métricas que no permiten calcular puntajes exactos. Adicionalmente requiere de recursos físicos como impresión en papel milimetrado para ayudar a una mejor cuantificación y requiere de una alta inversión de tiempo. Se consideró que su uso en la práctica clínica real puede ser muy limitado especialmente comparado con los otros dos modelos que presentan herramientas de cálculo sencillas y de fácil acceso. Por estas razones no se logró utilizar y aplicar a los pacientes de la cohorte de validación la escala de Xie-China con sus puntos de corte.

#### **6.2.5. Conclusiones del análisis del desempeño de la escala Xie-China al aplicarla a la cohorte RESCATE COVID-19:**

Se considera que la evaluación global del modelo de Xie-China tiene un buen desempeño al aplicarse en la cohorte RESCATE COVID-19. Todas sus variables mantienen una asociación significativa con la mortalidad intrahospitalaria al ajustarse en el modelo de regresión logística múltiple. Su capacidad de discriminación es aceptable (AUC 0.8) y las predicciones del modelo se ajustan estadística y gráficamente a las observadas. Sin embargo, la traducción del modelo a escenarios de práctica clínica es prácticamente imposible. El uso de la herramienta propuesta por los autores (nomograma) requiere un alto consumo de tiempo y recursos para calcular un puntaje con alto riesgo de sesgos de medición. Se tiene en cuenta que existen actualmente posibilidades de generar herramientas de fácil generalización y con menor probabilidad de error de medición (herramientas en línea, aplicaciones para dispositivos móviles, sistemas sencillos de puntos) que han hecho que en la práctica no sea usual el uso de nomogramas. Por esta razón se considera que a pesar de parecer un buen modelo, sus dificultades para la aplicación local lo hacen poco viable en nuestra población.

### **6.3. Proceso de evaluación del desempeño de la escala Bello-México al aplicarlo a la cohorte RESCATE COVID-19**

#### **6.3.1. Pasos 1-3 del proceso de evaluación del desempeño del modelo: Análisis univariado – bivariado y modelo de regresión logística múltiple:**

El modelo propuesto por Bello y colaboradores utiliza únicamente variables discretas dicotómicas codificadas como presencia/ausencia de la característica. Por esta razón en el componente descriptivo de la tabla 6 se reportan conteos y proporciones de los predictores en el grupo con y sin el desenlace. El grupo de pacientes que presentaron el evento tuvo una mayor representación de mayores de 65 años (55 – 18%), obesos (18 - 7.5%), con historia de diabetes (29 – 11%), EPOC (22 - 3.7%), ERC (12 – 3.7%) y con diagnóstico de neumonía (81 - 42%). Por su parte en el grupo de los pacientes sin evento hubo mayor representación de menores de 40 años (42 – 16%). En los dos grupos la prevalencia de inmunosuprimidos fue del 1.2%.

En el análisis bivariado siete de los 9 predictores resultaron con una asociación significativa con la probabilidad de morir durante la hospitalización. El antecedente de EPOC fue el predictor con mayor fuerza de asociación aumentando la probabilidad entre 4 y 12 veces del evento (OR: 7.7). Es seguido por el diagnóstico de neumonía y el antecedente de EPOC que muestran un aumento 5.8 y 5.3 veces la probabilidad del desenlace. La edad menor a 40 fue la única variable que mostró una relación inversa significativa con una reducción del 73%.

El modelo multivariado se realizó con todos los predictores originales del estudio a pesar de que el antecedente de diabetes diagnosticada antes de los 40 años y de inmunosupresión no cumplieran criterios estadísticos según el análisis bivariado ( $p > 0.2$ ). Se realizó de esta manera ya que se trata de un proceso de validación externa del modelo. La baja prevalencia de las dos características en la cohorte (entre 0.5-1.1%) y en los grupos dificulta la evaluación de sus contrastes. Teniendo en cuenta la proporción de datos perdidos con estas variables el modelo se aplicó a 1168 pacientes.

Tras el ajuste de los OR por las nueve variables del modelo, cuatro mantuvieron una asociación significativa con la probabilidad del evento (todas positivas). La magnitud de la asociación resultó más conservadora, pero se mantuvo la dirección

del efecto y la incertidumbre alrededor del estimado. Con este modelo se concluye que, al aplicarse a la cohorte local, los pacientes con diagnóstico de COVID-19 en quienes se realice diagnóstico de neumonía al ingreso tienen una probabilidad 4 veces mayor de fallecer durante la hospitalización que aquellos sin neumonía. De igual forma el antecedente de EPOC, la edad mayor o igual a 65 años triplican la probabilidad y la obesidad la dobla, independientemente de las otras variables. Los modelos de regresión logística y su ecuación de la regresión “logit” se encuentran disponibles en el Anexo 10.

**Tabla 6. Caracterización de los predictores propuestos por el modelo Bello-México y evaluación de su asociación con mortalidad intrahospitalaria en la cohorte RESCATE COVID-19.**

Predictor	Descripción		Análisis Bivariado (Regresión Logística Simple)			Análisis Multivariado† (Regresión Logística Múltiple)				
	General n= 1342 (%)	Mortalidad No n = 1145	Si n = 171	Prob. del comparador	OR crudo (IC 95%)	p	B	EE	OR ajustado (IC 95%)	p
Edad < 40	529 (39.3)	500 (42.7)	29 (16.9)	42.3	0.27 (0.18 – 0.41)	<0.01*	-0.15626	0.31024	0.85 (0.46 – 1.57)	0.61
Edad ≥ 65	317 (23.5)	222 (18.9)	95 (55.5)	16.0	5.34 (3.82 – 7.47)	<0.01*	1.11518	0.24683	3.05 (1.88 – 4.94)	<0.01**
Diabetes	179 (13.3)	129 (11.0)	50 (29.2)	23.2	3.32 (2.28 – 4.85)	<0.01*	0.01578	0.93829	1.01 (0.16 – 6.39)	0.98
Diabetes <40	7 (0.52)	5 (0.4)	2 (1.2)	23.2	3.44 (0.66 – 17.96)	0.14	1.46010	0.96163	4.30 (0.65 – 28.35)	0.12
Obesidad	119 (8.8)	88 (7.5)	31 (18.1)	25.9	2.71 (1.74 – 4.24)	<0.01*	0.83937	0.30123	2.31 (1.28 – 4.17)	<0.01**
Neumonía	642 (47.7)	500 (42.7)	139 (81.3)	9.5	5.82 (3.90 – 8.70)	<0.01*	1.43422	0.27217	4.19 (2.50 – 7.15)	<0.01**
ERC	64 (4.7)	43 (3.7)	21 (12.3)	26.6	3.66 (2.11 – 6.34)	<0.01*	-0.05720	0.48525	0.94 (0.36 – 2.44)	0.90
EPOC	82 (6.0)	43 (3.7)	39 (22.8)	23.5	7.72 (4.83 – 12.36)	<0.01*	1.18192	0.32634	3.26 (1.71 – 6.18)	<0.01**
Inmunosup	16 (1.1)	14 (1.2)	2 (1.2)	29.3	0.97 (0.21 – 4.32)	0.97	-0.63833	1.08021	0.53 (0.06 – 4.38)	0.55
<b>Bondad de ajuste de Hosmer-Lemeshow: p = 0.3254</b>										

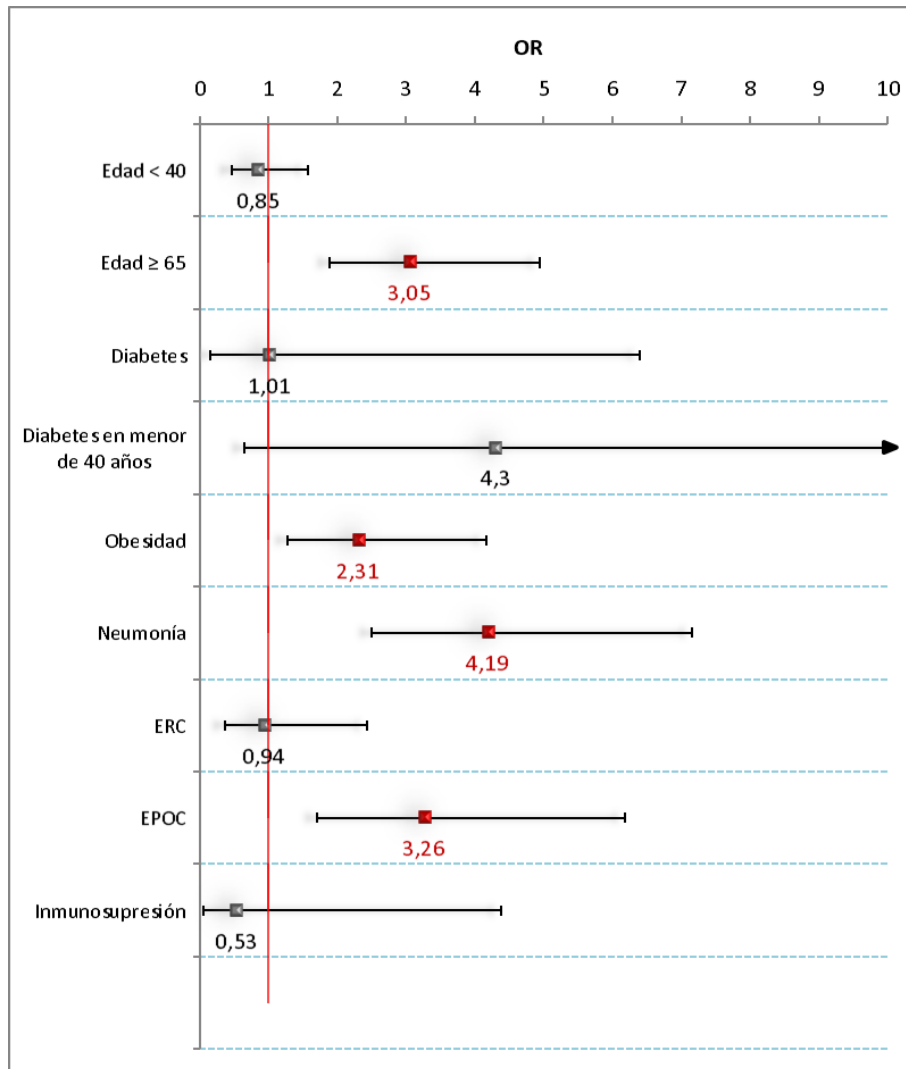
Para todas las variables se reporta su **conteo y proporción en porcentaje (%)**

B: Coeficiente (B1), EE: Error Estándar, ERC: Enfermedad Renal Crónica, EPOC: Enfermedad Pulmonar Obstrucciona crónica, Inmunosup: inmunosupresión.

\* Asociación con significancia estadística  $p < 0.05$

† Número de observaciones en el modelo multivariado: **1168**

**Figura 13. Gráfico de bosque del análisis multivariado de la escala Bello-México en la cohorte RESCATE COVID-19**



En rojo: Variables con asociación positiva significativa tras ajuste por confusión.

En gris: Variables con asociación no significativa.

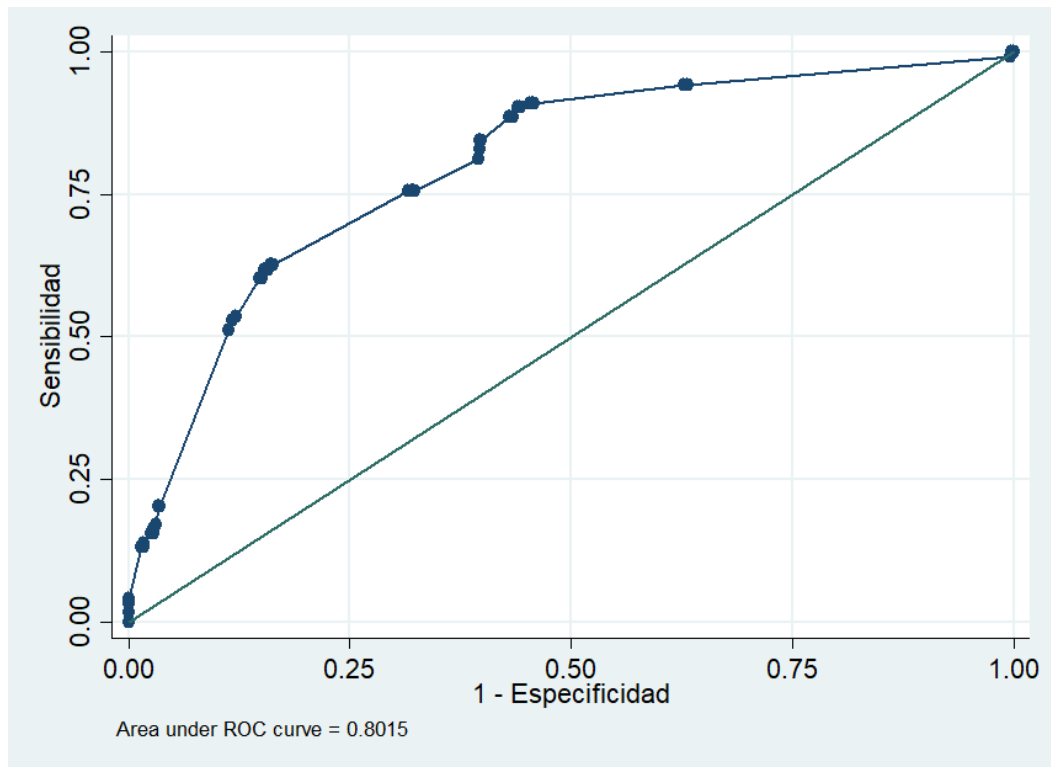
El tamaño de la figura central no representa ningún valor estadístico y se utiliza como referencia del estimado central de la asociación: OR.



### 6.3.2. Paso 4 de la evaluación del desempeño del modelo: Capacidad de discriminación.

Se realizó una curva receptor operador (ROC) según las probabilidades predichas por el modelo y la observada en la cohorte en la que se calcula la capacidad de discriminación con diferentes puntos de corte (Gráfico 8). Se observa que en el grupo de pacientes que fallecieron la probabilidad predicha por el modelo fue mayor que la de aquellos que egresaron vivos el 80% de las veces (AUC: 0.789). En otras palabras, de una muestra de parejas tomadas al azar de un paciente fallecido y uno que egresó vivo, este modelo le dará una probabilidad mayor al primero que al segundo en 80 de cada 100 pares de elecciones. Su área corresponde a una buena capacidad de discriminación.

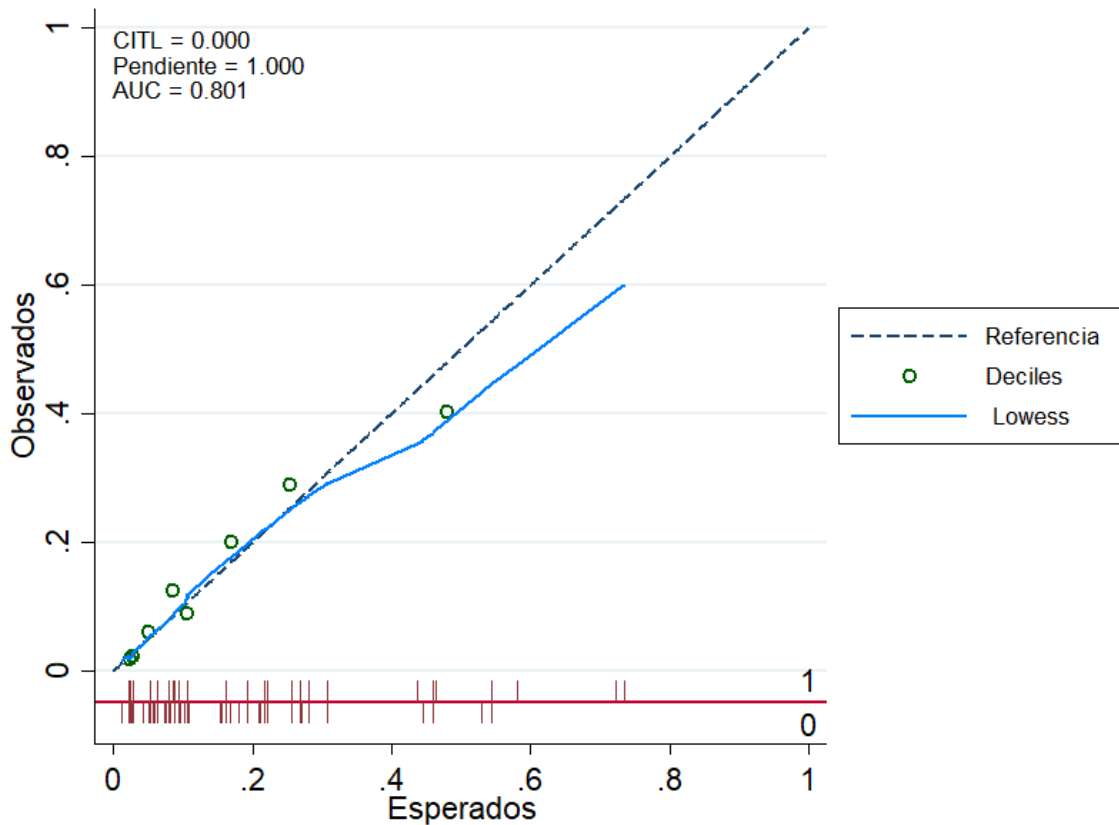
Figura 14. Curva ROC del modelo Bello-México para predicción de mortalidad intrahospitalaria en la cohorte de validación externa



### 6.3.3. Paso 5 de la evaluación del desempeño del modelo: Calibración.

Los estadísticos realizados corresponden con una buena calibración del modelo. La prueba de Bondad de ajuste Hosmer-Lemeshow resultó en un valor de  $p > 0.05$  ( $p = 0.3254$ ) que confirma un ajuste estadístico del modelo. De igual manera la calibración en promedio (CITL) fue muy cercana a 0 y la pendiente de la curva muy cercana a 1. En la valoración gráfica se observa que en los deciles bajos de probabilidad el modelo se ajusta muy bien a la línea de referencia, pero en los deciles mayores tiende a sobre estimar el riesgo.

Figura 15. Gráfico de calibración entre probabilidades esperadas por el modelo Bello-México y las observadas en la cohorte de validación según deciles de probabilidad esperada



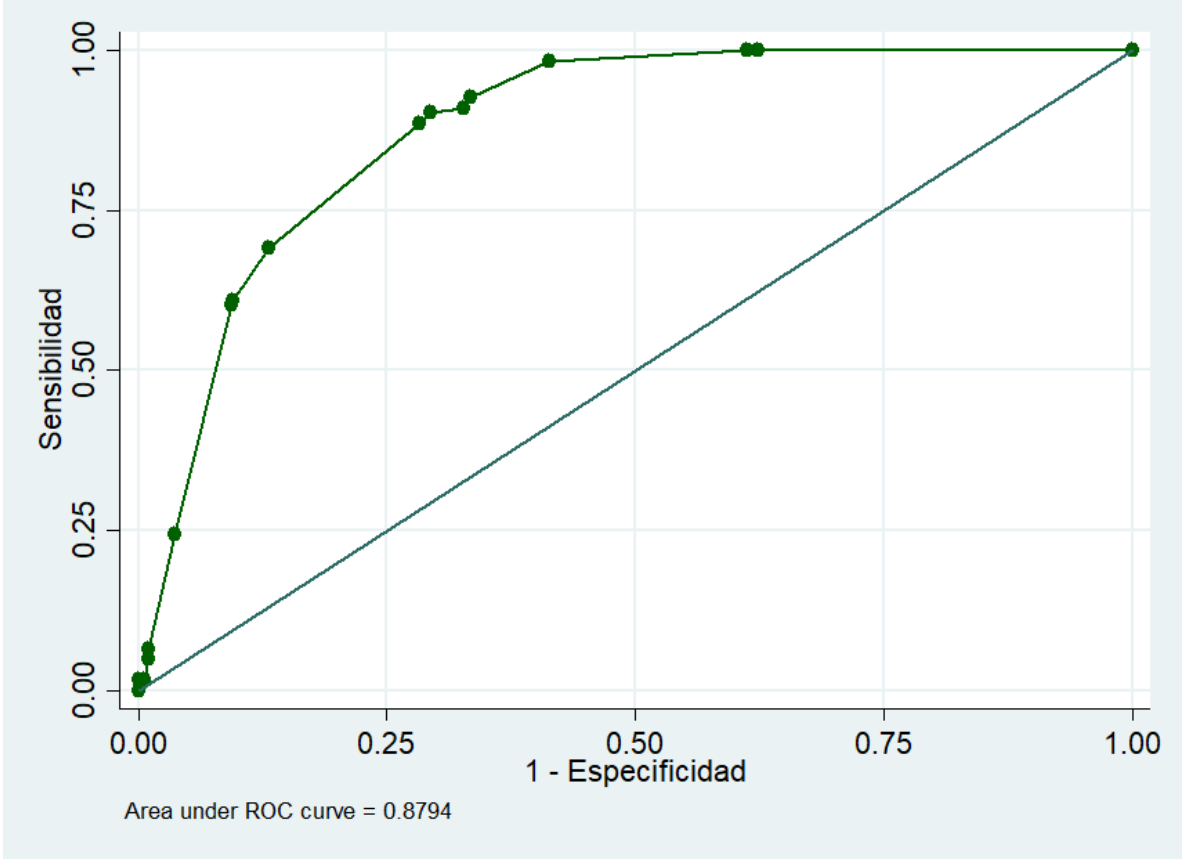
Prueba de Bondad de ajuste Hosmer-Lemeshow:  $p = 0.3254$

#### **6.3.4. Evaluación del desempeño de la escala Bello-México según los puntajes calculados a cada paciente de la cohorte RESCATE COVID-19:**

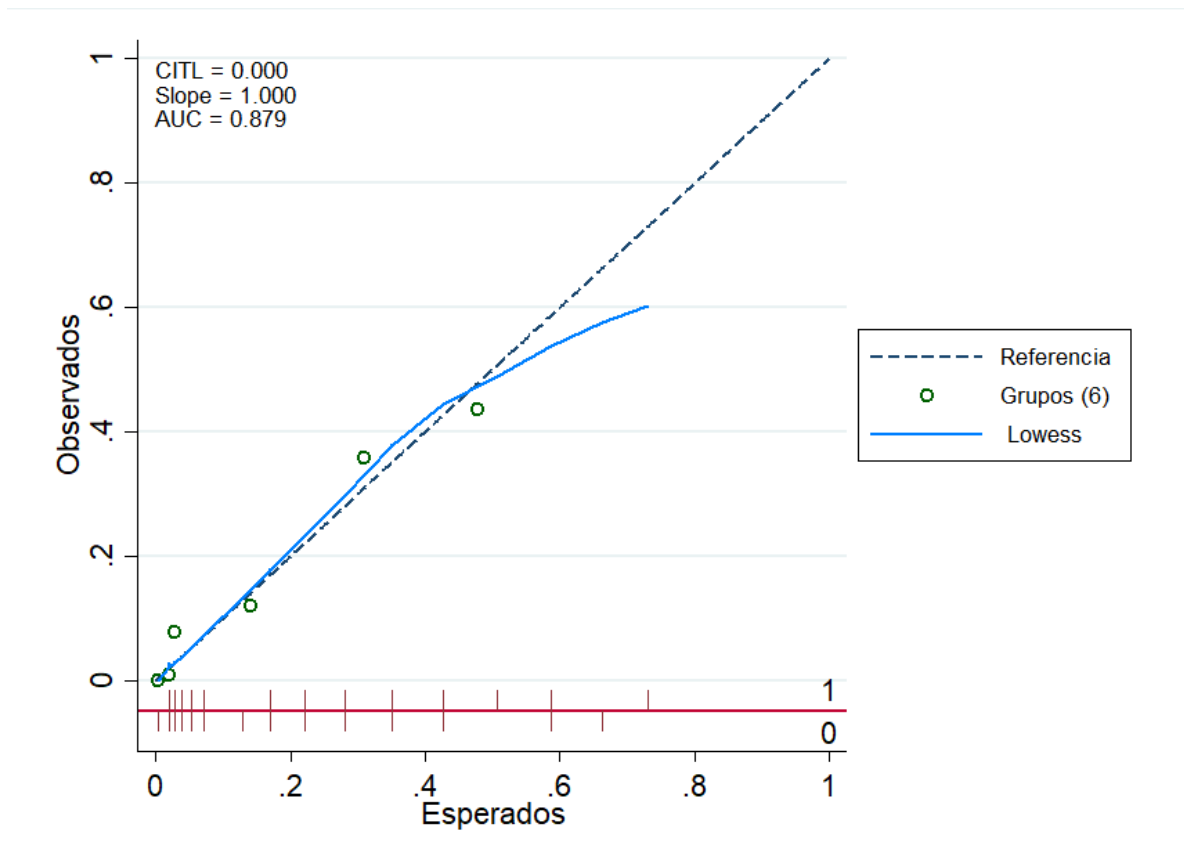
A cada uno de los 1171 pacientes con información disponible, se le realizó un cálculo del puntaje total de la escala según el sistema de puntuación disponible en la tabla 1. Posteriormente se generó una variable que contiene este valor para cada paciente y se realizó un modelo de regresión logística en la que se evalúa su asociación con el desenlace. Se obtuvo un OR de 1.38 (IC95%: 1.30 – 1.46) con  $p < 0.01$ . De esta manera por cada diferencia de un nivel de puntuación la probabilidad de mortalidad intrahospitalaria aumenta en un 38% (Ver anexo de modelos).

El área bajo la curva mejoró respecto a la evaluación del modelo original (AUC: 0.88) logrando capacidad de discriminación correcta muy buena (Gráfico 10). No fue posible agrupar las probabilidades en deciles (grupos de 10) por lo que no se logró realizar la prueba tradicional de Hosmer-Lemeshow. Se realizó la prueba con el máximo número de grupos disponibles (6) con lo cual se obtiene un resultado de  $p = 0.357$  que permite considerar un ajuste estadístico de los datos al modelo. El gráfico de calibración mantiene características muy similares a las analizadas en el gráfico del modelo.

Figura 16. Curva ROC del modelo Bello-México para predicción de mortalidad intrahospitalaria en la cohorte de validación externa según los puntajes de la escala:



**Figura 17. Gráfico de calibración entre probabilidades esperadas por los puntajes de la escala de Bello-México y la probabilidad observada en la cohorte de validación**



Prueba de bondad de ajuste de Hosmer-Lemeshow (en seis niveles)  $p=0.357$

### 6.3.5. Evaluación del desempeño de la escala Bello-México según los niveles de riesgo obtenidos a partir de los puntajes:

Se generó una variable categórica politómica con los niveles de riesgo propuestos por los autores según los puntajes de la escala. La variable tiene 5 niveles siendo la primera “muy bajo riesgo” ( $\leq 0$  puntos), seguida por “bajo riesgo” (1-3 puntos), “riesgo moderado” (4-6 puntos), “alto riesgo” (7-9 puntos) y “muy alto riesgo” ( $\geq 10$  puntos). Se realizó un modelo de regresión logística múltiple utilizando los diferentes niveles como variables indicadoras. De esta manera se obtuvieron OR como medidas de asociación entre la probabilidad de morir en cada nivel de riesgo comparado con el menor nivel (“muy bajo riesgo”).

**Tabla 7. Modelo de regresión logística múltiple (variables indicadoras) entre niveles de riesgo obtenidos a partir de los puntajes calculados por la escala Bello-México y mortalidad intrahospitalaria en la cohorte RESCATE-COVID-19.**

Predictor	Descripción		Regresión Logística ‡			
	Vivos al egreso		B	EE	OR (IC 95%)	p
	Si n= 1048 (89.5%)	No n= 123 (10.5%)				
<b>Muy bajo riesgo (≤ 0 puntos)</b>	614 (99.6)	2 (0.4)	-5.7268 †	0.7082	Ref.	N.A.
<b>Riesgo bajo (1-3 puntos)</b>	126 (92.6)	10 (7.4)	3.1932	0.7807	24.36 (5.3 – 112.5)	<0.01*
<b>Riesgo moderado (4-6 puntos)</b>	12 (85.7)	2 (14.3)	3.9351	1.0416	51.16 (6.6 – 394.1)	<0.01*
<b>Alto riesgo (7-9 puntos)</b>	198 (85.0)	35 (15.0)	3.9939	0.7316	54.26 (12.9 – 227.6)	<0.01*
<b>Muy alto riesgo (≥10 puntos)</b>	98 (57)	74 (43)	5.4459	0.7248	231.81 (55.9 – 959.6)	<0.01*
<b>Total</b>	<b>1171</b>		<b>Prueba de bondad de ajuste de Hosmer-Lemeshow: p = 1</b>			

B: Coeficiente (B1), EE: Error Estándar.

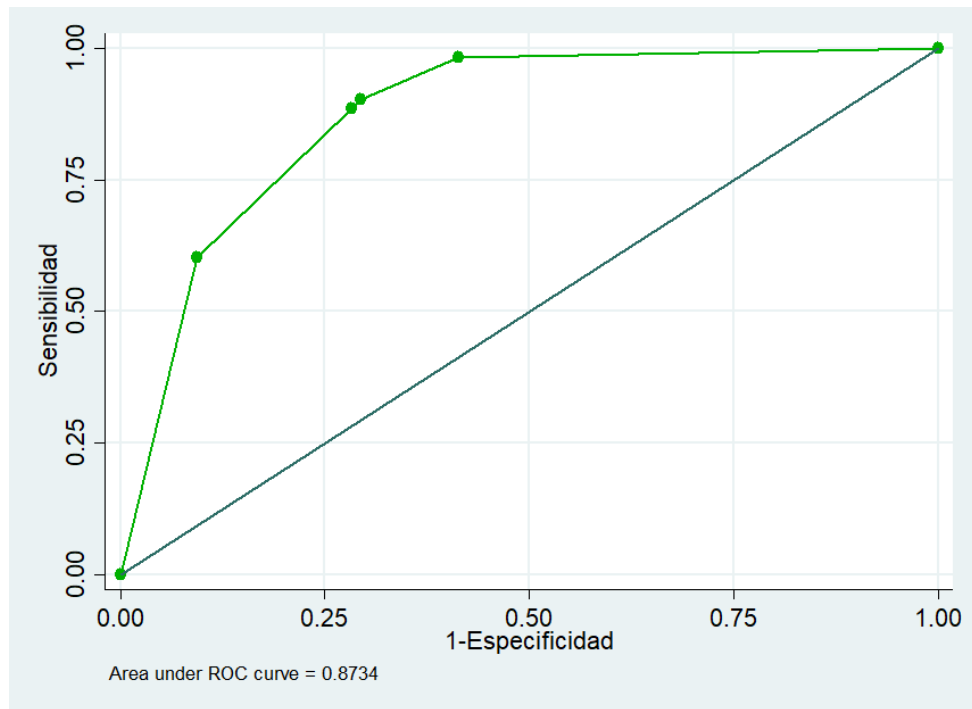
\* Asociación con significancia estadística  $p < 0.05$  tras ajustar efecto de covariables en el modelo.

† Probabilidad de muerte en el nivel de “muy bajo riesgo” es de **0.65% (Probabilidad para el intercepto)**.

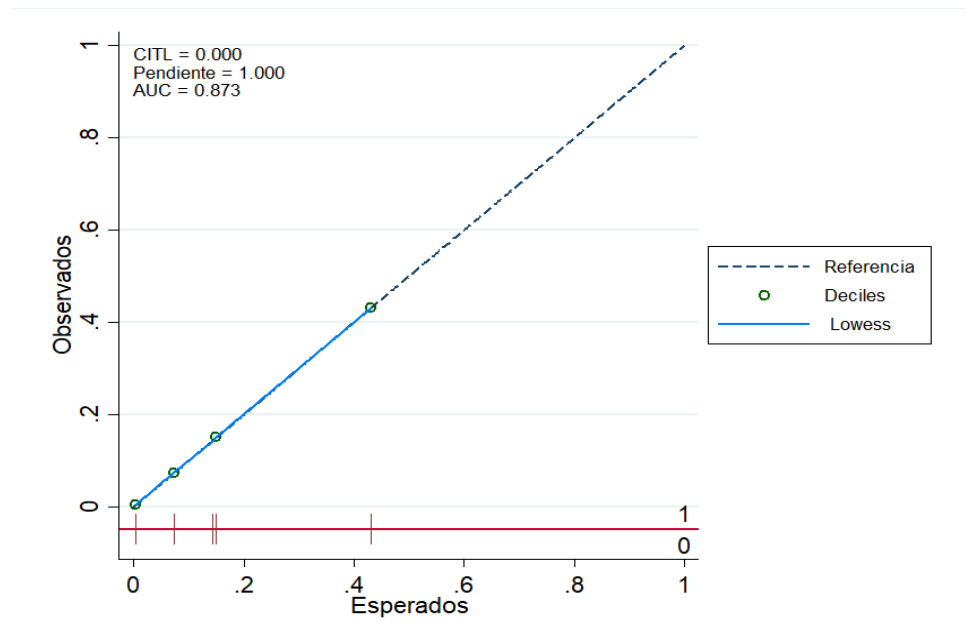
‡ Número de observaciones en el modelo: **1171**

En el modelo todos los niveles de riesgo tienen una asociación de proporcionalidad directa significativa con la probabilidad de muerte intrahospitalaria. Adicionalmente se aprecia un gradiente en la magnitud de los OR a través de los niveles de riesgo que le aporta información la capacidad de predicción de la escala a pesar de la incertidumbre progresivamente mayor en cada nivel. La capacidad de discriminación de la escala utilizando estos niveles se mantuvo muy buena como la obtenida con los puntajes (AUC: 0.87) y la calibración según estadísticos (prueba de HL con  $p=1$ , CITL=0 y Pendiente=1) y valoración gráfica también lo es.

**Figura 18. Curva ROC de la escala Bello-México para predicción de mortalidad intrahospitalaria en la cohorte de validación externa según los niveles de riesgo:**



**Figura 19. Gráfico de calibración entre probabilidades esperadas por los niveles de riesgo puntajes de la escala de Bello-México y la probabilidad observada en la cohorte de validación.**



Prueba de bondad de ajuste de Hosmer-Lemeshow:  $p = 1$

**Tabla 8. Comparativo de áreas bajo la curva ROC con sus respectivos intervalos de confianza del 95% correspondientes a cada uno de los modelos, sus escalas y los niveles de riesgo.**

	AUC ROC	IC95%	Observaciones
Zhang – China			
Modelo	0.789	0.709 – 0.869	317
Escala	0.743	0.655 – 0.831	
Niveles	0.678	0.585 – 0.771	
Xie - China			
Modelo	0.814	0.769 – 0.858	593
Bello - México			
Modelo	0.801	0.761 – 0.841	1168
Escala	0.879	0.855 – 0.903	
Niveles	0.873	0.848 – 0.898	



### **6.3.6. Conclusiones del análisis del desempeño de la escala Bello-México al aplicarla a la cohorte RESCATE COVID-19:**

Con esta información se considera que el modelo de Bello-México cumple con criterios de muy buen desempeño al evaluarse en la cohorte RESCATE COVID-19. El modelo multivariado resultó con 4/9 predictores que mantienen significancia tras ser ajustados por el efecto de todas las variables. Su capacidad de discriminación es buena ( $AUC > 0.8$ ) y cumple criterios estadísticos y gráficos de un buen ajuste entre las predicciones esperadas por el modelo y las observadas, especialmente en los deciles bajos de probabilidad. La magnitud de la calibración es consistente con la observada en su cohorte de derivación, e igualmente sus medidas de calibración, condiciones que aportan validez (interna y externa) a esta escala de pronóstico para pacientes con COVID-19.

Con la utilización de la escala y el cálculo de los puntajes para cada paciente el rendimiento de la escala se mantuvo. Su capacidad de discriminación siguió siendo buena utilizando el sistema de puntuación y los niveles de riesgo como predictores en los modelos de regresión. La calibración gráficamente se mantiene, aunque los estadísticos de HL no se lograron calcular con comparaciones de deciles de probabilidad por la reducción de los datos. El gradiente observado en la magnitud de la asociación a través de los diferentes niveles de riesgo aporta validez al constructo y al proceso de validación externa.

Además, los autores presentan un sistema de puntuación sencillo, fácilmente aplicable, para una escala que utiliza mayoritariamente como predictores a antecedentes clínicos y a información del abordaje clínico básico inicial. Su potencial de generalización se es alto y sus beneficios clínicos especialmente en la identificación de pacientes de bajo riesgo se consideran relevantes. Teniendo eso en cuenta se realizó un análisis de los mejores puntos de corte de la escala según su sensibilidad, especificidad y valores predictivos (Tablas 7 y 8).

Por tratarse de una escala sustentada en variables clínicas como antecedentes, valoración médica inicial y la toma de una radiografía de tórax, se considera útil en escenario de primer contacto médico. Es por esto que se selecciona un punto de corte de con énfasis en alta sensibilidad, baja tasa de falsos negativos y alto valor predictivo negativo para su uso clínico. De esta manera se cumple con el objetivo de tamización en el abordaje inicial que consiste principalmente en evitar pasar por alto a los pacientes con alto riesgo.

**Tabla 8. Evaluación de los puntos de corte de la escala Bello-México para su posible aplicación clínica en escenarios de primer contacto médico.**

Punto de corte (Puntaje final en escala)	Sensibilidad (%)	Especificidad (%)	Clasificación correcta (%)	LR+	LR-	VPP (%)	VPN (%)
≥ -6	100	0.00%	10.63%	1.0000		12.74	-
≥ -5	100	37.91%	44.51%	1.6107	0.0000	19.03	100
≥ 0	100	38.99%	45.47%	1.6390	0.0000	19.30	100
≥ 1	98.36%	58.58%	62.80%	2.3745	0.0280	25.74	99.59
≥ 2	92.62%	66.67%	69.43%	2.7787	0.1107	28.86	98.40
≥ 3	90.98%	67.35%	69.86%	2.7865	0.1339	28.91	98.08
≥ 4	90.98%	70.08%	72.30%	3.0407	0.1287	30.74	98.15
≥ 7	89.34%	71.15%	73.08%	3.0969	0.1498	31.13	97.85
≥ 8	69.67%	86.55%	84.76%	5.1800	0.3504	43.06	95.13
≥ 9	61.48%	90.25%	87.20%	6.3074	0.4268	47.93	94.13
≥ 10	60.66%	90.45%	87.28%	6.3503	0.4350	48.11	94.02
≥ 11	24.59%	96.20%	88.59%	6.4691	0.7839	48.58	89.73
≥ 12	6.56%	99.03%	89.20%	6.7278	0.9436	49.68	87.89
≥ 13	4.92%	99.03%	89.02%	5.0459	0.9602	42.54	87.70
≥ 14	1.64%	99.42%	89.02%	2.8033	0.9894	29.22	87.37
≥ 15	1.64%	100	89.55%		0.9836	-	99.85
> 15	0.00%	100	89.37%		1.0000	-	99.85

Nivel de riesgo según puntajes	Sensibilidad (%)	Especificidad (%)	Clasificación correcta (%)	LR+	LR-	VPP (%)	VPN (%)
Muy bajo riesgo (0 puntos)	100	0.00%	10.63%	1.0000	-	93.58	-
Riesgo bajo (1-3 puntos)	98.36%	58.58%	62.80%	2.3745	0.0280	25.74	99.59
Riesgo moderado (4-6 puntos)	90.98%	70.08%	72.30%	3.0407	0.1287	30.74	98.15
Alto riesgo (7-9 puntos)	89.34%	71.15%	73.08%	3.0969	0.1498	31.13	97.85
Muy alto riesgo (≥10 puntos)	60.66%	90.45%	87.28%	6.3503	0.4350	48.11	94.02

LR: Likelihood Ratio Positivo (+), Negativos (-)

VPP: Valor predictivo positivo; VPN: Valor predictivo positivo.

Se seleccionó el punto de corte de **4 puntos (nivel de riesgo moderado)** en la escala de Bello y colaboradores, como umbral para la toma de decisiones clínicas. De esta manera un paciente con puntajes menores tendrá un 98% de probabilidad de no presentar el evento. La probabilidad de presentarlo con puntajes superiores es del 30% y aumenta hasta el 48% en el nivel de muy alto riesgo. Así a decisión clínica de seguimiento ambulatorio, ampliación en los periodos de contacto, diferir toma de laboratorios o evitar remisiones a centros de mayor complejidad será mejor informada. Por otra parte, la incertidumbre sobre el riesgo de presentar el evento en pacientes con mayores puntajes se tendrá que discriminar mejor con otras herramientas de mayor especificidad.

#### **6.4. Proceso de evaluación de los modelos en la población de pacientes con RT-PCR para SARS-COV2 negativa de la cohorte rescate COVID-19.**

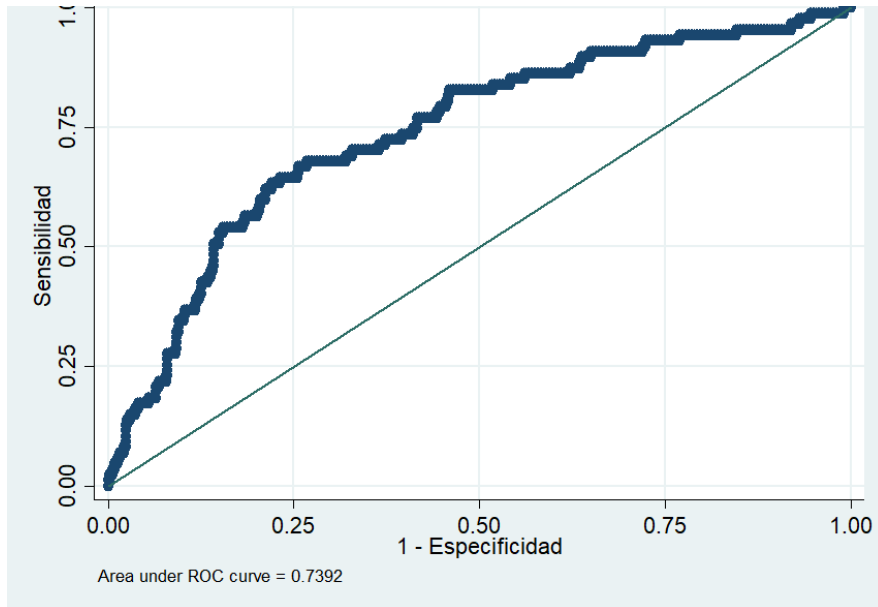
##### **6.4.1. Evaluación del desempeño de la escala Zhang-China en pacientes con RT-PCR negativa para SARS-COV2 de la cohorte RESCATE COVID-19**

El grupo de pacientes negativos en la cohorte RESCATE COVID-19 fue de 1242 al punto de corte del 31 de diciembre de 2020 y en éste se habían documentado 97 eventos (mortalidad: 7.8%). La caracterización de cada una de las variables en la población, así como el análisis bivariado y el modelo de regresión multivariado detallado, se encuentran disponibles en las tablas suplementarias del anexo 11. El modelo propuesto por Zhang y colaboradores no se logró evaluar en su totalidad en este grupo de pacientes ya que para ese momento no contaba con información sobre los valores de proteína C reactiva. Los recuentos celulares de interés (neutrófilos, linfocitos, y plaquetas) estaban disponibles en el 54% de los pacientes y la creatinina en el 45.3% de la población. El modelo de regresión logística multivariada se generó con todas las variables disponibles (no PCR) en 550 pacientes que contaban con la información disponible.

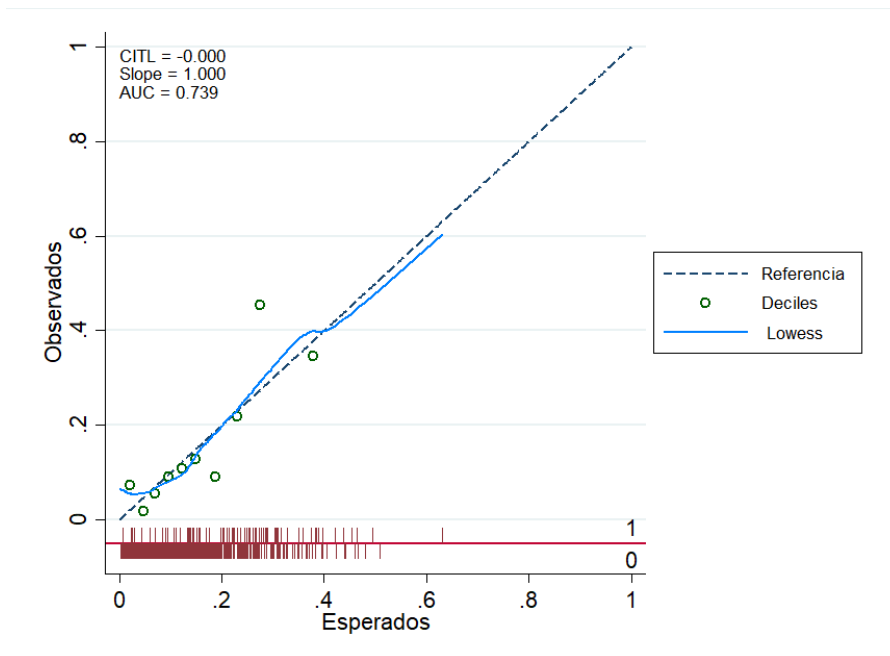
A pesar de que en los OR ajustados de 4 de los 6 predictores evaluados resultaron en una asociación significativa con la mortalidad intrahospitalaria (positiva: edad, creatinina y negativa: linfocitos y plaquetas) la capacidad de discriminación del modelo fue aceptable (AUC: 0.73). Su área bajo la curva ROC fue menor que al evaluarla en los pacientes confirmados.

De igual forma la calibración del modelo es deficiente. Estadísticamente no hay ajuste entre las probabilidades esperadas por el modelo y las observadas al comparar grupos de deciles (bondad de ajuste de Hosmer-Lemeshow:  $p < 0.01$ ).

**Figura 20. Curva ROC del modelo Zhang-China para predicción de mortalidad intrahospitalaria en la cohorte de validación externa de pacientes con RT-PCR para SARS-COV2 negativa de la cohorte RESCATE COVID-19.**



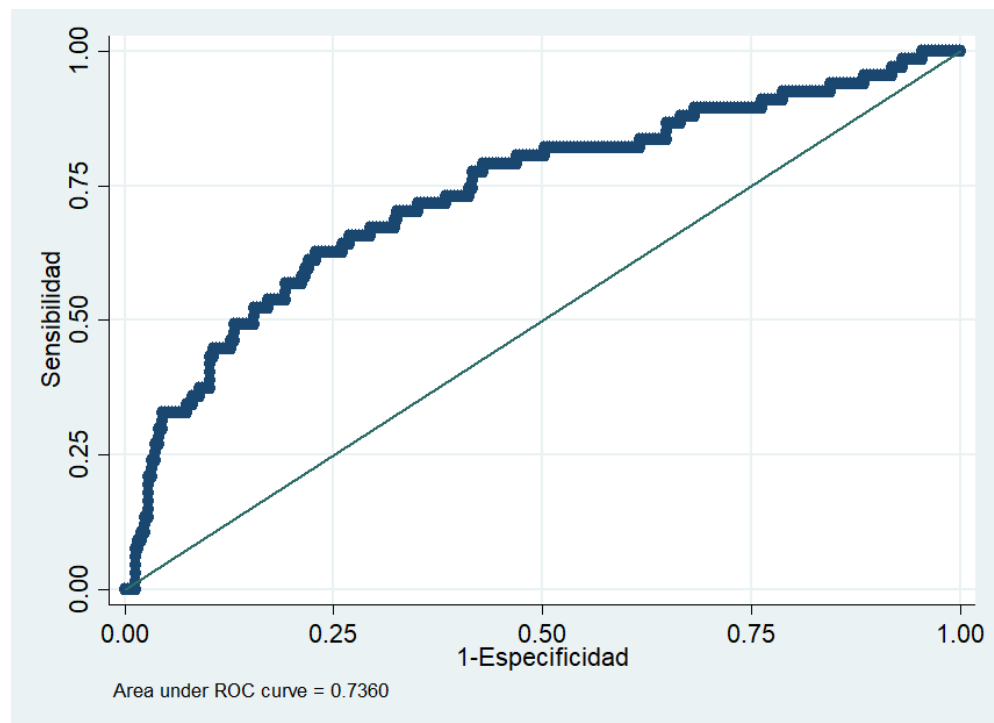
**Gráfico 15. Gráfico de calibración entre probabilidades esperadas por el modelo Zhang-China y las observadas en la cohorte de pacientes con RT-PCR negativa de la cohorte RESCATE COVID-19.**



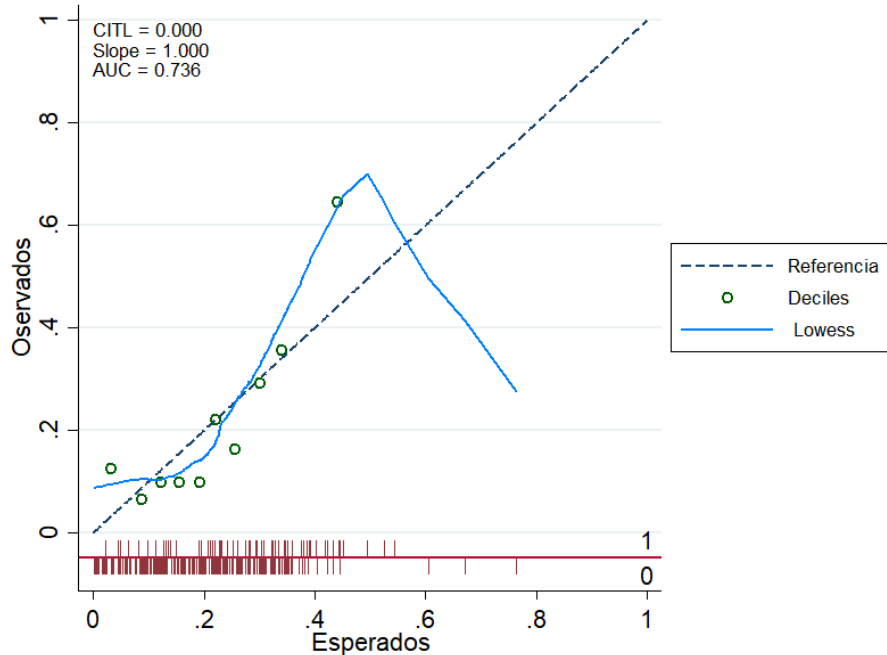
#### 6.4.2. Evaluación del desempeño de la escala Xie-China en pacientes con RT-PCR negativa para SARS-COV2 de la cohorte RESCATE COVID-19.

la descripción de la distribución de las variables dependientes del modelo Xie-China y su análisis de asociación (bivariado y multivariado) se encuentra en tablas detalladas en el anexo 11 La proporción de valores perdidos de la LDH fue del 73% y la de linfocitos del 46% por lo que el modelo de regresión se realizó en 312 pacientes. En este solamente uno de los predictores presentó una asociación significativa negativa con la probabilidad de muerte intrahospitalaria (OR ajustado de linfocitos: 0.46; IC95%0.29-0.71). Su capacidad de discriminación fue apenas aceptable (AUC: 0.73) con menor desempeño que al evaluarse en pacientes con diagnóstico confirmado de COVID-19. El modelo falló en su calibración de maneta estadística (HL  $p < 0.05$ ) y gráficamente.

**Figura 21. Curva ROC del modelo Xie-China para predicción de mortalidad intrahospitalaria en la cohorte de validación externa de pacientes con RT-PCR negativa de la cohorte RESCATE COVID-19.**



**Figura 22. Gráfico de calibración entre probabilidades esperadas por el modelo Xie-China y las observadas en la cohorte de pacientes con RT-PCR negativa de la cohorte RESCATE COVID-19.**



#### **6.4.3. Evaluación del desempeño de la escala Bello-México en pacientes con PCR negativa.**

La evaluación del modelo mexicano se logró realizar en 1110 pacientes ya que la proporción de valores perdidos para sus variables fue menor comparada con las propuestas con los modelos orientales. A pesar de esto en la regresión logística solamente uno de los predictores mostró una asociación significativa con el aumento de la probabilidad de muerte intrahospitalaria (neumonía: OR: 21). La capacidad de discriminación del modelo fue muy buena con un área bajo la curva ROC de 0.9. De igual forma se demostró ajuste estadístico de los datos al modelo (HL  $p > 0.5$ ) y en la evaluación gráfica los valores predichos y esperados se mantienen muy cercanos a la línea de referencia especialmente en los deciles bajos de probabilidad.

**Tabla 9. Asociación entre predictores del modelo Bello-México y mortalidad intrahospitalaria en el grupo de paciente con RT-PCR para SARS-COV2 negativa de la cohorte RESCATE COVID-19.**

Descripción				Análisis Bivariado (Regresión Logística Simple)			Análisis Multivariado † (Regresión Logística Múltiple)			
Predictor	General n= 1258	Mortalidad		Riesgo en el comparador (%)	OR crudo (IC 95%)	p	B	EE	OR ajustado (IC 95%)	p
		No n = 1145	Si n = 97							
Edad < 40	625 (49.6)	611 (53.3)	10 (10.3)	33	0.09 (0.05 – 0.19)	<0.01*	-0.51156	0.47696	0.59 (0.23 – 1.52)	0.2835
Edad > 65	338 (26.8)	262 (22.8)	69 (71.1)	6.3	8.23 (5.19 – 13.05)	<0.01*	0.79588	0.37752	2.21 (1.05 – 4.64)	0.0350
Diabetes	132 (10.5)	109 (9.5)	23 (23.7)	14.3	2.95 (1.77 – 4.90)	<0.01*	1.11580	1.39586	3.05 (0.19 – 47.07)	0.4241
Diabetes <40	7 (0.5)	6 (0.5)	1 (1.0)	14.3	2.31 (0.27 – 19.5)	0.439	N.A.	N.A.	N.A.	N.A.
Obesidad	40 (3.1)	34 (2.9)	6 (6.1)	16.4	2.15 (0.88 – 5.26)	0.092	0.97268	0.73284	2.64 (0.62 – 11.12)	0.1844
Neumonía	282 (22.4)	197 (17.2)	84 (86.6)	2.7	31.09 (16.99 – 56.88)	<0.01*	3.05928	0.38407	21.32 (10.03 – 45.24)	<0.01*
ERC	67 (5.3)	48 (4.1)	19 (19.6)	14.2	5.56 (3.12 – 9.93)	<0.01*	0.87715	0.50223	2.40 (0.89 – 6.43)	0.0807
EPOC	132 (10.5)	105 (9.2)	25 (25.7)	13.8	3.43 (2.09 – 5.65)	<0.01*	-0.06351	0.34605	0.93 (0.47 – 1.85)	0.8544
Inmunosup	21 (1.7)	18 (1.6)	3 (0.3)	16.7	1.99 (0.57 – 6.91)	0.274x	0.26286	0.77207	1.30 (0.28 – 5.90)	0.7335

Para todas las variables se reporta su **conteo y proporción en porcentaje (%)**

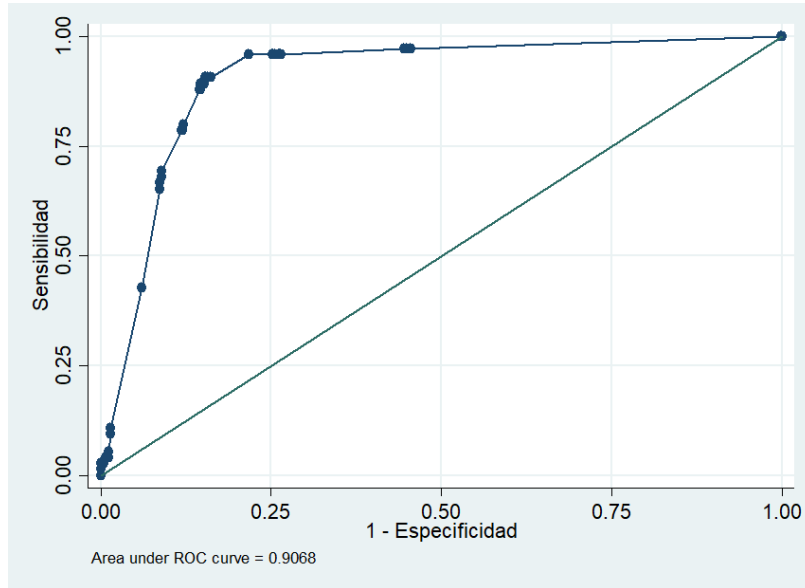
B: Coeficiente (B1), EE: Error Estándar, ERC: Enfermedad Renal Crónica, EPOC: Enfermedad Pulmonar Obstructiva crónica, Inmunosup: inmunosupresión.

\* Asociación con significancia estadística p<0.05

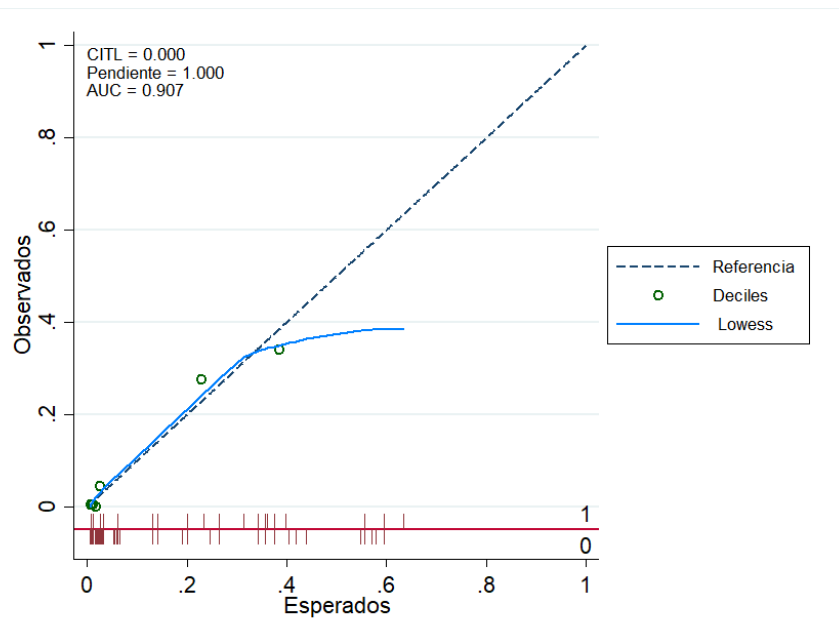
† Número de observaciones en el modelo multivariado: **1110**



**Figura 23. Curva ROC del modelo Bello-México para predicción de mortalidad intrahospitalaria en la cohorte de validación externa de pacientes con RT-PCR negativa de la cohorte RESCATE COVID-19.**



**Figura 24. Gráfico de calibración entre probabilidades esperadas por el modelo Bello-México y las observadas en la cohorte de pacientes con RT-PCR negativa de la cohorte RESCATE COVID-19.**



**Prueba de bondad de ajuste de Hosmer Lemeshow:  $p=0.2154$**

## 7. DISCUSIÓN

Mediante este estudio se logró documentar al menos una escala que podría utilizarse en la práctica clínica local como herramienta para la toma de decisiones durante la atención a la pandemia por COVID-19. El proceso para la selección de las escalas, su caracterización y análisis se realizó mediante procesos estructurados y rigurosos que se explican de manera detallada y de los cuales se deja evidencia en este documento. Los análisis se realizaron en una cohorte robusta de pacientes que se ha ensamblado gracias a la participación de seis instituciones hospitalarias de tres ciudades distintas de Colombia. Además, su curso ha sido liderado por un centro con reconocida experiencia y trayectoria en el campo de la investigación nacional. No se conoce de otro estudio a nivel nacional de validación de escalas de pronóstico en COVID-19.

El modelo final que proponen Bello y colaboradores incluye únicamente predictores clínicos, en su mayoría antecedentes de comorbilidades que se han comportado como factores de riesgo de manera consistente en diferentes estudios (42,71–73). La única variable derivada de la atención médica es el diagnóstico de neumonía que tiene implícita la necesidad de una imagen diagnóstica de tórax para completar la puntuación en la escala. Su cohorte de derivación ( $n > 50.000$ ) y su validación cruzada en el 20% de la muestra tienen un tamaño que da suficiente poder para evaluar contrastes y medidas de asociación.

La evaluación de la calidad de la escala realizada por Wynants y colaboradores resaltó la importancia de realizar procesos de validación externa para evaluar posibles estimaciones optimistas. Por esto resulta importante reconocer la similitud en los valores de las medidas de discriminación al analizar el modelo (AUC: 0.82 en la derivación y 0.81 en RESCATE) como un criterio de validez externa del modelo. Así mismo resulta llamativo que en nuestra cohorte la capacidad de discriminación del sistema de puntos sea superior al del estudio mexicano (AUC: 0.82 y 0.87) ya que generalmente los procesos de validación externa arrojan valores más bajos (66).

En la cohorte RESCATE COVID-19 se contaba con la mayoría de la información que se requería para la evaluación de esta escala, habiendo sido recolectada también de manera retrospectiva. La incidencia de mortalidad en el estudio original (10.1%) es similar a la observada en nuestra población (12.7%) y a la reportada en pacientes con diagnóstico confirmado que consulta a un servicio de urgencias (20). El proceso de validación se llevó a cabo en diferentes momentos de la pandemia,

en países diferentes y con una muestra heterogénea de instituciones (6 centros en diferentes ciudades) que probablemente recogen gran parte de la variabilidad de la práctica. Todos estos argumentos sobre las fortalezas del estudio en cuanto a riesgos de selección.

Múltiples autores han reconocido que la falta de evaluación de la calibración es una de las principales razones de la baja calidad analítica de los modelos publicados en COVID-19 (56,74). Dentro de las fortalezas de este estudio se reconoce el esfuerzo en analizar la calibración como una medida del desempeño del modelo más allá del uso de pruebas estadísticas como la de Hosmer-Lemeshow. En todos los casos se realizó un análisis estadístico y gráfico de la calibración para calificar la dirección de la sobre o subestimación (si estaba presente) y los cambios en las estimaciones a lo largo del espectro de las probabilidades graficadas (evaluación visual y de la pendiente). Con esto se pretende evaluar al modelo con mayor profundidad como recomiendan diferentes grupos expertos en la validación de modelos (62,69,75).

El buen desempeño del modelo mexicano en el grupo de pacientes negativos es uno de los hallazgos más llamativos de este estudio. En primera medida se debe reconocer que esta evaluación fue premeditada y manifiesta explícitamente como el último de los objetivos del proyecto. Se planeó de esta manera sustentados en la hipótesis de que podría haber similitudes entre los grupos de pacientes positivos y negativos teniendo en cuenta las dificultades técnicas para la toma y procesamiento de muestras a nivel nacional que probablemente resultaron errores de clasificación. Además, la cohorte se conforma de pacientes sintomáticos, definidos como casos sospechosos por criterios clínicos en quienes un primer resultado de PCR no descarta el diagnóstico.

Teniendo en cuenta esta probabilidad de clasificación errónea y la naturaleza de las variables del modelo de Bello (ninguna derivada de estudios invasivos o costosos), se considera que el hallazgo del buen desempeño de la escala tanto en pacientes negativos como positivos (sospechosos) lo convierte en una herramienta valiosa. Especialmente para estadios tempranos en el proceso de la atención en los que no se dispone aún del resultado de una prueba molecular.

Otras escalas que utilizan predictores clínicos sencillos y que han sido diseñadas para determinar pronóstico de patologías específicas como CURB-65 (en neumonía) o CHADS-VASc (en fibrilación auricular) han demostrado buen desempeño al evaluarse en otros escenarios. La primera por ejemplo ha

demostrado utilidad en predicción de mortalidad en pacientes con sepsis y la segunda en pacientes con obstrucción coronaria y fenómeno de no reflujo (76,77).

La utilización de un sistema de puntuación como una herramienta para la traducción clínica del modelo es otra a de las características a destacar de la escala mexicana. Además de su buen desempeño esto le aporta capacidad de generalización, contrario a lo ocurrido con el modelo de Xie cuya herramienta para la aplicación clínica no permitió su validación completa.

También es importante reconocer que el estudio cuenta con algunas limitaciones, especialmente con las derivadas del diseño retrospectivo en la cohorte de validación. La posibilidad de sesgos de selección y de información en la selección de los pacientes y en la caracterización de las variables está presente. Para el caso de las escalas que utilizan medidas de laboratorio, especialmente las que no se utilizan naturalmente en el abordaje clínico inicial, la presencia del valor en la base de datos puede introducir un sesgo de selección. Para el caso de la escala mexicana esto ocurre con la variable “neumonía” ya que es probable que a quienes se les haya solicitado una radiografía sean sistemáticamente diferentes de los pacientes a quienes no se les realizó.

Los sesgos de información se han manejado con la revisión periódica de la base de datos para corregir valores extremos o atípicos. Para efectos de este estudio se realizó un reporte de todos los valores perdidos de cada uno de los predictores de los modelos. En cada uno de los modelos se reportó el número de observaciones con las que se analizó que corresponden a los pacientes con valores disponibles para todas las variables. Sin embargo, no es posible asegurar que los valores perdidos en los predictores no sean sistemáticamente diferentes en quienes presentaron o no el evento. En el caso del modelo mexicano el riesgo de sesgos de información es bajo por la naturaleza de sus variables.

El efecto de la confusión siempre fue evaluado utilizando modelos de regresión lineal múltiple para obtener medidas de asociación ajustadas.

Finalmente se considera que la utilización del modelo mexicano en la práctica clínica puede informar al médico de primer contacto a tomar decisiones más seguras. Teniendo en cuenta que el mayor beneficio del punto de corte de seleccionado de 4 en la escala (para definir riesgo bajo Vs alto) es el de no dejar de intervenir a pacientes con alto riesgo de mortalidad. De esta manera es posible una mejor utilización de los servicios de salud en tiempos de pandemia.

## 8. Conclusiones

La escala de Bello-México resultó tener muy buen desempeño al aplicarse a los pacientes con diagnóstico de COVID-19 de la cohorte RESCATE-COVID19. Su sistema de puntuación sencillo, con buena capacidad de discriminación entre pacientes de alto y bajo riesgo, ofrece predicciones calibradas de las probabilidades de mortalidad intrahospitalaria. Teniendo en cuenta la naturaleza de sus predictores se considera útil en escenarios de atención temprana de pacientes con diagnóstico o sospecha de la enfermedad.

Con un punto de corte de 4 puntos la escala resulta útil como estrategia de tamización del pronóstico ya que puntajes menores tienen muy baja tasa de falsos negativos. Su uso puede informar mejor (aunque no reemplazar) la decisión clínica y optimizar la utilización de recursos. Es necesario evaluar su desempeño en otras poblaciones antes de generalizar su uso.

La escala de Zhang-China no logró un buen desempeño al aplicarse a la cohorte de RESCATE COVID-19 donde obtuvo mala capacidad de discriminación y mala calibración. El modelo de Xie-China, aunque prometedor por su desempeño, presenta una herramienta para cálculo de puntajes compleja, vulnerable de error de medición y demandante de recursos, que no la hace apta para su aplicación clínica local.

## 9. CLASIFICACIÓN DE LOS RESULTADOS DEL ESTUDIO SEGÚN COLCIENCIAS:

Como resultado del estudio logró una validación local de una escala de predicción del pronóstico de los pacientes consultan a servicios de salud por sospecha de infección por SARS-COV2 en una cohorte de pacientes colombianos. Teniendo en cuenta que estas se consideran herramientas tecnológicas para la toma de decisiones de la práctica clínica se considera que hacen parte de los estudios **“relacionados con desarrollos tecnológicos e innovación”** según la clasificación de Colciencias. Debido a que el estudio se desarrolla como un proyecto dentro del proceso de formación de un estudiante de maestría se considera también relacionado con el **“fortalecimiento de la capacidad científica institucional”** de la Universidad Autónoma de Bucaramanga (UNAB).

Como indicadores se espera generar la publicación de un artículo científico en el que se exponga el diseño, metodología y resultados del estudio con la intención de difundir la información y contribuir al conocimiento en el área. Así mismo se espera sea aprobado tras su sustentación como tesis de grado de maestría en métodos para la producción y aplicación del conocimiento científico en salud de la UNAB. Los potenciales beneficiarios de la actual propuesta de investigación son los pacientes infectados por SARS-CoV2, el personal de salud dedicado a su atención y los gestores de los recursos en salud locales.

## REFERENCIAS BIBLIOGRÁFICAS

1. INS IN de S. Coronavirus (COVID-19) en Colombia. 2020 [Internet]. El coronavirus (COVID-19) en Colombia 2020. 2020 [citado 30 de octubre de 2020]. Disponible en: <https://www.ins.gov.co/Noticias/paginas/coronavirus.aspx>
2. Johns Hopkins Univertisy. COVID-19 Dashboard Global Cases by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU) [Internet]. Johns Hopkins Coronavirus Resource Center. 2020 [citado 30 de octubre de 2020]. Disponible en: <https://coronavirus.jhu.edu/map.html>
3. OPS y OMS. Reporte de Situacion COVID-19 Colombia No. 168 - 26 de octubre 2020 - OPS/OMS | Organización Panamericana de la Salud [Internet]. 2020 [citado 30 de octubre de 2020]. Disponible en: <https://www.paho.org/es/documentos/reporte-situacion-covid-19-colombia-no-168-26-octubre-2020>
4. Dinero. Colombia duplicó el número de camas UCI, desde el inicio de la pandemia [Internet]. Camas UCI se duplicaron en Colombia desde el inicio de la pandemia. [citado 30 de octubre de 2020]. Disponible en: <http://www.dinero.com/empresas/confidencias-on-line/articulo/camas-uci-se-duplicaron-en-colombia-desde-el-inicio-de-la-pandemia/302316>
5. Reporte de Situacion COVID-19 Colombia No. 212 - 16 de marzo 2021 - OPS/OMS | Organización Panamericana de la Salud [Internet]. [citado 3 de abril de 2021]. Disponible en: <https://www.paho.org/es/documentos/reporte-situacion-covid-19-colombia-no-212-16-marzo-2021>
6. RECOVERY Collaborative Group, Horby P, Lim WS, Emberson JR, Mafham M, Bell JL, et al. Dexamethasone in Hospitalized Patients with Covid-19 - Preliminary Report. *N Engl J Med*. 17 de julio de 2020;
7. Deffereos SG, Giannopoulos G, Vrachatis DA, Siasos GD, Giotaki SG, Gargalianos P, et al. Effect of Colchicine vs Standard Care on Cardiac and Inflammatory Biomarkers and Clinical Outcomes in Patients Hospitalized With Coronavirus Disease 2019. *JAMA Netw Open* [Internet]. 24 de junio de 2020 [citado 30 de octubre de 2020];3(6). Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7315286/>
8. Paranjpe I, Fuster V, Lala A, Russak AJ, Glicksberg BS, Levin MA, et al. Association of Treatment Dose Anticoagulation With In-Hospital Survival Among Hospitalized Patients With COVID-19. *J Am Coll Cardiol*. 7 de julio de 2020;76(1):122-4.
9. Diaz-Quijano F, Silva J, Ganem F, Oliveira S, Vesga-Varela A, Croda J. A model to predict SARS-CoV-2 infection based on the first three- month surveillance data in Brazil. 2020.
10. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ*. 7 de abril de 2020;m1328.
11. Antioquia U de. Factores clínicos pronósticos de enfermedad grave y mortalidad en pacientes con COVID-19. Síntesis rápida. Factores Clínicos

Pronósticos Enferm Grave Mortal En Pacientes Con COVID-19 Sínt Rápida. 2020;57-57.

12. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. *Lancet Lond Engl*. 15 de 2020;395(10223):470-3.
13. Bogoch II, Watts A, Thomas-Bachli A, Huber C, Kraemer MUG, Khan K. Pneumonia of unknown aetiology in Wuhan, China: potential for international spread via commercial air travel. *J Travel Med*. 13 de 2020;27(2).
14. WHO. WHO. Virtual press conference on COVID-19 – 11 March 2020 [Internet]. WHO. World Health Organization; 2020 [citado 30 de octubre de 2020]. Disponible en: <http://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/>
15. Cheng VCC, Lau SKP, Woo PCY, Yuen KY. Severe Acute Respiratory Syndrome Coronavirus as an Agent of Emerging and Reemerging Infection. *Clin Microbiol Rev*. octubre de 2007;20(4):660-94.
16. Ren L-L, Wang Y-M, Wu Z-Q, Xiang Z-C, Guo L, Xu T, et al. Identification of a novel coronavirus causing severe pneumonia in human: a descriptive study. *Chin Med J (Engl)*. 5 de mayo de 2020;133(9):1015-24.
17. Wang W, Tang J, Wei F. Updated understanding of the outbreak of 2019 novel coronavirus (2019-nCoV) in Wuhan, China. *J Med Virol*. abril de 2020;92(4):441-7.
18. Holshue ML, DeBolt C, Lindquist S, Lofy KH, Wiesman J, Bruce H, et al. First Case of 2019 Novel Coronavirus in the United States. *N Engl J Med*. 5 de marzo de 2020;382(10):929-36.
19. MinSalud. Colombia confirma su primer caso de COVID-19 [Internet]. 2020 [citado 30 de octubre de 2020]. Disponible en: <https://www.minsalud.gov.co/Paginas/Colombia-confirma-su-primer-caso-de-COVID-19.aspx>
20. Reporte de Situación COVID-19 Colombia No. 213 - 24 de marzo 2021 - OPS/OMS | Organización Panamericana de la Salud [Internet]. [citado 2 de abril de 2021]. Disponible en: <https://www.paho.org/es/documentos/reporte-situacion-covid-19-colombia-no-213-24-marzo-2021>
21. INS IN de S. Coronavirus (COVID-19) en Colombia. 2021 [Internet]. El coronavirus (COVID-19) en Colombia 2021. 2021 [citado 2 de abril de 2021]. Disponible en: <https://www.ins.gov.co/Noticias/Paginas/Coronaviruss.aspx>
22. WHO. Laboratory testing for 2019 novel coronavirus (2019-nCoV) in suspected human cases [Internet]. 2020 [citado 30 de octubre de 2020]. Disponible en: <https://www.who.int/publications-detail-redirect/10665-331501>
23. Zhang Z-L, Hou Y-L, Li D-T, Li F-Z. Diagnostic efficacy of anti-SARS-CoV-2 IgG/IgM test for COVID-19: A meta-analysis. *J Med Virol* [Internet]. [citado 30 de octubre de 2020];n/a(n/a). Disponible en: <https://onlinelibrary.wiley.com/doi/abs/10.1002/jmv.26211>
24. Deeks JJ, Dinnes J, Takwoingi Y, Davenport C, Spijker R, Taylor-Phillips S, et al. Antibody tests for identification of current and past infection with SARS-CoV-2. *Cochrane Database Syst Rev* [Internet]. 2020 [citado 30 de octubre de 2020];(6). Disponible en:



<https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD013652/full?cookiesEnabled>

25. Liu R, Liu X, Han H, Shereen MA, Niu Z, Li D, et al. The comparative superiority of IgM-IgG antibody test to real-time reverse transcriptase PCR detection for SARS-CoV-2 infection diagnosis. medRxiv. 30 de marzo de 2020;2020.03.28.20045765.
26. Gestoso-Pecellín L, García-Flores Y, González-Quintana P, Marrero-Arencibia JL. Recomendaciones y uso de los diferentes tipos de test para detección de infección por SARS-COV-2. Enfermería Clin. febrero de 2021;31:S40-8.
27. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med. 30 de abril de 2020;382(18):1708-20.
28. Xu X-W, Wu X-X, Jiang X-G, Xu K-J, Ying L-J, Ma C-L, et al. Clinical findings in a group of patients infected with the 2019 novel coronavirus (SARS-Cov-2) outside of Wuhan, China: retrospective case series. BMJ [Internet]. 19 de febrero de 2020 [citado 30 de octubre de 2020];368. Disponible en: <https://www.bmj.com/content/368/bmj.m606>
29. Chung M, Bernheim A, Mei X, Zhang N, Huang M, Zeng X, et al. CT Imaging Features of 2019 Novel Coronavirus (2019-nCoV). Radiology. 2020;295(1):202-7.
30. Li Y, Xia L. Coronavirus Disease 2019 (COVID-19): Role of Chest CT in Diagnosis and Management. AJR Am J Roentgenol. 2020;214(6):1280-6.
31. Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, et al. Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. Radiology. 2020;296(2):E32-40.
32. Yuan M, Yin W, Tao Z, Tan W, Hu Y. Association of radiologic findings with mortality of patients infected with 2019 novel coronavirus in Wuhan, China. PLOS ONE. 19 de marzo de 2020;15(3):e0230548.
33. Confalonieri M, Salton F, Fabiano F. Acute respiratory distress syndrome. Eur Respir Rev Off J Eur Respir Soc. 30 de junio de 2017;26(144).
34. Matthay MA, Aldrich JM, Gotts JE. Treatment for severe acute respiratory distress syndrome from COVID-19. Lancet Respir Med. 2020;8(5):433-4.
35. Mejía F, Medina C, Cornejo E, Morello E, Vásquez S, Alave J, et al. Características clínicas y factores asociados a mortalidad en pacientes adultos hospitalizados por COVID-19 en un hospital público de Lima, Perú. 29 de junio de 2020 [citado 28 de marzo de 2021]; Disponible en: <https://preprints.scielo.org/index.php/scielo/preprint/view/858>
36. Chen C, Chen C, Yan JT, Zhou N, Zhao JP, Wang DW. [Analysis of myocardial injury in patients with COVID-19 and association between concomitant cardiovascular diseases and severity of COVID-19]. Zhonghua Xin Xue Guan Bing Za Zhi. 24 de julio de 2020;48(7):567-71.
37. Yang X, Yu Y, Xu J, Shu H, Xia J, Liu H, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. Lancet Respir Med. 2020;8(5):475-81.

38. Li H, Liu L, Zhang D, Xu J, Dai H, Tang N, et al. SARS-CoV-2 and viral sepsis: observations and hypotheses. *The Lancet*. 9 de mayo de 2020;395(10235):1517-20.
39. Cheng Y, Luo R, Wang K, Zhang M, Wang Z, Dong L, et al. Kidney impairment is associated with in-hospital death of COVID-19 patients. *medRxiv*. 20 de febrero de 2020;2020.02.18.20023242.
40. Bello-Chavolla OY, Bahena-López JP, Antonio-Villa NE, Vargas-Vázquez A, González-Díaz A, Márquez-Salinas A, 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*. 1 de agosto de 2020;105(8).
41. Yang J, Zheng Y, Gou X, Pu K, Chen Z, Guo Q, et al. Prevalence of comorbidities and its effects in patients infected with SARS-CoV-2: a systematic review and meta-analysis. *Int J Infect Dis IJID Off Publ Int Soc Infect Dis*. mayo de 2020;94:91-5.
42. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet*. 28 de marzo de 2020;395(10229):1054-62.
43. 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*. 07 de 2020;323(13):1239-42.
44. Lippi G, Plebani M, Henry BM. Thrombocytopenia is associated with severe coronavirus disease 2019 (COVID-19) infections: A meta-analysis. *Clin Chim Acta Int J Clin Chem*. julio de 2020;506:145-8.
45. Lippi G, Plebani M. Procalcitonin in patients with severe coronavirus disease 2019 (COVID-19): A meta-analysis. *Clin Chim Acta Int J Clin Chem*. junio de 2020;505:190-1.
46. Tang N, Li D, Wang X, Sun Z. Abnormal coagulation parameters are associated with poor prognosis in patients with novel coronavirus pneumonia. *J Thromb Haemost JTH*. 2020;18(4):844-7.
47. Vences MA, Ramos JJP, Otero P, Veramendi-Espinoza LE, Vega-Villafana M, Mogollón-Lavi J, et al. FACTORES ASOCIADOS A MORTALIDAD EN PACIENTES HOSPITALIZADOS CON COVID-19: COHORTE PROSPECTIVA EN EL HOSPITAL NACIONAL EDGARDO REBAGLIATI MARTINS. LIMA, PERÚ. 24 de septiembre de 2020 [citado 28 de marzo de 2021]; Disponible en: <https://preprints.scielo.org/index.php/scielo/preprint/view/1241>
48. de Mayo, Salazar MG, Espeche WG, Barochiner J, Leiva C, Ennis IL. Pandemia por COVID19: ¿Después qué? Control de los factores de riesgo para prevenir una segunda ola de morbilidad y mortalidad. *Sociedad Argentina de Hipertensión Arterial*. :10.
49. Belanger MJ, Hill MA, Angelidi AM, Dalamaga M, Sowers JR, Mantzoros CS. Covid-19 and Disparities in Nutrition and Obesity. *N Engl J Med*. 10 de septiembre de 2020;383(11):e69.
50. Petrova D, Salamanca-Fernández E, Rodríguez Barranco M, Navarro Pérez P, Jiménez Moleón JJ, Sánchez M-J. La obesidad como factor de riesgo en

personas con COVID-19: posibles mecanismos e implicaciones. *Aten Primaria*. 2020;52(7):496-500.

51. Manrique-Hernández EF, Moreno-Montoya J, Hurtado-Ortiz A, Prieto-Alvarado FE, Idrovo ÁJ. Desempeño del sistema de vigilancia colombiano durante la pandemia de COVID-19: evaluación rápida de los primeros 50 días. *Biomédica*. 12 de noviembre de 2020;40(Suppl 2):96-103.

52. Reporte de Situación COVID-19 Colombia No. 163- 20 de octubre 2020 - OPS/OMS | Organización Panamericana de la Salud [Internet]. [citado 3 de abril de 2021]. Disponible en: <https://www.paho.org/es/documentos/reporte-situacion-covid-19-colombia-no-163-20-octubre-2020>

53. Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *BMJ*. 23 de febrero de 2009;338:b375.

54. Emanuel EJ, Persad G, Upshur R, Thome B, Parker M, Glickman A, et al. Fair Allocation of Scarce Medical Resources in the Time of Covid-19. *N Engl J Med*. 21 de mayo de 2020;382(21):2049-55.

55. COVID Precise [Internet]. [citado 3 de abril de 2021]. Disponible en: <https://www.covprecise.org/project/>

56. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. *BMJ*. 7 de abril de 2020;369:m1328.

57. Zhang H, Shi T, Wu X, Zhang X, Wang K, Bean D, et al. Risk prediction for poor outcome and death in hospital in-patients with COVID-19: derivation in Wuhan, China and external validation in London, UK. *medRxiv*. 3 de mayo de 2020;2020.04.28.20082222.

58. Xie J, Hungerford D, Chen H, Abrams ST, Li S, Wang G, et al. Development and external validation of a prognostic multivariable model on admission for hospitalized patients with COVID-19. *medRxiv*. 7 de abril de 2020;2020.03.28.20045997.

59. Bello-Chavolla OY, Bahena-López JP, Antonio-Villa NE, Vargas-Vázquez A, González-Díaz A, Márquez-Salinas A, 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*. 01 de 2020;105(8).

60. Herramienta para el cálculo en línea de riesgo de mortalidad según el modelo Zhang-China [Internet]. [citado 29 de marzo de 2021]. Disponible en: <https://covid.datahelps.life/about/>

61. F. Wolff R, G.M. Moons K, D. Riley R, F. Whiting P, Westwood M, S. Collins G, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann Intern Med* [Internet]. 1 de enero de 2019 [citado 28 de octubre de 2020]; Disponible en: <https://www.acpjournals.org/doi/abs/10.7326/M18-1376>

62. Moons KGM, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart Br Card Soc*. mayo de 2012;98(9):691-8.

63. Bleeker SE, Moll HA, Steyerberg EW, Donders ART, Derksen-Lubsen G, Grobbee DE, et al. External validation is necessary in prediction research: a clinical example. *J Clin Epidemiol*. septiembre de 2003;56(9):826-32.
64. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiol Camb Mass*. enero de 2010;21(1):128-38.
65. Magliano DJ, Sacre JW, Harding JL, Gregg EW, Zimmet PZ, Shaw JE. Young-onset type 2 diabetes mellitus - implications for morbidity and mortality. *Nat Rev Endocrinol*. junio de 2020;16(6):321-31.
66. Alba AC, Agoritsas T, Walsh M, Hanna S, Iorio A, Devereaux PJ, et al. Discrimination and Calibration of Clinical Prediction Models: Users' Guides to the Medical Literature. *JAMA*. 10 de octubre de 2017;318(14):1377-84.
67. Hosmer DW, Hosmer T, Cessie SL, Lemeshow S. A Comparison of Goodness-of-Fit Tests for the Logistic Regression Model. *Stat Med*. 1997;16(9):965-80.
68. Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW. A calibration hierarchy for risk models was defined: from utopia to empirical data. *J Clin Epidemiol*. 2016;74:167-76.
69. On behalf of Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative, Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. *BMC Med*. diciembre de 2019;17(1):230.
70. Documentos Técnicos covid-19 [Internet]. [citado 26 de octubre de 2020]. Disponible en:  
<https://www.minsalud.gov.co/salud/publica/PET/Paginas/Documentos-tecnicos-covid-19.aspx>
71. Alqahtani JS, Oyelade T, Aldhahir AM, Alghamdi SM, Almehmadi M, Alqahtani AS, et al. Prevalence, Severity and Mortality associated with COPD and Smoking in patients with COVID-19: A Rapid Systematic Review and Meta-Analysis. *PLoS ONE* [Internet]. 11 de mayo de 2020 [citado 4 de abril de 2021];15(5). Disponible en:  
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7213702/>
72. Pei G, Zhang Z, Peng J, Liu L, Zhang C, Yu C, et al. Renal Involvement and Early Prognosis in Patients with COVID-19 Pneumonia. *J Am Soc Nephrol JASN*. junio de 2020;31(6):1157-65.
73. Pal R, Bhadada SK. COVID-19 and diabetes mellitus: An unholy interaction of two pandemics. *Diabetes Metab Syndr*. 2020;14(4):513-7.
74. Leeuwenberg AM, Schuit E. Prediction models for COVID-19 clinical decision making. *Lancet Digit Health*. octubre de 2020;2(10):e496-7.
75. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J*. 1 de agosto de 2014;35(29):1925-31.
76. Yamamoto S, Yamazaki S, Shimizu T, Takeshima T, Fukuma S, Yamamoto Y, et al. Prognostic utility of serum CRP levels in combination with CURB-65 in patients with clinically suspected sepsis: a decision curve analysis. *BMJ Open*

[Internet]. 28 de abril de 2015 [citado 4 de abril de 2021];5(4). Disponible en:  
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4420934/>

77. Barman HA, Kahyaoglu S, Durmaz E, Atici A, Gulsen K, Tugrul S, et al. The CHADS-VASc score is a predictor of no-reflow in patients with non-ST-segment elevation myocardial infarction. *Coron Artery Dis.* enero de 2020;31(1):7-12.

## ANEXOS

**Anexo 1:** Carta de aval de la Clínica Chicamocha para proyecto RESCATE-COVID-19.



Bucaramanga, agosto 18 de 2020

Doctor  
CUSTODIO ALBERTO RUIZ BEDOYA  
Médico Especialista en Medicina Interna  
Clínica Chicamocha S.A.  
Ciudad

Estimado Doctor Custodio Ruiz:

En primer lugar, queremos agradecer su interés en realizar procesos de investigación científica en nuestra institución, los cuales esperamos ayuden a mejorar la atención que damos a nuestros pacientes.

En relación con su solicitud de aval para realizar el proyecto de investigación "RESCATE COVID-19: Respuesta de Conocimiento para la Atenuación de la Epidemia por COVID-19" en la Clínica Chicamocha S.A, como parte de la red de centros asistenciales que aportarán información al proyecto en mención, me permito comunicarle que, tanto el Comité de Investigación como el comité de Ética en Investigación de la Clínica Chicamocha S.A. en sus sesiones ordinarias llevadas a cabo los días 30 de julio y 11 de agosto 2020, respectivamente, analizaron el proyecto presentado e hicieron las siguientes consideraciones y decisiones:

1. El líder, coordinador y responsable del proyecto en la Clínica Chicamocha S.A. será el Dr. Custodio Alberto Ruiz Bedoya
2. Todos los participantes del equipo de investigación de la Clínica Chicamocha S.A. se comprometen a allegar al Comité de Ética en Investigación su certificado de aprobación de un curso de investigación con seres humanos o de buenas prácticas clínicas.
3. El Dr. Custodio Ruiz queda comprometido a firmar la correspondiente acta de inicio del proyecto y a cumplir con lo establecido en la misma.
4. El Dr. Custodio Ruiz queda comprometido a informar al Comité de Ética en Investigación de la institución, de cualquier modificación o desviación del protocolo de estudio que fue aprobado por el mismo.
5. El Dr. Custodio Ruiz velará porque se salvaguarde la confidencialidad de la información a la cual tendrá acceso tanto él como los demás miembros del equipo de investigación de la Clínica Chicamocha S.A. y de otras instituciones participantes en el estudio RESCATE.

6. La Clínica Chicamocha S.A. se guarda el derecho de usar la información que se origine en la institución, para los fines que considere pertinente y siguiendo los procedimientos establecidos por ella.
7. Todos los participantes del equipo de investigación de la Clínica Chicamocha S.A. se comprometen a crear su CvLAC en la plataforma de Minciencias y a vincularlo al Grupo de Investigación de la Clínica Chicamocha S.A.
8. El Dr. Custodio Ruiz tendrá como filiación institucional a la Clínica Chicamocha S.A. en todo momento y para efectos de la participación suya y de la institución en la generación de productos académicos.
9. La Clínica Chicamocha S.A. no asumirá costos adicionales a los implícitos en la atención usual de los pacientes, incluido aquí el costo del mantenimiento del sistema de información y de historias clínicas de la institución.
10. El Dr. Custodio Ruiz y los demás miembros del equipo de investigación exoneran a la Clínica Chicamocha S.A. de cualquier pago por el trabajo que deban realizar con ocasión de su participación en el proyecto en mención. Para esto, cada uno deberá enviar una carta en la que conste que su participación en el proyecto la harán en horario no laboral y en forma ad honorem.

De acuerdo con el cumplimiento de las anteriores consideraciones y el análisis del proyecto, el concepto del Comité de Investigación y el comité de Ética en Investigación es: **APROBADO**.

Cordial saludo,



Dra. LINA MARÍA VERA CALA  
Presidente  
Comité de Investigación  
Comité de Ética en Investigación  
Clínica Chicamocha S.A.

**Anexo 2.** Proceso de selección de los modelos de predicción documentados por la revisión sistemática realizada por Wynants

#	Ref	Authors	Dataset Setting	Study design	Age mean $\pm$ se, median (IQR), or range	Sex (% male)	Outcome	Sample size total  No of participants for model development set (No with outcome)	Predictors in final model	Modelling technique	Type of validation	Sample size: total  No of participants for model validation (No with outcome)	Performance* (C index, sensitivity (%), specificity (%), PPV/ NPV (%), calibration slope, other (95% CI, if reported))	Selection Process (Razon for in/out)
1	9	Bai, Fang, et al.	China Wuhan Pulmonary Hospital, January 3 to February 13.  Inpatients at admission with mild confirmed covid-19	Retrospective cohort	53 $\pm$ 13	50%	Deterioration into severe/critical disease (period unspecified)	133 (54)	Combination of demographics, signs and symptoms, laboratory results and features derived from CT images	Neural net (Multilayer perceptron + long short term memory vs. logistic regression, linear discriminant analysis, support vector machine, multilayer perceptron)	Unclear	N.A.	C index 0.95 (0.94 to 0.97)	Out – No mortality outcome
2	19	Caramelo, Ferreira, et al.	China. Simulated based on data from Wuhan. December 8 to	Simulation	Unclear	41%	Mortality (period unspecified) *2	Unknown	Age, sex, presence of any comorbidity (hypertension, diabetes, cardiovascular disease, chronic respiratory	Logistic regression	N.A.	Not reported	Not reported	Out – Simulation



			February 11						disease, cancer) *3					
3		Gong, Ou, et al:	China, Guangzhou Eighth People's Hospital, Zhongnan Hospital of Wuhan University, Third Affiliated Hospital of Sun Yat-sen University, January 20 to March 2	Retrospective cohort	Dev.: 49 Val.1: 52 Val.2: 42	Dev.: 47% Val.1: 44% Val.2: 50%	Severe COVID-19 infection (minimum 15 day)		Age, serum LDH, CRP, variation of red blood cell distribution width, blood urea nitrogen, albumin, direct bilirubin	LASSO followed by logistic regression (vs. LASSO followed by decision tree, random forest or support vector machine)				<b>Out – No mortality outcome</b>
4	20	Lu, Hu, et al.	China Wuhan Hankou Hospital, January 21 to February 4.  Inpatients at admission with suspected or confirmed covid-19	Retrospective cohort  One center	55 (IQR 39-66)	44%	Mortality (12 day)	577 (44)	Age, C-reactive protein	Cox regression	Not reported	Not reported	Not reported	In? – ACP Scale, Simple  <b>Out - Predictors are not available in RESCATE</b>
5	21	Qi, Jiang, et al.	China, 5 hospitals from Ankang, Lishui, Zhenjiang, Lanzhou, Linxia between January 23 to February 8; date of	Prospective cohort	38 (IQR 26 to 47)	55%	Hospital stay >10 days	26 (20)	6 features derived from CT images *3	Logistic regression	Fivefold cross validation	N.A.	C index 0.92	<b>Out – No mortality outcome</b>

			last follow-up February 20.  Inpatients with confirmed covid-19 at admission											
6	21	Qi, Jiang, et al.	China, 5 hospitals from Ankang, Lishui, Zhenjiang, Lanzhou, Linxia between January 23 to February 8; date of last follow-up February 20.  inpatients with confirmed covid-19 at admission.  Inpatients with confirmed covid-19 at admission	Prospective cohort	38 (IQR 26 to 47)	55%	Hospital stay >10 days	26 (20)	6 features derived from CT images *3	Tree-based (random forest)	Fivefold cross validation	N.A	C index 0.96	<b>Out – No mortality outcome</b>
7	37	Shi, Yu, et al.	China, Hangzhou, Zhejiang Province, study dates unspecified, follow-up until	Retrospective cohort	46 ± 19	53%	Death or severe COVID-19 (period unspecified)	478 (49)	Age (dichotomized), sex, hypertension	Multivariate model (not specified)	Validation in less severe cases	66 (15)	Not reported	<b>Out - Predictors are not available in RESCATE</b>

			February 17.  inpatients with confirmed covid-19 at admission											
8	7	Xie, Hungerford, et al	China, Tongji Hospital and Jinyintan Hospital, admitted between January and February  inpatients with confirmed covid-19 at admission	Retrospective cohort	Tongji: 65 (IQR 54 to 73) Jinyintan: 56 (IQR 47 to 68)	Tongji: 52% Jinyintan: 35%	Mortality (in hospital)	299 (155)	Age, LDH, lymphocyte count, SPO2	Logistic regression	External validation (other Chinese centre)	130 (69)	C index 0.98 (0.96 to 1.00),  calibration slope 2.5 (1.7 to 3.7)	In ***
9	22	Yan, Zhang, et al.	China, Tongji Hospital Wuhan, January 10 to February 18  inpatients suspected of covid-19	Retrospective cohort	59 ± 16	59%	Mortality (period unspecified)	375 (174)	Lactic dehydrogenase, lymphocyte count, high sensitivity C-reactive protein	Tree-based (XGBoost)	Temporal validation, selecting only severe cases	29 (17)	Sensitivity 92, PPV 95	Out – Machine Learning
10	23	Yuan, Yin, et al.	China, Hubei Public Health Clinical Centre; central Hospital of Wuhan,	Retrospective cohort	60 (IQR 47 to 69)	44%	Mortality (period unspecified)	Not applicable	Clinical scorings of CT images (zone, left/right, location, attenuation, distribution of affected parenchyma)	NA (external validation)	External validation of existing model	27 (10)	C index 0.90 (0.87 to 0.93)	Out - Predictors are not available in RESCATE

			January 1 to January 25.  inpatients with confirmed covid-19 at admission											
1 1	60	Huang, Cai et al.	China, Guangzhou 8th People's Hospital, January 20 to February 29.  inpatients with confirmed covid-19 at admission	Retrospective cohort	45 ± 19	50%	severe symptoms 3 days after admission	125 (32)	Underlying diseases, fast respiratory rate >24/min, elevated CRP-level (> 10mg/dL), elevated lactate dehydrogenase level (> 250U/L)	Logistic regression	Apparent performance only	Not applicable	C index 0.99 (0.97 to 1.00),  sensitivity 91, specificity 96	<b>Out – No mortality outcome</b>
1 2	61	Pourhomayoun, Shakibi et al.	76 countries. Publicly available data on Github from Cov2019 Data Working Group with confirmed cases worldwide in a variety of settings.  inpatients with confirmed covid-19	Unclear	57 (mean)	Unclear	in-hospital mortality (period unspecified)	Unknown	Unknown	Neural net (Neural Networks vs. Support Vector Machine (SVM), Random Forest, Decision Tree, Logistic Regression, and KNearest Neighbour (KNN))	10-fold cross validation	N.A.	C index 0.96, sensitivity 90, High specificity 97	<b>Out – Machine Learning process for modeling</b>

1 3	44	Sakar, Chakrabarti	Several Countries. Kaggle data with cases from 22 countries in Asia, Australia, Europe, North America, January 13 to February 28.  inpatients with covid-19 symptoms	Unclear	Unclear	Unclear	death vs recovery (period unspecified)	80 (37)	Age, days from symptom onset to hospitalisation, from Wuhan, sex, visit to Wuhan	Tree-based (Random forest)	Apparent performance only	N.A.	C index 0.97	<b>Out – Machine Learning process for modeling</b>
1 4	42	Wang, Zha et al.	China. Renmin Hospital of Wuhan University, Henan Provincial People's Hospital, Beijing Youan Hospital of Capital Medical University, Huangshi Central Hospital).  inpatients with confirmed covid-19	Retrospective cohort	Dev. 51 ± 19; Val. 50 ± 19, 48 ± 14	Dev. 48%; Val. 47%, 51%	length of hospital stay	301 (not applicable)	Age and CT features	Cox regression	Not reported	N.A.	Not reported	<b>Out – No mortality outcome</b>
1 5	62	Zeng, Li et al.	Third People's hospital of Shenzhen (Shenzhen)	Retrospective cohort	Severe: 58.7 (11), nonsevere : 46.1 (14.1)	Severe: 62% non-severe: 44%	severe disease progression (period	338 (76)	CT features	LASSO followed by Fine and Gray	Cross validation (number	N.A.	C index 0.88	<b>Out – No mortality outcome</b>

			, China) from January 11 to February 29.  inpatients with confirmed covid-19				unspecifie d)			(for both analysis)	of folds unclear)			
1 6	62	Zeng, Li et al.	Third People's hospital of Shenzhen (Shenzhen , China) from January 11 to February 29.  inpatients with confirmed covid-19	Retrospec tive cohort	Severe: 58.7 (11), nonsevere : 46.1 (14.1)	Severe: 62% non- severe: 44%	severe disease progressio n (period unspecifie d)	338 (76)	CT features and laboratory markers	LASSO followed by Fine and Gray (for both analysis)	Cross validation (number of folds unclear)	N.A.	C index 0.88	<b>Out – No mortality outcome</b>
1 7	63	Al - Najjar, Al-Rousan	South Korea. Official time series data from the Korea Centers for Disease Control and Prevention (KCDC), Unclear start date of inclusion to 9 March	Registry data	Unclear	Unclear	recovery from covid19 (period unspecifie d)	466 (40)	birth year (age), sex, country, group, infection reason, confirmed date	Neural net (one hidden layer and gradient descent as an optimizati on algorithm )	Training test split	193 (14)	Sensitivity 43, specificity 98	<b>Out – No mortality outcome</b>

			target population unclear											
18	63	Al - Najjar, Al-Rousan	South Korea. Official time series data from the Korea Centers for Disease Control and Prevention (KCDC), Unclear start date of inclusion to 9 March  target population unclear	Registry data	Unclear	Unclear	mortality (period unspecified)	463 (25)	Age, sex, country, region, infection reason, confirmed date	Neural net (one hidden layer and gradient descent as an optimization algorithm)	Training test split	191 (7)	Sensitivity 86, specificity 100	<b>Out – Machine Learning process for modeling</b>
19	67	Barda, Riesel et al	Israel. Data warehouse of Clalit Health Services (payer-provider), 29 January to 8 April inclusion, follow-up until 22 April (covid cases for external validation)  patients with confirmed covid-19	Retrospective cohort	41 ± 21.26	49%	mortality (period unspecified)	735,000 (8251)	Age, sex, pack years, COPD, number of wheezing/dyspnea diagnoses, albumin, red cell distribution width, c-reactive peptide, urea, lymphocyte, chloride, creatinine, high density lipoprotein, duration of hospitalizations, count of hospitalizations, count of ambulance rides, count of sulfonamide dispenses, count of anticholinergic dispenses, count	Tree-based (gradient boosted tree)	Other (specify in column CL) January 29 to April 8 inclusion, follow-up until April 22 2020 (covid cases for external validation)	3176 (87)	C index 0.94 (0.92 to 0.96), sensitivity 90 (83 to 96), PPV 17 (14 to 21)	<b>Out – Machine Learning process for modeling</b>  Used indirect population information . No Clarity in final predictors of the model. Difficulty reconstructing modeling.

									of glucocorticoid dispenses, chronic respiratory disease, cardiovascular disease, diabetes, malignancy, hypertension					
20	70	Bello-Chavolla, Bahena Lopez et al	Mexico. General Directorate of Epidemiology of Mexican Ministry of Health, dates not reported.  Confirmed COVI-19 patients presenting at GP	Surveillance data	47 ± 16	58	30-day mortality	12424 (1137)	Age, pregnancy, diabetes, obesity, pneumonia, CKD, COPD, immunosuppression	Cox proportional hazards regression	Training test split	3105 (297)	C index 0.80, Somer's D 0.60	In – Sociodemographic risk factors. Power.
21	75	Carr, Bendayan et al	United Kingdom. King's College Hospital NHS Foundation Trust hospital, 1 March 1 to 5 April.  Inpatients with confirmed covid-19	Retrospective cohort	67 ± 28	55	progression to severe covid-19 (period unspecified)	452 (159)	Age, National Early Warning Score (NEWS) 2, CRP, neutrophil, eGFR, albumin	Logistic regression	Temporal validation	256 (59)	C index 0.73, sensitivity 46, specificity 87	Out – No mortality outcome
22	77	Chassagnon, Vakalopoulos, et al.	France. Unclear, 4 March to 29 March	Retrospective cohort	63 ± 16	67	composite, 4-day intubation or mortality	383 (84)	Unclear	Support Vector Machines, Decision Trees,	External validation (new)	95 (26)	Sensitivity 88, specificity 74	Out – Machine Learning process for modeling



			inpatients with confirmed covid-19							Random Forests, AdaBoost, and Gaussian Naive Bayes. Prediction made by winner takes all principle (Lasso/Ridge)	centres, France)			
23	79	Colombi, Bodini et al	Italy. "Guglielmo da Saliceto" Hospital, Piacenza, Italy, 17 February to 10 March.  inpatients with confirmed covid-19	Retrospective cohort	Unclear	75	ICU admission or inhospital (period unspecified)	236 (108)	Age, cardiovascular comorbidities, median platelet count, CRP, visual assessment of well aerated lung %	Logistic regression	Apparent performance only	N.A.	C index 0.86 (0.81 to 0.90), sensitivity 72 (63 to 80), specificity 81 (73 to 88) PPV 70 (61 to 78), NPV 78 (72 to 83)	Out - Predictors are not available in RESCATE
24	79	Colombi, Bodini et al	Italy. "Guglielmo da Saliceto" Hospital, Piacenza, Italy, 17 February to 10 March  inpatients with confirmed covid-19	Retrospective cohort	Unclear	75	ICU admission or inhospital mortality (period unspecified)	236 (108)	Age, cardiovascular comorbidities, median platelet count, LDH, CRP, software assessment of well aerated lung absolute volume, adipose tissue	Logistic regression	Apparent performance only	N.A.	C index 0.86 (0.81 to 0.90), sensitivity 75 (66 to 83), specificity 81 (73 to 88), PPV 70 (61 to 78), NPV 78 (72 to 83)	Out - Predictors are not available in RESCATE
25	80	Das, Mishra, et al	South Korea. Data shared by Korea	Retrospective cohort	Unclear	44	ICU admission or inhospital	3022 (61)	Age, sex, province, date of diagnosis, place	Gradient boosting algorithm	Training test split	604 (12)	C index 0.97	Out - Machine Learning

			Centers for Disease Control and Prevention, 20 January to 7 April 7				mortality (period unspecified)		of exposure to covid-19					process for modeling
			inpatients with confirmed covid-19											
26	119	Gong, Ou et al.	China. Target population unclear.	Retrospective cohort	¿	?	15-day progression to severe covid-19	189 (28)	Age, direct bilirubin, red cell distribution width, blood urea nitrogen, CRP, lactate dehydrogenase, albumin	Logistic regression	External validation (new centres, China)	165 (40)	C index 0.85 (0.79 to 0.92), sensitivity 78, specificity 78	Out – No mortality outcome
27	83	Guo, Liu et al.	China. 35 hospitals in Guangdong Province and Hubei Province, 27 December 2019, to 4 March 2020.	Retrospective cohort	42 (IQR 31 to 57)	48	14-day progression to severe covid-19	818 (24)	Age, chronic illness, neutrophil to lymphocyte ratio, CRP, D-dimer	Cox proportional hazards regression	External validation (new centres, China)	320 (38)	C index 0.78 (0.70 to 0.87)	Out – No mortality outcome
			inpatients with confirmed covid-19											
28	84	Hu, Liu et al	China. Sino-French New City Branch of	Dev: Retrospective cohort, Val: Unclear	68 ± 10 in nonsurvivors; 61 ± 13 in survivors	58	in-hospital mortality (period unspecified)	183 (68)	Age, high-sensitivity CRP, lymphocyte count, Ddimer	Logistic regression	External validation (new	64 (31)	C index 0.88, sensitivity 84,	Out – Machine Learning process for modeling

			Tongji Hospital (n=183; Dev) and Optical Valley Branch of Tongji Hospital (n=64; Val), 28 January 2020 to 11 March 2020 (dev)		in nonsurvivors						centres, China)		specificity 79	
29	85	Hu, Yao et al	China. Electronic patient care reports from an emergency medical team that collected in Wuhan from 7 February 7 to 7 March	Retrospective cohort	61 ± 16	51	in-hospital mortality (period unspecified)	N.A.	Modified Early Warning Score (MEWS): heart rate, systolic blood pressure, respiratory rate, body temperature, consciousness	?	External validation only	105 (19)	C index 0.68 (0.58 to 0.77), sensitivity 68, specificity 65, PPV 30, NPV 90	In? - Evaluated MEWS – REMS as clinical scores we don't have in RESCATE. But maybe we can calculate them.  Out - Predictors are not available in RESCATE
30	85	Hu, Yao et al	China. Electronic patient care reports from an emergency medical team that	Retrospective cohort	61 ± 16	51	in-hospital mortality (period unspecified)	N.A.	Rapid Emergency Medicine Score (REMS): mean arterial pressure, pulse rate, respiratory rate, oxygen	?	External validation only	105 (19)	C index 0.84 (0.76 to 0.91), sensitivity 89, specificity	In? - Evaluated MEWS – REMS as clinical scores we don't have in RESCATE.

			collected in Wuhan from 7 February 7 to 7 March						saturation, GCS, age				70, PPV 40, NPV 97	But maybe we can calculate them.
			inpatients with confirmed covid-19											Out - Predictors are not available in RESCATE
31	88	Ji, Zhang et al	China. Fuyang Second People's Hospital (FYSPH), Anhui; the fifth medical center of Chinese PLA general hospital (PLAGH), Beijing, Unclear; data collected from 20 January through 22 February.	Retrospective cohort	44 ± 16	56	10-day progression to severe COVID-19	208 (40)	Comorbidity, age, lymphocyte count, lactate dehydrogenase	Cox proportional hazards regression	Internal validation by resampling (bootstrap)	N.A.	C index 0.91 (0.86 to 0.94), sensitivity 95 (83 to 99), specificity 78 (71 to 84)	Out – No mortality outcome
32	89	Jiang, Coffee et al	China. Wenzhou Central Hospital and Cangnan People's Hospital in Wenzhou,	Prospective cohort	43 (IQR 32, 49)	62	acute respiratory distress syndrome	53 (5)	Alanine aminotransferase, myalgias, hemoglobin, gender, temp, Na+, K+, lymphocyte count, creatinine,	Logistic regression	10-fold cross validation	N.A.	Classification accuracy 50%	Out – Machine Learning process for modeling

			China, dates not reported  inpatients with confirmed covid-19						age, white blood count					
33	89	Jiang, Coffee et al	China. Wenzhou Central Hospital and Cangnan People's Hospital in Wenzhou, China, dates not reported  inpatients with confirmed covid-19	Prospective cohort	43 (IQR 32, 49)	62	acute respiratory distress syndrome	53 (5)	Alanine aminotransferase , myalgias, hemoglobin, gender, temp, Na+, K+, lymphocyte count, creatinine, age, white blood count	K nearest neighbour	10-fold cross validation	N.A.	Classification accuracy 80%	<b>Out – Machine Learning process for modeling</b>
34	89	Jiang, Coffee et al	China. Wenzhou Central Hospital and Cangnan People's Hospital in Wenzhou, China, dates not reported  inpatients with confirmed covid-19	Prospective cohort	43 (IQR 32, 49)	62	acute respiratory distress syndrome	53 (5)	Alanine aminotransferase , myalgias, hemoglobin, gender, temp, Na+, K+, lymphocyte count, creatinine, age, white blood count	Tree-based (decision tree (gain ratio))	10-fold cross validation	N.A.	Classification accuracy 70%	<b>Out – Machine Learning process for modeling</b>
35	89	Jiang, Coffee et al	China. Wenzhou Central Hospital and	Prospective cohort	43 (IQR 32, 49)	62	acute respiratory distress syndrome	53 (5)	Alanine aminotransferase , myalgias, hemoglobin, gender, temp,	Tree-based (decision tree (gini index) )	10-fold cross validation	N.A.	Classification accuracy 70%	<b>Out – Machine Learning process for modeling</b>

			Cangnan People's Hospital in Wenzhou, China, dates not reported  inpatients with confirmed covid-19						Na+, K+, lymphocyte count, creatinine, age, white blood count					
36	89	Jiang, Coffee et al	China. Wenzhou Central Hospital and Cangnan People's Hospital in Wenzhou, China, dates not reported  inpatients with confirmed covid-19	Prospective cohort	43 (IQR 32, 49)	62	acute respiratory distress syndrome	53 (5)	Alanine aminotransferase, myalgias, hemoglobin, gender, temp, Na+, K+, lymphocyte count, creatinine, age, white blood count	Tree-based (random forest)	10-fold cross validation	N.A.	Classification accuracy 70%	<b>Out – Machine Learning process for modeling</b>
37	89	Jiang, Coffee et al	China. Wenzhou Central Hospital and Cangnan People's Hospital in Wenzhou, China, dates not reported  inpatients with confirmed covid-19	Prospective cohort	43 (IQR 32, 49)	62	acute respiratory distress syndrome	53 (5)	Alanine aminotransferase, myalgias, hemoglobin, gender, temp, Na+, K+, lymphocyte count, creatinine, age, white blood count	Support vector machine	10-fold cross validation	N.A.	Classification accuracy 80%	<b>Out – Machine Learning process for modeling</b>

38	96	Levy, Richardson et al.	USA. Northwell Health acute care hospitals (New York metropolitan area), 1 March 1 to 12 April.  Inpatients with confirmed covid-19	Retrospective cohort	65 (IQR 54-77)	65	in-hospital mortality (period unspecified)	Unknown	Age, serum blood urea nitrogen, emergency severity index, red cell distribution width, absolute neutrophil count, serum bicarbonate, glucose	Logistic regression (Lasso/Ridge)	Leave-one-out cross validation	N.A.	C index 0.83	In? Sodium and RDW aren't available in RESCATE  Out - Predictors are not available in RESCATE
39	96	Levy, Richardson et al.	USA. Northwell Health acute care hospitals (New York metropolitan area), 1 March 1 to 12 April.  Inpatients with confirmed covid-19	Retrospective cohort	65 (IQR 54-77)	58	in-hospital mortality (period unspecified)	11095	SOFA score	Logistic regression (Lasso/Ridge)	External validation only	Unclear	C index 0.73	In? SOFA in RESCATE – Recalculate  Out - Predictors are not available in RESCATE
40	96	Levy, Richardson et al.	USA. Northwell Health acute care hospitals (New York metropolitan area), 1 March 1 to 12 April.  Inpatients with	Retrospective cohort	Unclear	58	in-hospital mortality (period unspecified)	N.A.	CURB-65 score	Logistic regression (Lasso/Ridge)	External validation only	Unclear	C index 0.74	In? Calculate CURB-65 in RESCATE  Out - Predictors are not

			confirmed covid-19											available in RESCATE
4 1	96	Levy, Richardson et al.	USA. Northwell Health acute care hospitals (New York metropolitan area), 1 March 1 to 12 April.  Inpatients with confirmed covid-19	Retrospective cohort	Unclear	58	in-hospital mortality (period unspecified)	N.A.	SOFA+ score	Logistic regression (Lasso/Ridge)	External validation only	Unclear	C index 0.83	In? SOFA in RESCATE – Recalculate  Out - Predictors are not available in RESCATE
4 2	98	Liu, Fang et al.	China.  inpatients with confirmed covid-19	¿	¿	¿	in-hospital mortality (period unspecified)	340 (30)	Age, underlying disease status, helper T cells, Helper T cells and Suppressor T cells ratio	Logistic regression	Apparent performance only	N.A.	McFadden pseudo R-squared 0.35	Out – Predictors are not available in RESCATE
4 3	100	McRae, Simmons et al.	China. development: Wuhan, China Hospital, external validation: Hospital in Shenzhen, China, dates unclear.  inpatients with confirmed covid-19	Dev: Retrospective cohort, Val: Unclear	Unclear	discharged: 44, died: 70	in-hospital mortality (period unspecified)	160 (43)	Age, sex, cardiac troponin I, CRP, procalcitonin, myoglobin	Logistic regression (Lasso)	New centres in China, case series	12 (unknown)	C index 0.94 (0.89 to 0.99)	Out - Predictors are not available in RESCATE



4	108	Singh, Valley et al	USA. Michigan Medicine, between 9 March, 2020 and 7 April, 2020.  Inpatients with confirmed covid-19	Dev: Retrospective cohort , Val: Registry data	61 (IQR 50 to 71)	58	ICU-level care, mechanical ventilation or inhospital mortality (period unspecified)	Unknown	Epic Deterioration Index	N.A.	External validation only	174 (61)	C index 0.76 (0.68 to 0.84), sensitivity 39 PPV 80	<b>Out - Predictors are not available in RESCATE</b>
4	112	Vaid, Somani et al	USA. Mount Sinai Health System, 9 March to 11 April.  inpatients with confirmed covid-19	Unclear	62 (IQR 49 to 92)	58	intubation, discharge to hospice care or mortality (period unspecified)	1225 (37)	Sex, race, ethnicity, age, hypertension, atrial fibrillation, coronary artery disease, heart failure, stroke, chronic kidney disease, diabetes, asthma, COPD, cancer, heart rate, pulse, oximetry, respiration rate, temperature, systolic blood pressure, diastolic blood pressure, body weight, sodium, potassium, creatinine, lactate, white blood cells, lymphocyte percentage, haemoglobin, red blood cell distribution width, platelets, alanine, aminotransferase , aspartate, aminotransferase , albumin, total bilirubin, prothrombin time,	Tree-based (XGBoost)	External validation, new centres (USA)	1830 (unknown)	C index 0.84, sensitivity 86, specificity 82	<b>Out – Machine Learning.</b>

									partial thromboplastin time, PCO2, pH, CRP, ferritin, Ddimer, creatinine phosphokinase, lactate dehydrogenase, procalcitonin, troponin I					
4 6	113	Vazquez Guillamet, Vazquez Guillamet et al.	USA. Medical Decision Network, Charlottesville & Barnes Jewish Hospital dataset, 1 January 2015 to 30 September 2015 & 2016 to 2019.  target population unclear	Dev: Retrospective cohort Val: Unclear	Unclear	47	in-hospital mortality (period unspecified)	2122 (429)	Age, immunosuppression, COPD, congestive heart failure, BMI, sex, time to mechanical ventilation (days), length of hospital stay prior to hospital admission, PaO2/FiO2, Glasgow coma scale, maximum Heart rate, maximum respiratory rate, minimum mean arterial blood pressure, maximum temperature, minimum albumin, minimum pH	Logistic regression	External validation, new centres (USA)	1175 (154)	C index 0.81, PPV 55, NPV 89	In? - <b>Differences in selection group: Only ICU patients. Not all with COVID-19</b>  <b>Differences in characteristics of predictors: "Max HR – RR" or "Min pH"</b>  <b>Out - Derived from indirect population or in specific scenarios</b>
4 7	113	Vazquez Guillamet, Vazquez Guillamet et al.	USA. Medical Decision Network, Charlottesville &	Retrospective cohort	Unclear	47	mechanical ventilation > 96 hours	2167 (158)	Age, immunosuppression, COPD, congestive heart failure, BMI, sex, time to	Logistic regression	Training test split	1063 (96)	C index 0.81	In?  <b>Differences in selection group: Only</b>

			Barnes Jewish Hospital dataset, 1 January 2015 to 30 September 2015 & 16 to 2019.  target population unclear						mechanical ventilation (days), length of hospital stay prior to hospital admission, PaO2 /FiO2, Glasgow coma scale, maximum Heart rate, maximum respiratory rate, minimum mean arterial blood pressure, maximum temperature, minimum albumin, minimum pH					ICU patients. Not all with COVID-19  Differences in characteristics of predictors: "Max HR – RR" or "Min pH"  Out - Derived from indirect population or in specific scenarios
48	113	Vazquez Guillamet, Vazquez Guillamet et al.	USA. Medical Decision Network, Charlottesville & Barnes Jewish Hospital dataset, 1 January 2015 to 30 September 2015 & 2016 to 2019.  target population unclear	Retrospective cohort	Unclear	47	mechanical ventilation > 96 hours	1169 (141)	Age, immunosuppression, COPD, congestive heart failure, BMI, sex, time to mechanical ventilation (days), length of hospital stay prior to hospital admission, PaO2 /FiO2, Glasgow coma scale, maximum Heart rate, maximum respiratory rate, minimum mean arterial blood pressure, maximum temperature, minimum	Logistic regression	Training test split	619 (90)	C index 0.78	In? - Differences in selection group: Only ICU patients. Not all with COVID-19  Differences in characteristics of predictors: "Max HR – RR" or "Min pH"

									albumin, minimum pH					Out - Derived from indirect population or in specific scenarios
4 9	116	Zhang, Shi et al	China and United Kingdom. Wuhan Sixth Hospital ;Taikang Tongji Hopsital; King's College Hospital, 1 February to 23 February, 1 March to 8 April.  inpatients with confirmed covid-19	Unclear	61 (IQR 50 to 68)	49	in hospital mortality (period unspecifie d)	653 (20)	Age, sex,	Logistic regressio n (Lasso)	External validation (new centres, different country)	226 (77)	C index 0.75, sensitivity 23, specificity 95, PPV 69, NPV 71	IN
5 0	116	Zhang, Shi et al	China and United Kingdom. Wuhan Sixth Hospital ;Taikang Tongji Hopsital; King's College Hospital, 1	Unclear	61 (IQR 50 to 68)	49	ARDS, intubation or ECMO, ICU admission, in hospital mortality (period unspecifie d)	768 (72)	Age, sex, chronic lung disease, diabetes mellitus, malignancy, cough, dyspnoea, immunocomprom ised, hypertension, heart disease, chronic renal disease, fever, fatigue, diarrhoea	Logistic regressio n (Lasso)	Repeated five-fold cross validation	N.A.	C index 0.80, sensitivity 9, specificity 99 PPV 53, NPV 91	Out - Predictors are not available in RESCATE

#	Ref	Authors	Dataset Setting	Study design	Age mean ± se, median (IQR), or range	Sex (% male)	Outcome	Sample size total No of 125iarrhea 125125a for model development set (No with outcome)	Predictors in final model	Modelling technique	Type of validation	Sample size: total No of participants for model validation (No with outcome)	Performance* (C index, sensitivity (%), specificity (%), PPV/ NPV (%), calibration slope, other (95% CI, if reported))	Selection Process (Razon for in/out)
			February to 23 February, 1 March to 8 April.  inpatients with confirmed covid-19											
1	9	Bai, Fang, et al.	China Wuhan Pulmonary Hospital, January 3 to February 13.  Inpatients at admission with mild confirmed covid-19	Retrospective cohort	53 ± 13	50%	Deterioration into severe/critical disease (period unspecified)	133 (54)	Combination of demographics, signs and symptoms, laboratory results and features derived from CT images	Neural net (Multilayer perceptron + long short term memory vs. logistic regression, linear discriminant analysis, support vector machine, multilayer	Unclear	N.A.	C index 0.95 (0.94 to 0.97)	Out – No mortality outcome

										perceptro n)				
2	19	Caramelo, Ferreira, et al.	China. Simulated based on data from Wuhan. December 8 to February 11	Simulation	Unclear	41%	Mortality (period unspecified) *2	Unknown	Age, sex, presence of any comorbidity (hypertension, diabetes, cardiovascular disease, chronic respiratory disease, cancer) *3	Logistic regression	N.A.	Not reported	Not reported	<b>Out – Simulation</b>
3		Gong, Ou, et al:	China, Guangzhou Eighth People's Hospital, Zhongnan Hospital of Wuhan University, Third Affiliated Hospital of Sun Yat-sen University, January 20 to March 2	Retrospective cohort	Dev.: 49 Val.1: 52 Val.2: 42	Dev.: 47% Val.1: 44% Val.2: 50%	Severe COVID-19 infection (minimum 15 day)		Age, serum LDH, CRP, variation of red blood cell distribution width, blood urea nitrogen, albumin, direct bilirubin	LASSO followed by logistic regression (vs. LASSO followed by decision tree, random forest or support vector machine)				<b>Out – No mortality outcome</b>
4	20	Lu, Hu, et al.	China Wuhan Hankou Hospital, January 21 to February 4.  Inpatients at admission with suspected or confirmed covid-19	Retrospective cohort  One center	55 (IQR 39-66)	44%	Mortality (12 day)	577 (44)	Age, C-reactive protein	Cox regression	Not reported	Not reported	Not reported	In? – ACP Scale – Simple – Easy. NON Validated. NO PEER REVIEW ***

5	21	Qi, Jiang, et al.	China, 5 hospitals from Ankang, Lishui, Zhenjiang, Lanzhou, Linxia between January 23 to February 8; date of last follow-up February 20.  Inpatients with confirmed covid-19 at admission	Prospective cohort	38 (IQR 26 to 47)	55%	Hospital stay >10 days	26 (20)	6 features derived from CT images *3	Logistic regression	Fivefold cross validation	N.A.	C index 0.92	Out – No mortality outcome
	21	Qi, Jiang, et al.	China, 5 hospitals from Ankang, Lishui, Zhenjiang, Lanzhou, Linxia between January 23 to February 8; date of last follow-up February 20.  Inpatients with confirmed covid-19 at admission.  Inpatients with	Prospective cohort	38 (IQR 26 to 47)	55%	Hospital stay >10 days	26 (20)	6 features derived from CT images *3	Tree-based (random forest)	Fivefold cross validation	N.A	C index 0.96	Out – No mortality outcome

			confirmed covid-19 at admission											
6	37	Shi, Yu, et al.	China, Hangzhou, Zhejiang Province, study dates unspecified, follow-up until February 17.  Inpatients with confirmed covid-19 at admission	Retrospective cohort	46 ± 19	53%	Death or severe COVID-19 (period unspecified)	478 (49)	Age (dichotomized), sex, hypertension	Multivariate model (not specified)	Validation in 128ia severe cases	66 (15)	Not reported	In? Low power.  Validation?  No report of S-E-C-statistic
7	7	Xie, Hungerford, et al	China, Tongji Hospital and Jinyintan Hospital, admitted between January and February  inpatients with confirmed covid-19 at admission	Retrospective cohort	Tongji: 65 (IQR 54 to 73) Jinyintan: 56 (IQR 47 to 68)	Tongji: 52% Jinyintan: 35%	Mortality (in hospital)	299 (155)	Age, LDH, lymphocyte count, SPO2	Logistic regression	External validation (other Chinese centre)	130 (69)	C index 0.98 (0.96 to 1.00),  calibration slope 2.5 (1.7 to 3.7)	In ***
8	22	Yan, Zhang, et al.	China, Tongji Hospital Wuhan, January 10 to February 18	Retrospective cohort	59 ± 16	59%	Mortality (period unspecified)	375 (174)	Lactic dehydrogenase, lymphocyte count, high sensitivity C-reactive protein	Tree-based (XGBoost)	Temporal validation, selecting only severe cases	29 (17)	Sensitivity 92, PPV 95	Out – Machine Learning



			inpatients suspected of covid-19											
9	23	Yuan, Yin, et al.	China, Hubei Public Health Clinical Centre; central Hospital of Wuhan, January 1 to January 25.  Inpatients with confirmed covid-19 at admission	Retrospective cohort	60 (IQR 47 to 69)	44%	Mortality (period unspecified)	Not 129 diarrhea 129	Clinical scorings of CT images (zone, left/right, location, attenuation, distribution of affected parenchyma)	NA (external validation)	External validation of existing model	27 (10)	C index 0.90 (0.87 to 0.93)	Out – Requires CT scoring system
10	60	Huang, Cai et al.	China, Guangzhou 8 <sup>th</sup> People's Hospital, January 20 to February 29.  Inpatients with confirmed covid-19 at admission	Retrospective cohort	45 ± 19	50%	severe symptoms 3 days after admission	125 (32)	Underlying diseases, fast respiratory rate >24/min, elevated CRP-level (> 10mg/dL), elevated lactate dehydrogenase level (> 250U/L)	Logistic regression	Apparent performance only	Not 129 diarrhea 129	C index 0.99 (0.97 to 1.00),  sensitivity 91, specificity 96	Out – No mortality outcome
11	61	Pourhomayoun, Shakibi et al.	76 countries. Publicly available data on Github from Cov2019 Data Working Group with	Unclear	57 (mean)	Unclear	in-hospital mortality (period unspecified)	Unknown	Unknown	Neural net (Neural Networks vs. Support Vector Machine (SVM), Random Forest,	10-fold cross validation	N.A.	C index 0.96, sensitivity 90, High specificity 97	Out – Machine Learning process for modeling

			confirmed cases worldwide in a variety of settings.  Inpatients with confirmed covid-19							Decision Tree, Logistic Regression, and Knearest Neighbour (KNN))				
1 2	44	Sakar, Chakrabarti	Several Countries. Kaggle data with cases from 22 countries in Asia, Australia, Europe, North America, January 13 to February 28.  Inpatients with covid-19 symptoms	Unclear	Unclear	Unclear	death vs recovery (period unspecified)	80 (37)	Age, days from symptom onset to 130iarrrhea130130ation, from Wuhan, sex, visit to Wuhan	Tree-based (Random forest)	Apparent performance only	N.A.	C index 0.97	<b>Out – Machine Learning process for modeling</b>
1 3	42	Wang, Zha et al.	China. Renmin Hospital of Wuhan University, Henan Provincial People's Hospital, Beijing Youan Hospital of Capital Medical University, Huangshi Central Hospital).	Retrospective cohort	Dev. 51 ± 19; Val. 50 ± 19, 48 ± 14	Dev. 48%; Val. 47%, 51%	length of hospital stay	301 (not 130iarrrhea 130)	Age and CT features	Cox regression	Not reported	N.A.	Not reported	<b>Out – No mortality outcome</b>

			Inpatients with confirmed covid-19											
14	62	Zeng, Li et al.	Third People's hospital of Shenzhen (Shenzhen, China) from January 11 to February 29.  Inpatients with confirmed covid-19	Retrospective cohort	Severe: 58.7 (11), nonsevere: 46.1 (14.1)	Severe: 62% non-severe: 44%	severe disease progression (period unspecified)	338 (76)	CT features	LASSO followed by Fine and Gray (for both analysis)	Cross validation (number of folds unclear)	N.A.	C index 0.88	Out – No mortality outcome
	62	Zeng, Li et al.	Third People's hospital of Shenzhen (Shenzhen, China) from January 11 to February 29.  Inpatients with confirmed covid-19	Retrospective cohort	Severe: 58.7 (11), nonsevere: 46.1 (14.1)	Severe: 62% non-severe: 44%	severe disease progression (period unspecified)	338 (76)	CT features and laboratory markers	LASSO followed by Fine and Gray (for both analysis)	Cross validation (number of folds unclear)	N.A.	C index 0.88	Out – No mortality outcome
15	63	Al – Najjar, Al-Rousan	South Korea. Official time series data from the Korea Centers for Disease Control and Prevention	Registry data	Unclear	Unclear	recovery from covid19 (period unspecified)	466 (40)	birth year (age), sex, country, group, infection reason, confirmed date	Neural net (one hidden layer and gradient descent as an optimization)	Training test 131iarr	193 (14)	Sensitivity 43, specificity 98	Out – No mortality outcome

			(KCDC), Unclear start date of inclusion to 9 March							algorithm )				
1 5	63	Al – Najjar, Al-Rousan	South Korea. Official time series data from the Korea Centers for Disease Control and Prevention (KCDC), Unclear start date of inclusion to 9 March	Registry data	Unclear	Unclear	mortality (period unspecifie d)	463 (25)	Age, sex, country, region, infection reason, confirmed date	Neural net (one hidden layer and gradient descent as an optimizati on algorithm )	Training test 132iarr	191 (7)	Sensitivity 86, specificity 100	<b>Out – Machine Learning process for modeling</b>
1 6	67	Barda, Riesel et al	Israel. Data warehouse of Clalit Health Services (payer- provider), 29 January to 8 April inclusion, follow-up until 22 April (covid cases for	Retrospec tive cohort	41 ± 21.26	49%	mortality (period unspecifie d)	735,000 (8251)	Age, sex, pack years, COPD, number of wheezing/dyspne a diagnoses, albumin, red cell distribution width, c-reactive peptide, urea, lymphocyte, chloride, creatinine, high density lipoprotein, duration of hospitalizations,	Tree- based (gradient boosted tree)	Other (specify in column CL) January 29 to April 8 inclusion, follow-up until April 22 2020 (covid cases for external validation)	3176 (87)	C index 0.94 (0.92 to 0.96) , sensitivity 90 (83 to 96), PPV 17 (14 to 21)	<b>Out – nonindividu al assessment</b>

			external validation)  patients with confirmed covid-19						count of hospitalizations, count of ambulance rides, count of sulfonamide dispenses, count of anticholinergic dispenses, count of glucocorticoid dispenses, chronic respiratory disease, cardiovascular disease, diabetes, malignancy, hypertension					
17	70	Bello-Chavolla, Bahena Lopez et al	Mexico. General Directorate of Epidemiology of Mexican Ministry of Health, dates not reported.  Confirmed COVI-19 patients presenting at GP	Surveillance data	47 ± 16	58	30-day mortality	12424 (1137)	Age, pregnancy, diabetes, obesity, pneumonia, CKD, COPD, immunosuppression	Cox proportional hazards regression	Training test 133iarr	3105 (297)	C index 0.80, Somer's D 0.60	In – Sociodemographic risk factors. Power.
18	75	Carr, Bendayan et al	United Kingdom. King's College Hospital NHS Foundation Trust hospital, 1 March 1 to 5 April.	Retrospective cohort	67 ± 28	55	progression to severe covid-19 (period unspecified)	452 (159)	Age, National Early Warning Score (NEWS) 2, CRP, neutrophil, eGFR, albumin	Logistic regression	Temporal validation	256 (59)	C index 0.73, sensitivity 46, specificity 87	Out – No mortality outcome

			Inpatients with confirmed covid-19											
19	77	Chassagnon, Vakalopoulos, et al.	France. Unclear, 4 March to 29 March  inpatients with confirmed covid-19	Retrospective cohort	63 ± 16	67	composite, 4-day intubation or mortality	383 (84)	Unclear	Support Vector Machines, Decision Trees, Random Forests, AdaBoost, and Gaussian I Bayes. Prediction made by winner takes all principle (Lasso/Ridge)	External validation (new centres, France)	95 (26)	Sensitivity 88, specificity 74	<b>Out – Machine Learning process for modeling</b>
20	79	Colombi, Bodini et al	Italy. “Guglielmo da Saliceto” Hospital, Piacenza, Italy, 17 February to 10 March.  Inpatients with confirmed covid-19	Retrospective cohort	Unclear	75	ICU admission or inhospital (period unspecified)	236 (108)	Age, cardiovascular comorbidities, median platelet count, CRP, visual assessment of well aerated lung %	Logistic regression	Apparent performance only	N.A.	C index 0.86 (0.81 to 0.90), sensitivity 72 (63 to 80), specificity 81 (73 to 88) PPV 70 (61 to 78), NPV 78 (72 to 83)	<b>Out – Requires CT scoring system by Radiologist</b>
20	79	Colombi, Bodini et al	Italy. “Guglielmo da Saliceto” Hospital, Piacenza, Italy, 17 February to 10 March	Retrospective cohort	Unclear	75	ICU admission or inhospital mortality (period unspecified)	236 (108)	Age, cardiovascular comorbidities, median platelet count, LDH, CRP, software assessment of well aerated lung absolute volume, adipose tissue	Logistic regression	Apparent performance only	N.A.	C index 0.86 (0.81 to 0.90), sensitivity 75 (66 to 83), specificity 81 (73 to 88), PPV 70 (61 to 78),	<b>Out – Requires CT scoring system by Radiologist</b>

			inpatients with confirmed covid-19										NPV 78 (72 to 83)	
21	80	Das, Mishra, et al	South Korea. Data shared by Korea Centers for Disease Control and Prevention, 20 January to 7 April 7  inpatients with confirmed covid-19	Retrospective cohort	Unclear	44	ICU admission or inhospital mortality (period unspecified)	3022 (61)	Age, sex, province, date of diagnosis, place of exposure to covid-19	Gradient boosting algorithm	Training test 135iarr	604 (12)	C index 0.97	<b>Out – Machine Learning process for modeling</b>
	119	Gong, Ou et al.	China. Target population unclear.	Retrospective cohort	¿	?	15-day progression to severe covid-19	189 (28)	Age, direct bilirubin, red cell distribution width, blood urea nitrogen, CRP, lactate dehydrogenase, albumin	Logistic regression	External validation (new centres, China)	165 (40)	C index 0.85 (0.79 to 0.92), sensitivity 78, specificity 78	<b>Out – No mortality outcome</b>
22	83	Guo, Liu et al.	China. 35 hospitals in Guangdong Province and Hubei Province, 27 December 2019, to 4 March 2020.  Inpatients with confirmed covid-19	Retrospective cohort	42 (IQR 31 to 57)	48	14-day progression to severe covid-19	818 (24)	Age, chronic illness, neutrophil to lymphocyte ratio, CRP, D-dimer	Cox proportional hazards regression	External validation (new centres, China)	320 (38)	C index 0.78 (0.70 to 0.87)	<b>Out – No mortality outcome *** Estudio bien hecho pero solo evalúa progresión o curso clínico desfavorable ***</b>  <b>Score EWAS</b>

23	84	Hu, Liu et al	China. Sino-French New City Branch of Tongji Hospital (n=183; Dev) and Optical Valley Branch of Tongji Hospital (n=64; Val), 28 January 2020 to 11 March 2020 (dev)  inpatients with confirmed covid-19	Dev: Retrospective cohort, Val: Unclear	68 ± 10 in nonsurvivors; 61 ± 13 in survivors in nonsurvivors	58	in-hospital mortality (period unspecified)	183 (68)	Age, high-sensitivity CRP, lymphocyte count, Ddimer	Logistic regression	External validation (new centres, China)	64 (31)	C index 0.88, sensitivity 84, specificity 79	<b>Out – Machine Learning process for modeling</b>
24	85	Hu, Yao et al	China. Electronic patient care reports from an emergency medical team that collected in Wuhan from 7 February 7 to 7 March  inpatients with confirmed covid-19	Retrospective cohort	61 ± 16	51	in-hospital mortality (period unspecified)	N.A.	Modified Early Warning Score (MEWS): heart rate, systolic blood pressure, respiratory rate, body temperature, consciousness	?	External validation only	105 (19)	C index 0.68 (0.58 to 0.77), sensitivity 68, specificity 65, PPV 30, NPV 90	<b>** Evaluated MEWS – REMS as clinical scores we don't have in RESCATE. But maybe we can calculate them.</b>



	85	Hu, Yao et al	China. Electronic patient care reports from an emergency medical team that collected in Wuhan from 7 February 7 to 7 March  inpatients with confirmed covid-19	Retrospective cohort	61 ± 16	51	in-hospital mortality (period unspecified)	N.A.	Rapid Emergency Medicine Score (REMS): mean arterial pressure, pulse rate, respiratory rate, oxygen saturation, GCS, age	?	External validation only	105 (19)	C index 0.84 (0.76 to 0.91), sensitivity 89, specificity 70, PPV 40, NPV 97	** Evaluated MEWS – REMS as clinical scores we don't have in RESCATE. But maybe we can calculate them.
25	88	Ji, Zhang et al	China. Fuyang Second People's Hospital (FYSPH), Anhui; the fifth medical center of Chinese PLA general hospital (PLAGH), Beijing, Unclear; data collected from 20 January through 22 February.  Inpatients with confirmed covid-19	Retrospective cohort	44 ± 16	56	10-day progression to severe COVID-19	208 (40)	Comorbidity, age, lymphocyte count, lactate dehydrogenase	Cox proportional hazards regression	Internal validation by resampling (bootstrap)	N.A.	C index 0.91 (0.86 to 0.94), sensitivity 95 (83 to 99), specificity 78 (71 to 84)	Out – No mortality outcome

26	89	Jiang, Coffee et al	China. Wenzhou Central Hospital and Cangnan People's Hospital in Wenzhou, China, dates not reported  inpatients with confirmed covid-19	Prospective cohort	43 (IQR 32, 49)	62	acute respiratory distress 138 diarrhea	53 (5)	Alanine aminotransferase, myalgias, hemoglobin, gender, temp, Na+, K+, lymphocyte count, creatinine, age, white blood count	Logistic regression	10-fold cross validation	N.A.	Classification accuracy 50%	Out – Machine Learning process for modeling
	89	Jiang, Coffee et al	China. Wenzhou Central Hospital and Cangnan People's Hospital in Wenzhou, China, dates not reported  inpatients with confirmed covid-19	Prospective cohort	43 (IQR 32, 49)	62	acute respiratory distress 138 diarrhea	53 (5)	Alanine aminotransferase, myalgias, hemoglobin, gender, temp, Na+, K+, lymphocyte count, creatinine, age, white blood count	K nearest neighbor	10-fold cross validation	N.A.	Classification accuracy 80%	Out – Machine Learning process for modeling
	89	Jiang, Coffee et al	China. Wenzhou Central Hospital and Cangnan People's Hospital in Wenzhou, China, dates not reported  inpatients with	Prospective cohort	43 (IQR 32, 49)	62	acute respiratory distress 138 diarrhea	53 (5)	Alanine aminotransferase, myalgias, hemoglobin, gender, temp, Na+, K+, lymphocyte count, creatinine, age, white blood count	Tree-based (decision tree (gain ratio))	10-fold cross validation	N.A.	Classification accuracy 70%	Out – Machine Learning process for modeling

			confirmed covid-19											
89	Jiang, Coffee et al	China. Wenzhou Central Hospital and Cangnan People's Hospital in Wenzhou, China, dates not reported	inpatients with confirmed covid-19	Prospective cohort	43 (IQR 32, 49)	62	acute respiratory distress 139 diarrhea	53 (5)	Alanine aminotransferase, myalgias, hemoglobin, gender, temp, Na+, K+, lymphocyte count, creatinine, age, white blood count	Tree-based (decision tree (gini index))	10-fold cross validation	N.A.	Classification accuracy 70%	Out – Machine Learning process for modeling
89	Jiang, Coffee et al	China. Wenzhou Central Hospital and Cangnan People's Hospital in Wenzhou, China, dates not reported	inpatients with confirmed covid-19	Prospective cohort	43 (IQR 32, 49)	62	acute respiratory distress 139 diarrhea	53 (5)	Alanine aminotransferase, myalgias, hemoglobin, gender, temp, Na+, K+, lymphocyte count, creatinine, age, white blood count	Tree-based (random forest)	10-fold cross validation	N.A.	Classification accuracy 70%	Out – Machine Learning process for modeling
89	Jiang, Coffee et al	China. Wenzhou Central Hospital and Cangnan People's Hospital in Wenzhou, China,		Prospective cohort	43 (IQR 32, 49)	62	acute respiratory distress 139 diarrhea	53 (5)	Alanine aminotransferase, myalgias, hemoglobin, gender, temp, Na+, K+, lymphocyte count, creatinine, age, white blood count	Support vector machine	10-fold cross validation	N.A.	Classification accuracy 80%	Out – Machine Learning process for modeling

			dates not reported											
27	96	Levy, Richardson et al.	USA. Northwell Health acute care hospitals (New York metropolitan area), 1 March 1 to 12 April.  Inpatients with confirmed covid-19	Retrospective cohort	65 (IQR 54-77)	65	in-hospital mortality (period unspecified)	Unknown	Age, serum blood urea nitrogen, emergency severity index, red cell distribution width, absolute neutrophil count, serum bicarbonate, glucose	Logistic regression (Lasso/Ridge)	Leave-one-out cross validation	N.A.	C index 0.83	<b>**In? Sodium and RDW aren't available in RESCATE....</b>  <b>There's a new paper-</b>
	96	Levy, Richardson et al.	USA. Northwell Health acute care hospitals (New York metropolitan area), 1 March 1 to 12 April.  Inpatients with confirmed covid-19	Retrospective cohort	65 (IQR 54-77)	58	in-hospital mortality (period unspecified)	11095	SOFA score	Logistic regression (Lasso/Ridge)	External validation only	Unclear	C index 0.73	<b>**In? SOFA in RESCATE – Recalculate? Compare?</b>
	96	Levy, Richardson et al.	USA. Northwell Health acute care hospitals (New York metropolitan area), 1	Retrospective cohort	Unclear	58	in-hospital mortality (period unspecified)	N.A.	CURB-65 score	Logistic regression (Lasso/Ridge)	External validation only	Unclear	C index 0.74	<b>**In? Calculate CURB-65 in RESCATE</b>

			March 1 to 12 April.  Inpatients with confirmed covid-19											
	96	Levy, Richardson et al.	USA. Northwell Health acute care hospitals (New York metropolitan area), 1 March 1 to 12 April.  Inpatients with confirmed covid-19	Retrospective cohort	Unclear	58	in-hospital mortality (period unspecified)	N.A.	SOFA+ score	Logistic regression (Lasso/Ridge)	External validation only	Unclear	C index 0.83	<b>**In? SOFA in RESCATE – Recalculate? Compare?</b>
?	98	Liu, Fang et al.	China.  Inpatients with confirmed covid-19	¿	¿	¿	in-hospital mortality (period unspecified)	340 (30)	Age, underlying disease status, helper T cells, Helper T cells and Suppressor T cells ratio	Logistic regression	Apparent performance only	N.A.	McFadden pseudo R-squared 0.35	<b>Out – Predictors are not available in RESCATE</b>
28	100	McRae, Simmons et al.	China. Development: Wuhan, China Hospital, external validation: Hospital in Shenzhen, China, dates unclear.  Inpatients with	Dev: Retrospective cohort, Val: Unclear	Unclear	discharged: 44, died: 70	in-hospital mortality (period unspecified)	160 (43)	Age, sex, cardiac troponin I, CRP, procalcitonin, myoglobin	Logistic regression (Lasso)	New centres in China, case series	12 (unknown)	C index 0.94 (0.89 to 0.99)	<b>Out - Validation of a programmable bio nano chip (p-BNC) with the capacity to learn and adapting it to the task of assessing COVID-19 disease severity</b>

			confirmed covid-19											
29	108	Singh, Valley et al	USA. Michigan Medicine, between 9 March, 2020 and 7 April, 2020.  Inpatients with confirmed covid-19	Dev: Retrospective cohort, Val: Registry data	61 (IQR 50 to 71)	58	ICU-level care, mechanical ventilation or inhospital mortality (period unspecified)	Unknown	Epic Deterioration Index	N.A.	External validation only	174 (61)	C index 0.76 (0.68 to 0.84), sensitivity 39 PPV 80	<b>** Out.</b>  External validation of Epic Deterioration Index (EDI)** Artificial Intelligence non clear origins.  There's a new version.
30	112	Vaid, Somani et al	USA. Mount Sinai Health System, 9 March to 11 April.  Inpatients with confirmed covid-19	Unclear	62 (IQR 49 to 92)	58	intubation, discharge to hospice care or mortality (period unspecified)	1225 (37)	Sex, race, ethnicity, age, hypertension, atrial fibrillation, coronary artery disease, heart failure, stroke, chronic kidney disease, diabetes, asthma, COPD, cancer, heart rate, pulse, oximetry, respiration rate, temperature, systolic blood pressure, diastolic blood pressure, body weight, sodium, potassium, creatinine, lactate, white blood cells, lymphocyte percentage, haemoglobin, red blood cell	Tree-based (XGBoost)	External validation, new centres (USA)	1830 (unknown)	C index 0.84, sensitivity 86, specificity 82	<b>Out – Machine Learning.</b>

									distribution width, platelets, alanine, aminotransferase, aspartate, aminotransferase, albumin, total bilirubin, prothrombin time, partial thromboplastin time, PCO2, pH, CRP, ferritin, Ddimer, creatinine phosphokinase, lactate dehydrogenase, procalcitonin, troponin I					
31	113	Vazquez Guillamet, Vazquez Guillamet et al.	USA. Medical Decision Network, Charlottesville & Barnes Jewish Hospital dataset, 1 January 2015 to 30 September 2015 & 2016 to 2019.  Target population 143iarrhe	Dev: Retrospective cohort, Val: Unclear	Unclear	47	in-hospital mortality (period unspecified)	2122 (429)	Age, immunosuppression, COPD, congestive heart failure, BMI, sex, time to mechanical ventilation (days), length of hospital stay prior to hospital admission, PaO2 /FiO2, Glasgow coma scale, maximum Heart rate, maximum respiratory rate, minimum mean arterial blood pressure, maximum temperature, minimum albumin, minimum pH	Logistic regression	External validation, new centres (USA)	1175 (154)	C index 0.81, PPV 55, NPV 89	In? - Differences in selection group: Only ICU patients. Not all with COVID-19  Differences in characteristics of predictors: "Max HR - RR" or "Min pH"
32	113	Vazquez Guillamet, Vazquez Guillamet et al.	USA. Medical Decision Network, Charlottesville & Barnes	Retrospective cohort	Unclear	47	mechanical ventilation > 96 hours	2167 (158)	Age, immunosuppression, COPD, congestive heart failure, BMI, sex, time to mechanical	Logistic regression	Training test 143iarr	1063 (96)	C index 0.81	In? - Differences in selection group: Only ICU

			Jewish Hospital dataset, 1 January 2015 to 30 September 2015 & 2016 to 2019.						ventilation (days), length of hospital stay prior to hospital admission, PaO2 /FiO2, Glasgow coma scale, maximum Heart rate, maximum respiratory rate, minimum mean arterial blood pressure, maximum temperature, minimum albumin, minimum pH					patients. Not all with COVID-19
			Target population 144iarrhe											Differences in characteristics of predictors: "Max HR – RR" or "Min pH"
3 3	113	Vazquez Guillamet, Vazquez Guillamet et al.	USA. Medical Decision Network, Charlotteville & Barnes Jewish Hospital dataset, 1 January 2015 to 30 September 2015 & 2016 to 2019.	Retrospective cohort	Unclear	47	mechanical ventilation > 96 hours	1169 (141)	Age, immunosuppression, COPD, congestive heart failure, BMI, sex, time to mechanical ventilation (days), length of hospital stay prior to hospital admission, PaO2 /FiO2, Glasgow coma scale, maximum Heart rate, maximum respiratory rate, minimum mean arterial blood pressure, maximum temperature, minimum albumin, minimum pH	Logistic regression	Training test 144iarr	619 (90)	C index 0.78	In? - Differences in selection group: Only ICU patients. Not all with COVID-19
			Target population 144iarrhe											Differences in characteristics of predictors: "Max HR – RR" or "Min pH"
3 4	116	Zhang, Shi et al	China and United Kingdom. Wuhan Sixth Hospital ;Taikang	Unclear	61 (IQR 50 to 68)	49	in hospital mortality (period unspecified)	653 (20)	Age, sex,	Logistic regression (Lasso)	External validation (new centres, different country)	226 (77)	C index 0.75, sensitivity 23, specificity	IN***



			Tongji Hopsital; King's College Hospital, 1 February to 23 February, 1 March to 8 April.  Inpatients with confirmed covid-19										95, PPV 69, NPV 71	
		Zhang, Shi et al	China and United Kingdom. Wuhan Sixth Hospital ;Taikang Tongji Hopsital; King's College Hospital, 1 February to 23 February, 1 March to 8 April.  Inpatients with confirmed covid-19	Unclear	61 (IQR 50 to 68)	49	ARDS, intubation or ECMO, ICU admission, in hospital mortality (period unspecified)	768 (72)	Age, sex, chronic lung disease, diabetes mellitus, malignancy, cough, 145iarrhe, immunocompromised, hypertension, heart disease, chronic renal disease, fever, fatigue, 145iarrhea	Logistic regression (Lasso)	Repeated five-fold cross validation	N.A.	C index 0.80, sensitivity 99, specificity 99 PPV 53, NPV 91	<b>Out – No mortality outcome</b>
		Zhang, Shi et al	China and United Kingdom. Wuhan Sixth Hospital ;Taikang	Unclear	61 (IQR 50 to 68)	49	ARDS, intubation or ECMO, ICU admission, in hospital	653 (58)	Age, sex, neutrophil count, lymphocyte count, platelet count, CRP, creatinine	Logistic regression (Lasso)	External validation (new centres, different country)	226 (97)	C index 0.72, sensitivity 40, specificity	<b>Out – No mortality outcome</b>

			Tongji Hopsital; King's College Hospital, 1 February to 23 February, 1 March to 8 April.				mortality (period unspecifie d)						85, PPV 67, NPV 65	
			Inpatients with confirmed covid-19											

### Anexo 3: Proceso de caracterización variables

#### Proceso de depuración de variables:

Se utilizó la herramienta STATA 14.

1. Se definió la variable “edad” como (fecha ingreso paciente – fecha nacimiento: en números enteros).
2. La variable “sexopte” se renombró como “sex” y se modificó el código de 1= masculino, 2 = femenino por 1= masculino y 0= femenino. (recode (2=0) (1=1)).
3. La variable “resultadopcr” se renombró como “pcr” y se modificó el código de 1= positiva, 2 = negativa por 1= positiva, 0= negativa. (recode (2=0) (1=1))
4. Se renombraron las siguientes variables:
  - antecedentes\_\_1 = asma
  - antecedentes\_\_2 = epoc
  - antecedentes\_\_3 = otroresp
  - antecedentes\_\_4 = dm2
  - antecedentes\_\_5 = vih
  - antecedentes\_\_6 = cardiopat
  - antecedentes\_\_7 = cancer
  - antecedentes\_\_8 = dnt
  - antecedentes\_\_9 = obesidad
  - antecedentes\_\_10 = erc
  - antecedentes\_\_11 = tabaq
  - antecedentes\_\_12 = tbc
  - antecedentes\_\_13 = hta
  - antecedentes\_\_14 = embarazo
  - antecedentes\_\_15 = inmunosupre
  - antecedentes\_\_16 = fcoinmunosupre
  - antecedentes\_\_17 = iecara2
  - antecedentes\_\_18 = aines
  - antecedentes\_\_19 = otrofco
  - antecedentes\_\_20 = nofco
5. Todas las variables de síntomas al ingreso se renombraron retirándoles el “ingreso” ya que únicamente se utilizarán esas variables como síntomas en las escalas de predicción.
  - Fiebreingreso = fiebre
  - Tosingreso = tos
  - Fatigaingreso = fatiga
  - Expectoraingreso = expectora
  - Disneaingreso = disnea

- Disgeusiaingreso = disgeusia
- Cefaleaingreso = cefalea
- Dolorabdingreso = dolorabd
- Diarreaingreso = diarrea
- Nauseaingreso = nausea
- Odinofagiaingreso = odinofag
- Anosmiaingreso = anosmia

6. A todas las variables anteriores codificadas como “síntomas al ingreso” se les modificó el código de 1= positiva, 2 = negativa por 1= positiva, 0= negativa. (recode (2=0) (1=1)).

7. Se renombraron las siguientes variables de escenarios de manejo:

- “hospitaliza” por “hospitalización” (También se codificó como (2=0) (1=1))
- “serviciohosp\_1” por “hospipiso”
- “serviciohosp\_2” por “ucinter”
- “serviciohosp\_3” por “uci”

8. Se renombraron las siguientes variables de “curso clínico desfavorable”

- “procedimientohospital\_\_9” por “vasopinotrop”
- “procedimientohospital\_\_10” por “vmi”
- “complicacioneshospital\_\_2” por “miocarditis”
- “complicacioneshospital\_\_3” por “coinfec”
- “complicacioneshospital\_\_4” por “Ira”: lesión renal aguda
- “complicacioneshospital\_\_5” por “nocomplic”: no complicaciones
- “otrascomplicaciones\_\_1” por “sdra”
- “otrascomplicaciones\_\_2” por “tep”
- “otrascomplicaciones\_\_3” por “tvp”

9. Se mantiene clasificación de variables politómicas:

- Variable “gravedadadulto” como nominal politómica y se :

1: Enfermedad no complicada; 2: Neumonía leve; 3: Neumonía grave; 4: Síndrome de dificultad respiratoria aguda SDRA;5: Sépsis; 6: Choque séptico.

- Variable “gravedadnino” como nominal politómica:

1: Asintomático; 2: Leve; 3: Grave; 4: Crítico

**Anexo 4:** Protocolo Proyecto RESCATE COVID-19: Respuesta de Conocimiento para la Atenuación de la Epidemia por COVID-19.

Acceso en línea:


<http://www.subredsur.gov.co/sites/default/files/documentos/PROYECTO%20RESCATE%20COVID%2019.pdf>

**Proyecto RESCATE COVID-19: Respuesta de Conocimiento para la Atenuación de la Epidemia por COVID-19**

**Institución proponente y centro coordinador nacional:** Fundación Cardioinfantil - Instituto de Cardiología

<b>Información general del proyecto</b>
<p><b>Título del proyecto:</b> RESCATE COVID-19: Respuesta de Conocimiento para la Atenuación de la Epidemia por COVID-19</p>
<p><b>Grupos de Investigación FCI-IC:</b> Grupo de investigación en Cardiología - COL0008737 Grupo de Investigación en la Práctica Clínica de Cuidado y Rehabilitación de Enfermería (Practicare) - COL0193475</p>
<p><b>Grupos de otras instituciones:</b> RAM (Red para el Uso Adecuado de Medicamentos)- Universidad Nacional de Colombia - COL0063479 Evaluación de tecnologías y políticas en salud - Universidad Nacional de Colombia - COL0003964 Grupo Académico de Epidemiología Clínica – Universidad de Antioquia - COL0007121 Grupo de Investigación en Salud Pública, Educación y Profesionalismo (S.E.P.) – Universidad de los Andes - COL0112922</p>
<p><b>Investigador principal:</b> Juan Carlos Villar Centeno <b>Coinvestigadores FCI-IC:</b> Karen Julieth Moreno Medina, Olga Lucía Cortés Ferreira, Eliana Yineth Váquiro Herrera.</p>
<p><b>Coinvestigadores de otras instituciones:</b> Hernando Guillermo Gaitán Duarte y Claudia Vacca González (Universidad Nacional de Colombia), Claudia Marcela Vélez y Daniel Felipe Patiño Lugo (Universidad de Antioquia), Juan José Yepes Núñez y Juan Carlos Briceño (Universidad de los Andes).</p>
<p><b>Profesionales del equipo formulador (FCI-IC):</b> Anamaria Muñoz Flórez, Ángela Manuela Balcázar, Edgar Camilo Barrera Garavito, Lina Góez, Luis David Sáenz Pérez, Luz Ángela Torres López, María Ximena Rojas Reyes,</p>
<p><b>Contacto</b> Correo electrónico: <a href="mailto:investigacionesfci@cardioinfantil.org">investigacionesfci@cardioinfantil.org</a> Teléfono-Extensión: 6672727 ext 73205</p>
<p><i>El proyecto RESCATE ha sido desarrollado por la Fundación Cardioinfantil -IC y sus contenidos son propiedad intelectual de este grupo. Se considera un delito reproducir o utilizar la información y los datos de este protocolo para otro propósito fuera del estudio sin previa autorización de la oficina del proyecto</i></p>

**Anexo 5:** Declaración de confidencialidad firmada por Investigador principal de la Clínica Chicamocha.

	<b>FORMATO DE DECLARACIÓN DE CONFIDENCIALIDAD</b>	Código: Por definir
		Versión: 1
	<b>PROCESO: Investigaciones</b>	Vigencia: 3 años
		Fecha de Vencimiento: 08/2023

## DECLARACIÓN DE CONFIDENCIALIDAD

Yo, Custodio Alberto Ruiz Bedoya identificado con cédula de ciudadanía No.1001282792, actuando en nombre propio como integrante del equipo de investigación de la institución Clínica Chicamocha, que actúa como centro participante del proyecto **"RESCATE COVID-19: Respuesta de conocimiento para la atenuación de la epidemia por COVID-19"**, coordinado por la **Fundación Cardioinfantil-Instituto de Cardiología**, por medio del presente documento declaro libre y espontáneamente que:

- a) Seré receptor de información relacionada con el proyecto de referencia y que, en virtud de lo anterior, conoceré y dispondré de información que se considera absolutamente confidencial.
- b) Que, con el fin de proteger la referida información, me comprometo y acepto las disposiciones del presente documento, el cual se regulará por las normas legales aplicables y en especial por las siguientes disposiciones:

1. Me comprometo a cumplir con lo establecido en la Ley Estatutaria 1582 de 2012, en la cual se dictan disposiciones generales para la protección de datos personales, así como lo estipulado en la normativa que regula la Ley. De igual manera me comprometo a cumplir con las buenas prácticas clínicas en investigación y los demás lineamientos nacionales e institucionales que enmarcan el desarrollo del proyecto RESCATE COVID-19

2. Toda la información que sea suministrada o a la que se tenga acceso en relación con las actividades del proyecto, es información reservada y confidencial (en adelante la "información confidencial") y podrá ser usada únicamente para los fines de desarrollar y cumplir el objeto del proyecto.

3. Acepto que estoy obligado a no revelar, divulgar, exhibir, mostrar o comunicar a terceros la información confidencial, sin autorización previa, expresa y escrita de los centros participantes en la ejecución del proyecto.

4. Me comprometo a proteger y mantener en secreto la información confidencial, usándola con el mismo grado de precaución y seguridad que utilizo para proteger mi propia información confidencial de similar importancia y en ningún caso recibirá menor cuidado que el razonable.

5. Permitiré el uso de la información confidencial solo por aquellas personas que han sido autorizadas previa y expresamente y que tengan necesidad de conocerla única y exclusivamente para los propósitos del proyecto y los inmersos en el Acuerdo de Cooperación firmado entre las instituciones para el proyecto "RESCATE COVID-19: Respuesta de conocimiento para la atenuación de la epidemia por COVID-19"


6. Me abstendré de revelar la información confidencial en forma directa o indirecta a terceros diferentes del personal previa y expresamente autorizado para recibir la información confidencial de que trata el presente documento y los inmersos en el Acuerdo de Cooperación firmado para el proyecto "RESCATE COVID-19: Respuesta de conocimiento para la atenuación de la epidemia por COVID-19".

7. Me abstendré de copiar, reproducir o duplicar por cualquier medio, total o parcialmente, la información confidencial, o elaborar o divulgar resúmenes o extractos, sin la previa autorización expresa y por escrito del equipo de administración de datos del proyecto.

8. Como receptor de la información confidencial del proyecto, acepto que estoy obligado a consultar a la Fundación Cardioinfantil-Instituto de Cardiología y a obtener su consentimiento escrito, antes de emitir cualquier declaración pública, noticia o información de prensa con respecto del proyecto, salvo cuando sea necesario de acuerdo con las leyes aplicables o con órdenes de autoridades competentes.

9. Me comprometo a que el uso de usuario y contraseña asignadas, por el equipo de datos del proyecto, para el ingreso de datos a las plataformas de registro será personal e intransferible.

No aplicará a aquella porción de la información confidencial que se encuentre en los siguientes supuestos:

	<b>FORMATO DE DECLARACIÓN DE CONFIDENCIALIDAD</b>	Código: Por definir
		Versión: 1
	<b>PROCESO: Investigaciones</b>	Vigencia: 3 años
		Fecha de Vencimiento: 08/2023

- a) Información que sea o pase a ser de dominio público sin que haya mediado incumplimiento de los receptores.
- b) Información que se encontraba ya en poder del receptor.
- c) Información que deba ser suministrada a las autoridades administrativas o judiciales, en cumplimiento de un deber u obligación legal o en cumplimiento de una orden judicial o administrativa emanada de una autoridad competente, en cuyo caso los receptores de la información comunicarán oportunamente a los propietarios la existencia del deber legal o de la orden judicial o administrativa de revelar o entregar la información, con el fin de que los propietarios puedan obtener cualquier orden o medida para proteger su información o, en cualquier otra forma, puedan obtener una protección de sus intereses en tal información. En todo caso tiene que hacerse esta gestión antes que se entregue la información a la autoridad competente, siempre y cuando sea posible.

Las obligaciones de confidencialidad previstas en la presente declaración, estarán vigentes a partir de la fecha de su suscripción y cinco (5) años más.

La violación o inobservancia a cualquiera de las disposiciones de que trata el presente documento o el uso indebido de la información constituirán causal de incumplimiento de la presente declaración y dará lugar a la imposición de las sanciones establecidas en la Ley, sin perjuicio de las acciones administrativas, civiles y penales según corresponda que pueda iniciar las partes o terceros afectados por esta situación.



Firma:  
Nombre: Custodio Alberto Ruiz Bedoya  
Cédula: 1001282792  
Rol en el equipo de investigación: Investigador Principal  
Institución: Clínica Chicamocha  
Fecha (dd/mm/aaaa): 01/09/2020



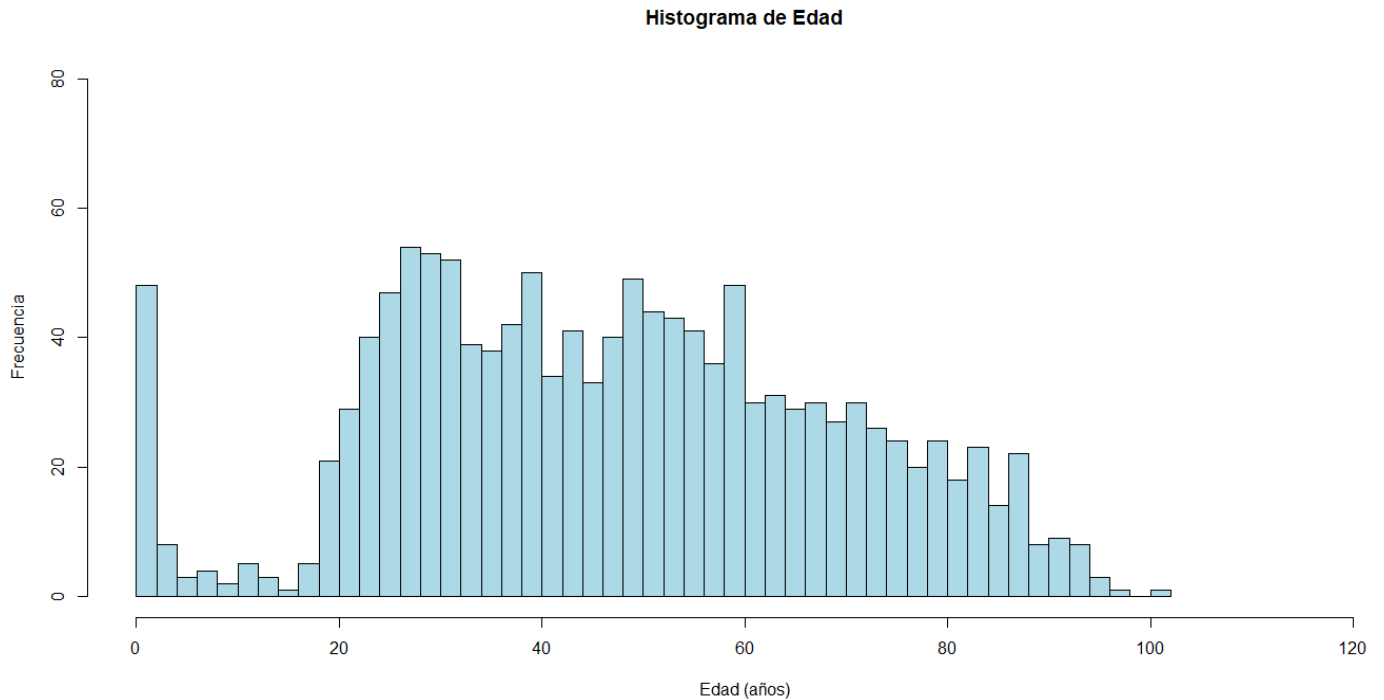


**Anexo 7: Caracterización de las variables continuas incluidas en los modelos:**

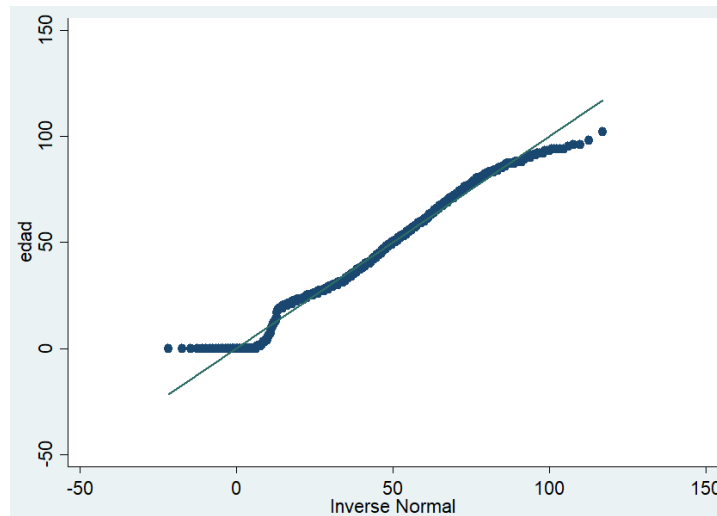
**1.1. Modelo Zhang-China:**

**1.1.1. Edad:**

- Histograma:



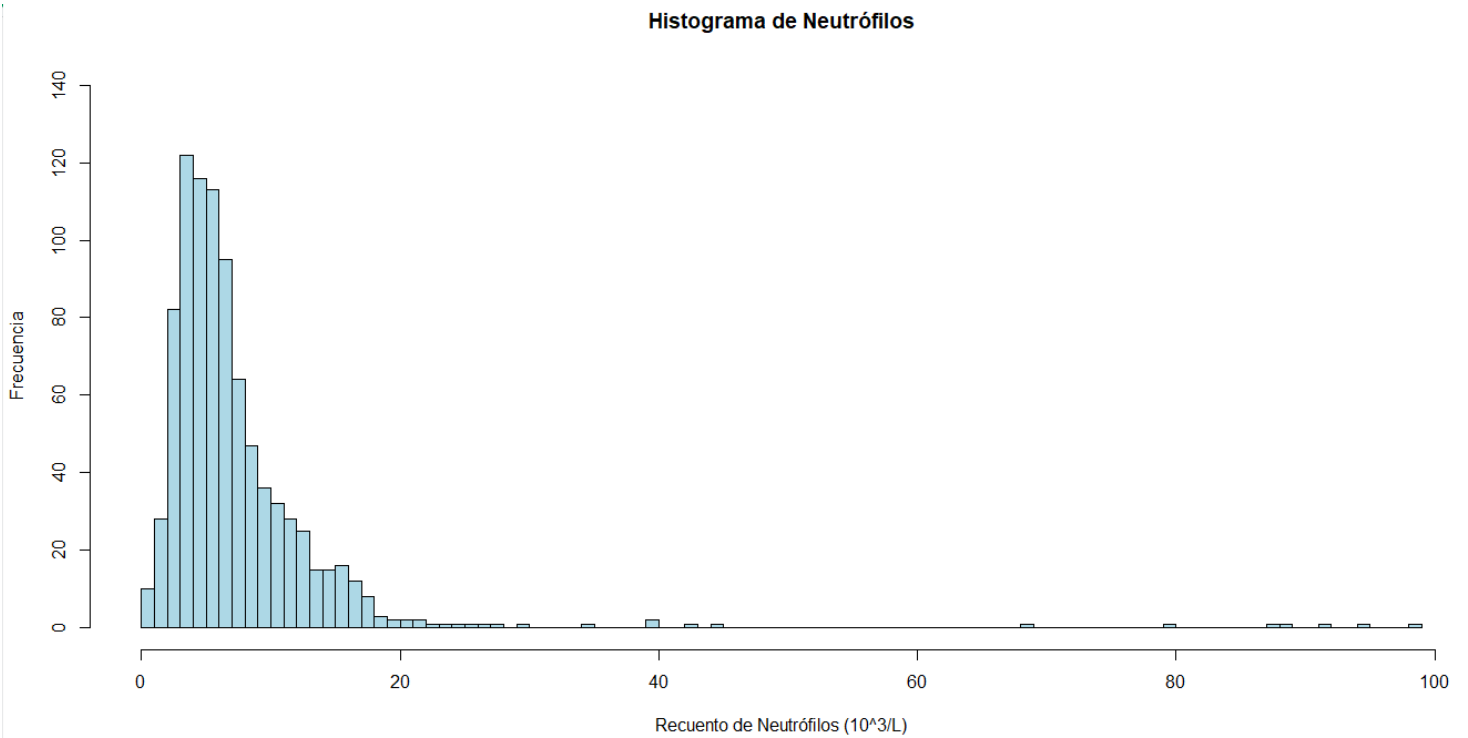
- Gráfico de normalidad:



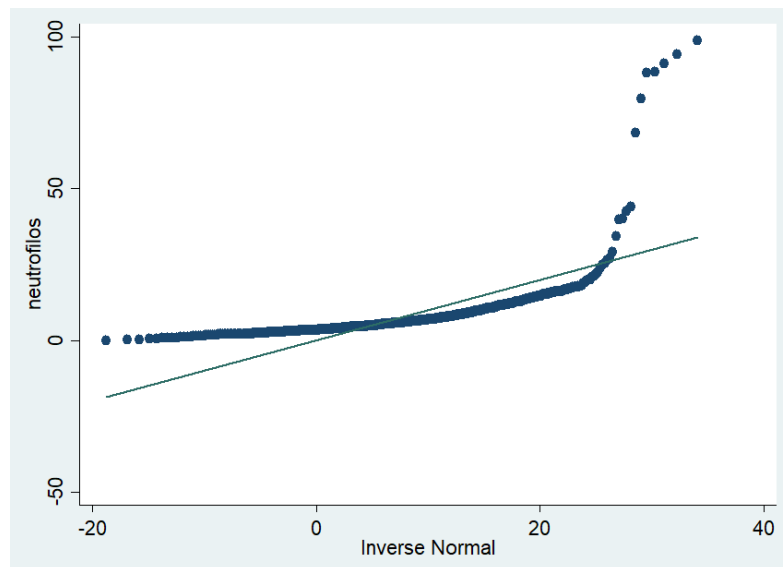
- Prueba de normalidad de Shapiro Wilk:  $p < 0.01$ . Los datos no se ajustan a una distribución normal.  $H_0$  = Los datos no son diferentes de una distribución normal.

### 1.1.2. Rec. Neutrófilos ( $10^3/L$ ):

- Histograma:



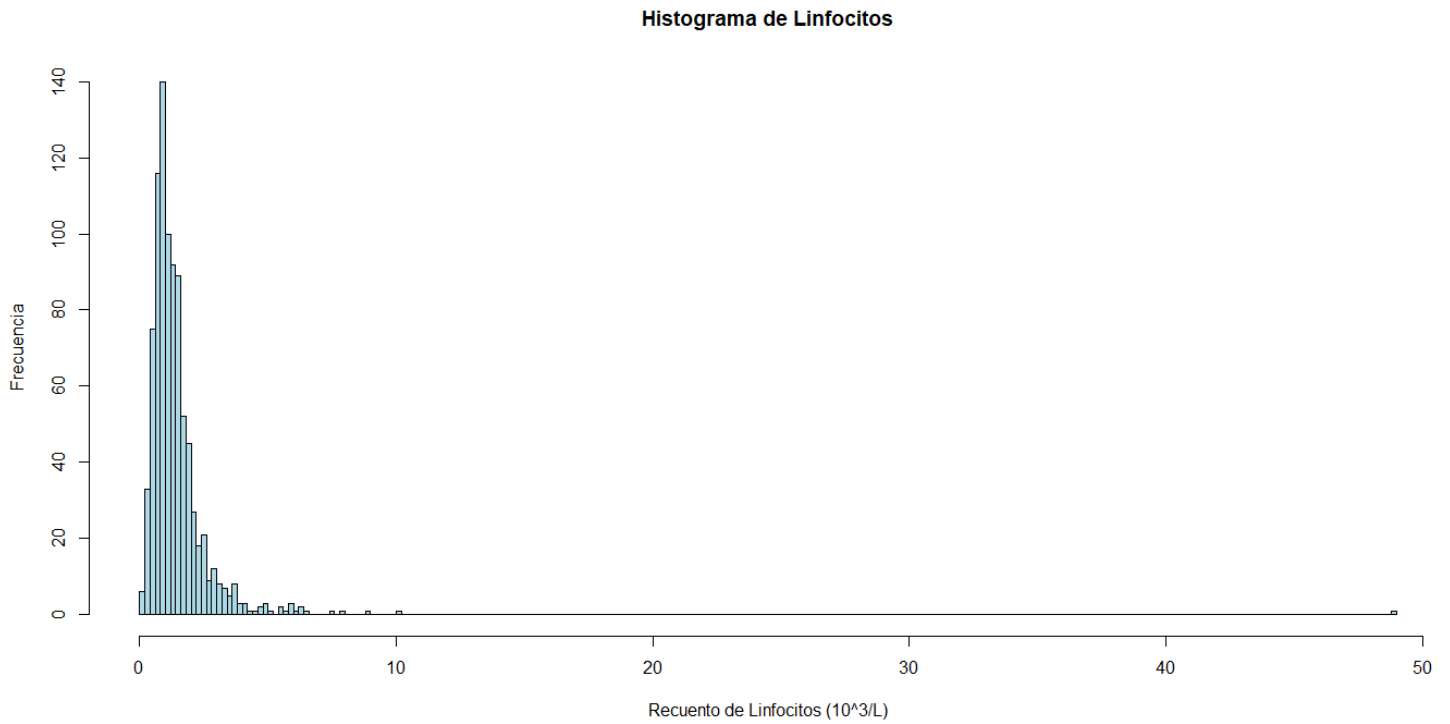
- Gráfico de normalidad:



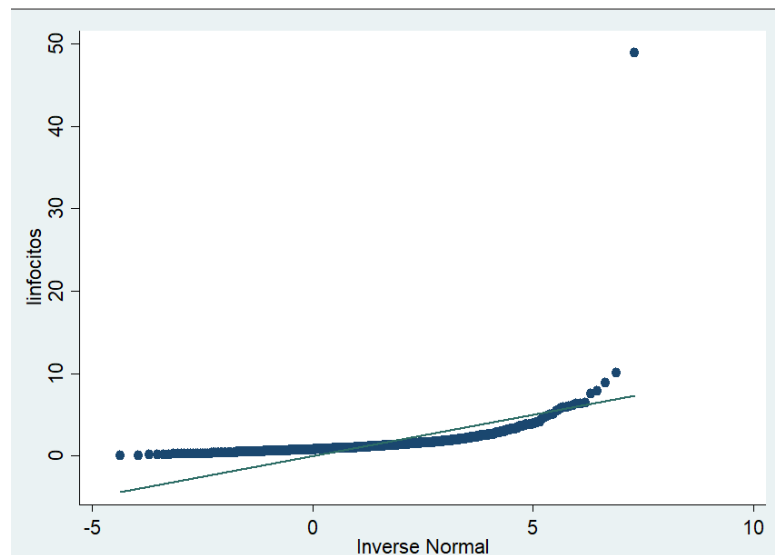
- Prueba de normalidad de Shapiro Wilk:  $p < 0.01$ . Los datos no se ajustan a una distribución normal.  $H_0$  = Los datos no son diferentes de una distribución normal

### 1.1.3. Rec. Linfocitos ( $10^3/L$ ):

- Histograma:



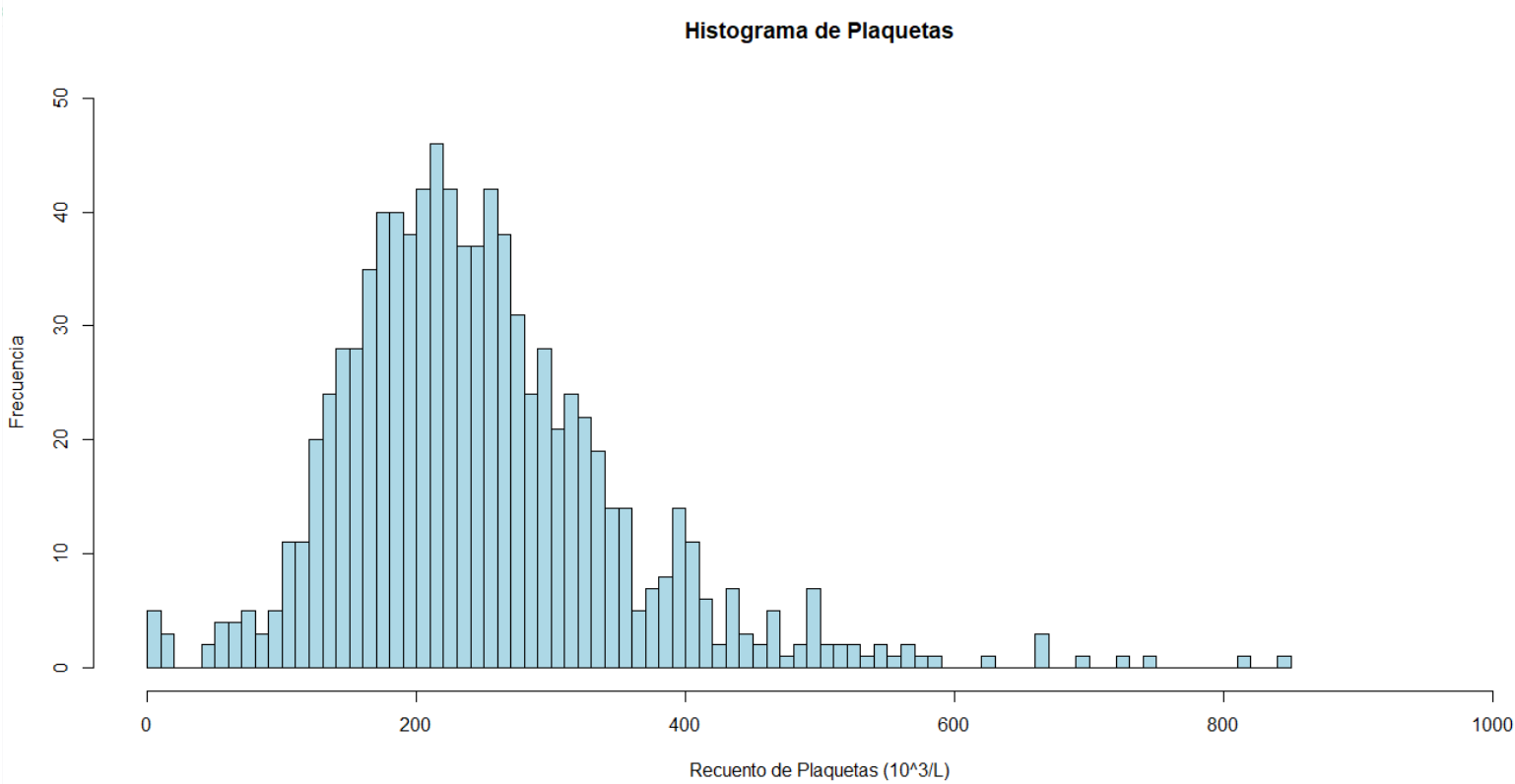
- Gráfico de normalidad:



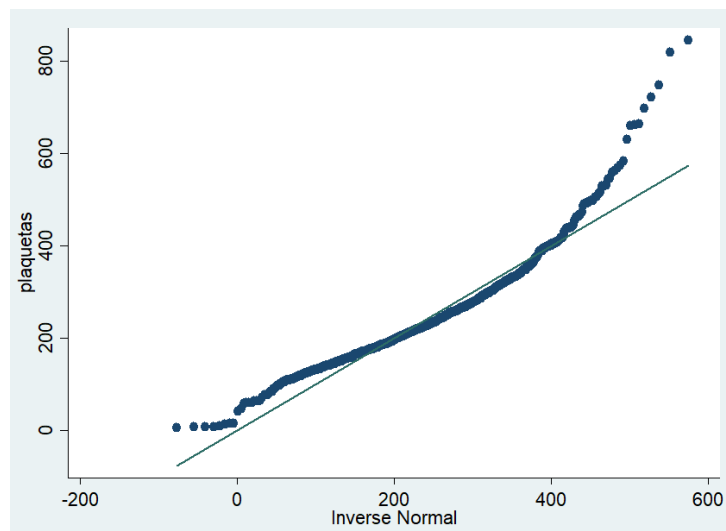
- Prueba de normalidad de Shapiro Wilk:  $p < 0.01$ . Los datos no se ajustan a una distribución normal.  $H_0$  = Los datos no son diferentes de una distribución normal.

### 1.1.4. Rec. Plaquetas ( $10^3/L$ ):

- Histograma:



- Gráfico de normalidad:

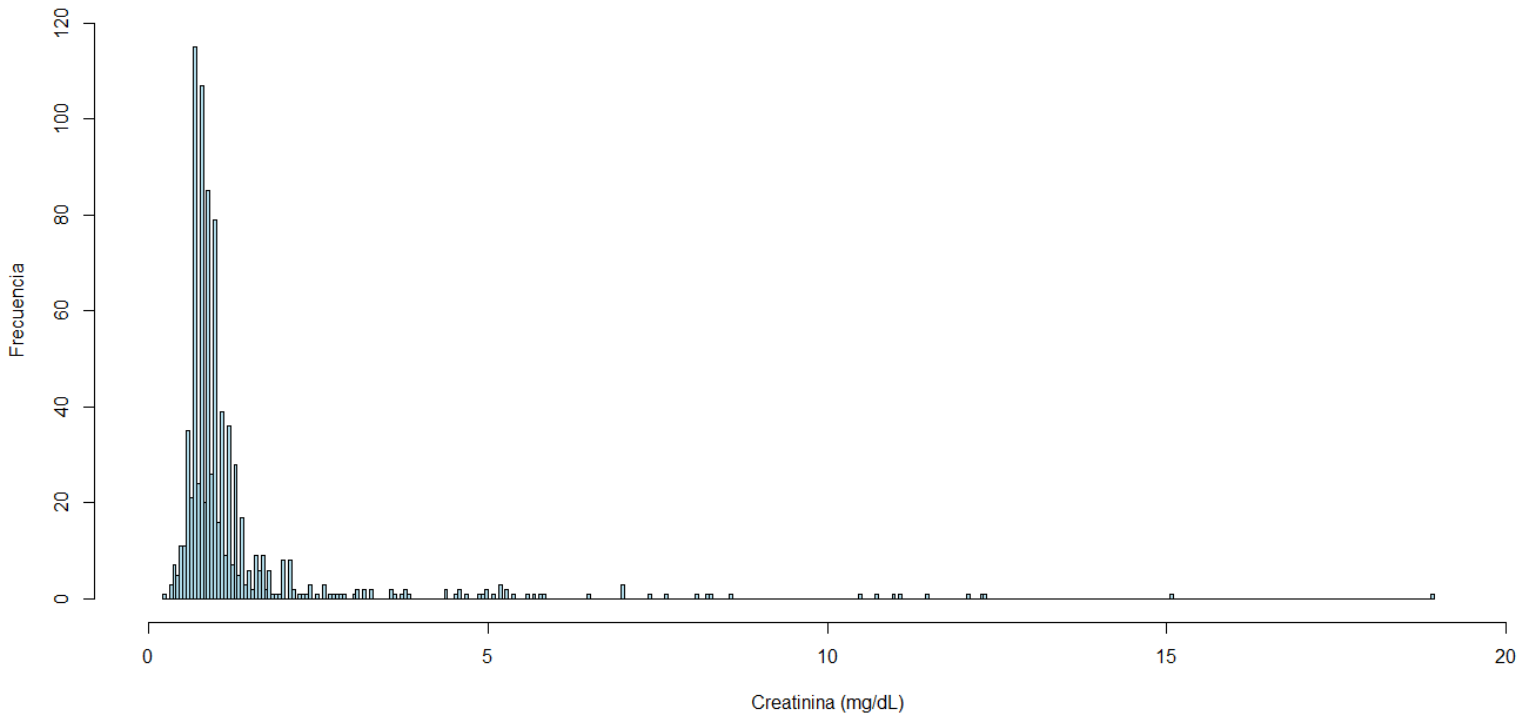


- Prueba de normalidad de Shapiro Wilk:  $p < 0.01$ . Los datos no se ajustan a una distribución normal.  $H_0$  = Los datos no son diferentes de una distribución normal.

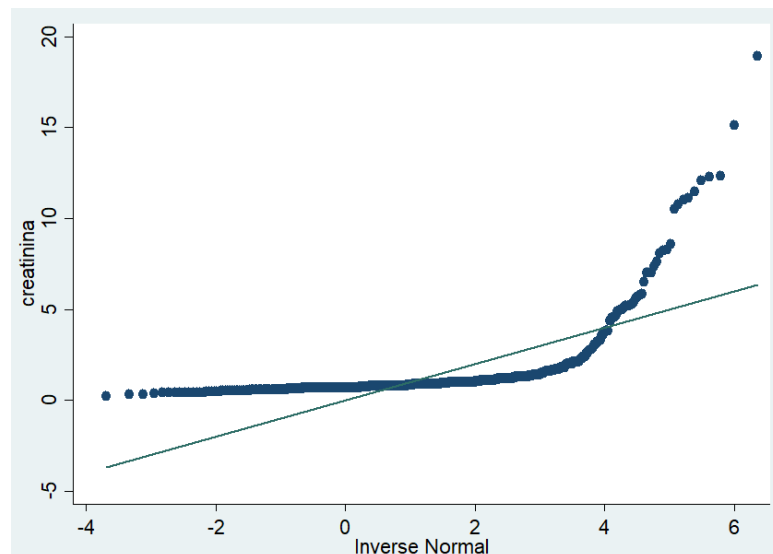
### 1.1.5. Creatinina (mg/dL):

- Histograma:

Histograma de Creatinina



- Gráfico de normalidad:

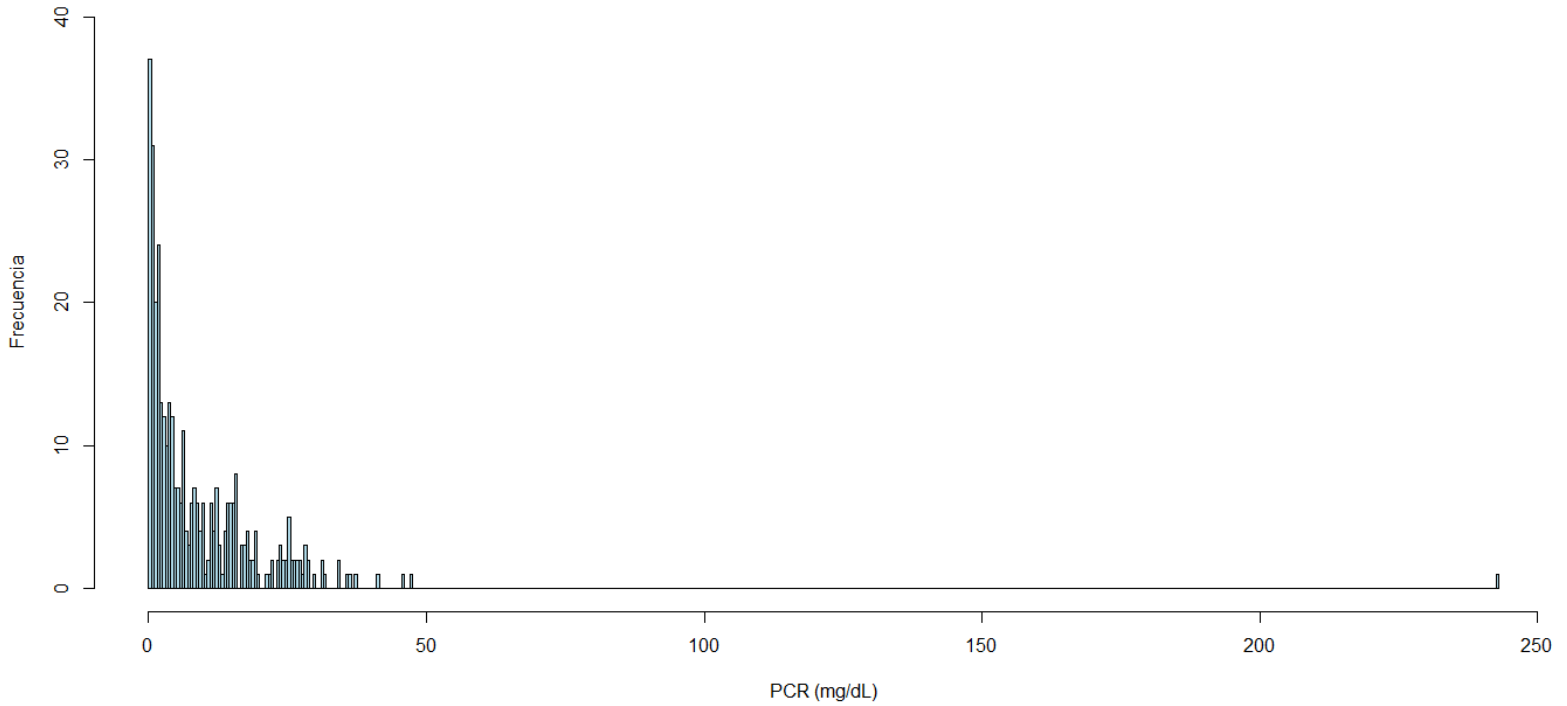


- Prueba de normalidad de Shapiro Wilk:  $p < 0.01$ . Los datos no se ajustan a una distribución normal.  $H_0$  = Los datos no son diferentes de una distribución normal.

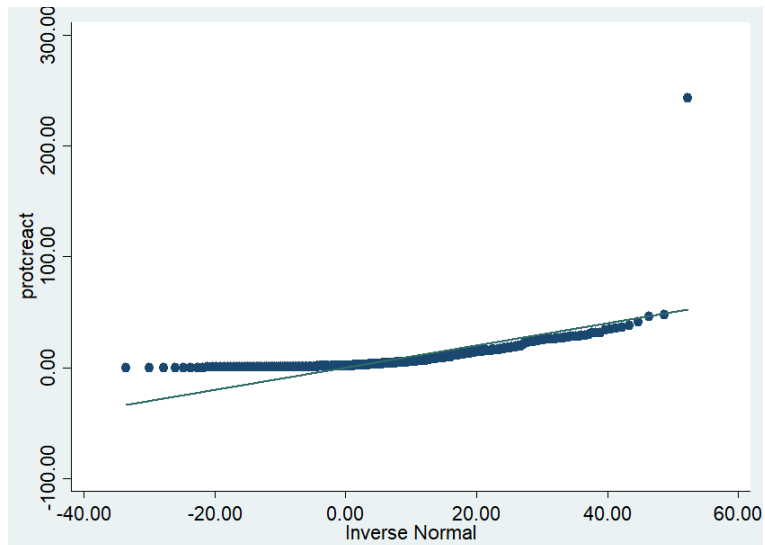
### 1.1.6. PCR (mg/dl):

- Histograma:

Histograma de PCR



- Gráfico de normalidad:

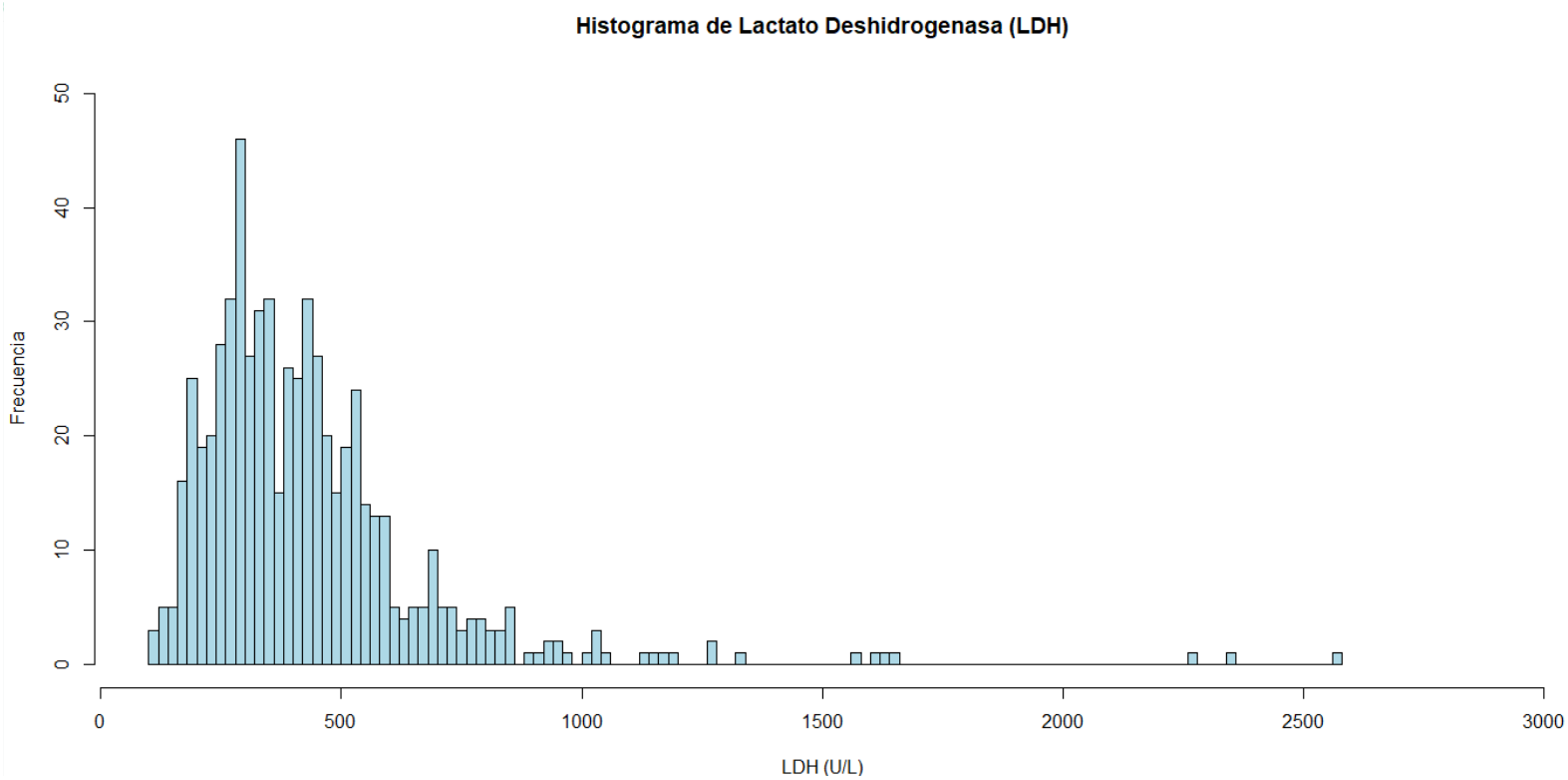


- Prueba de normalidad de Shapiro Wilk:  $p < 0.01$ . Los datos no se ajustan a una distribución normal.  $H_0$  = Los datos no son diferentes de una distribución normal

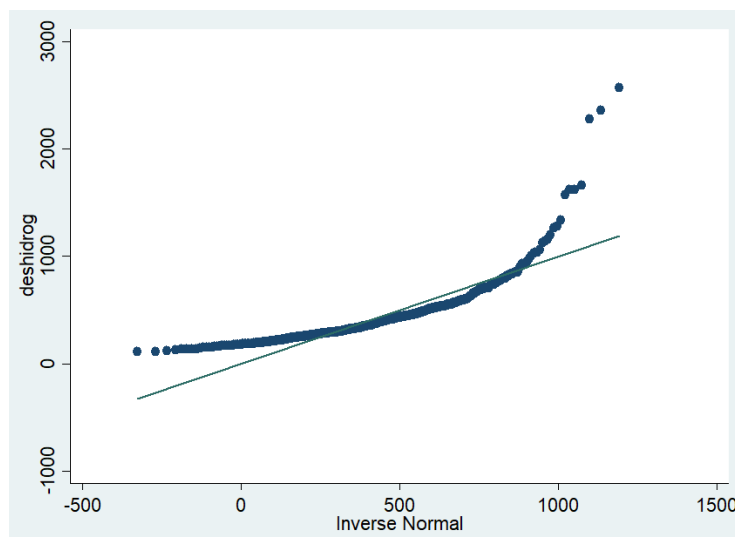
## 1.2. Modelo Xie-China:

### 1.2.1. Deshidrogenasa láctica:

- Histograma:



- Gráfico de normalidad:

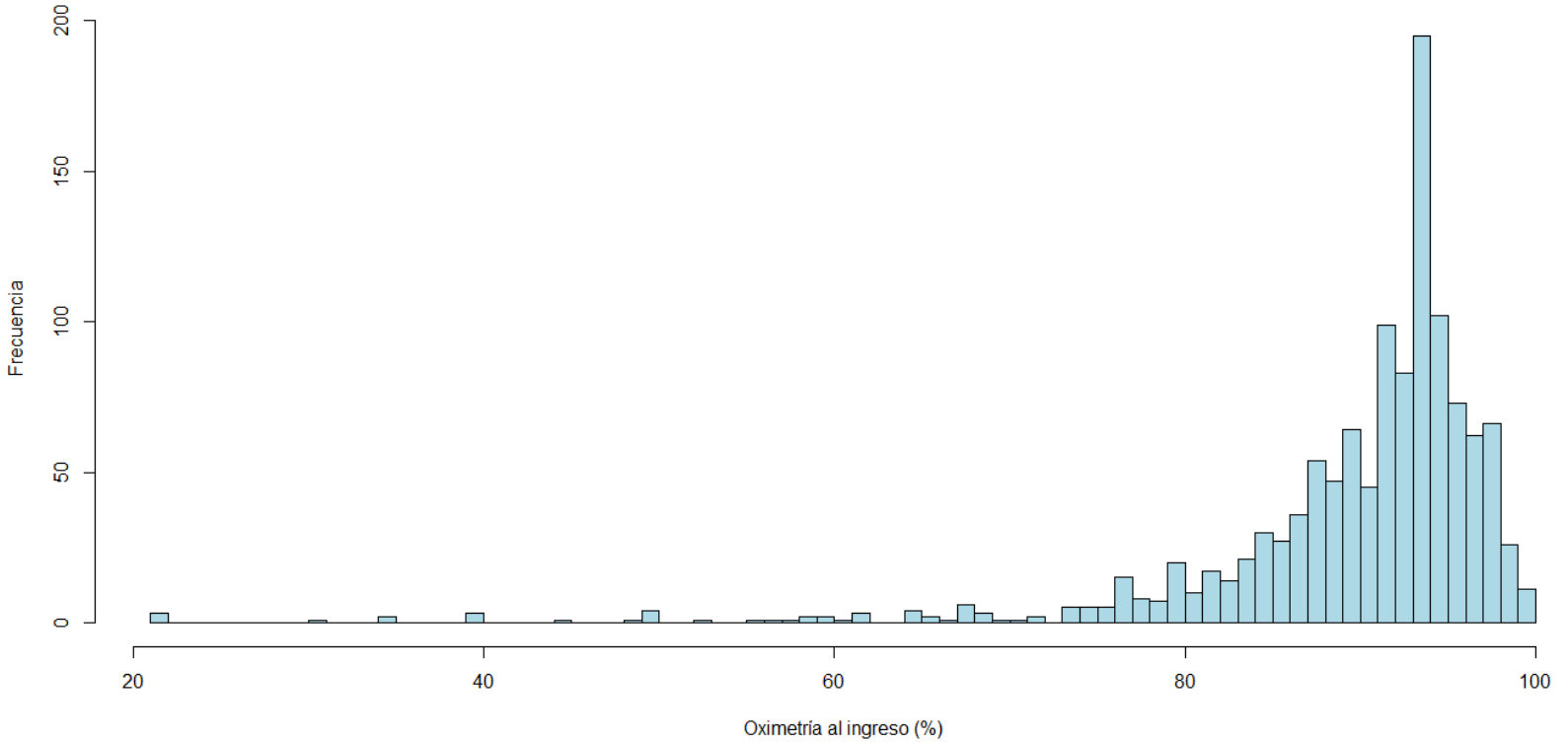


- Prueba de normalidad de Shapiro Wilk:  $p < 0.01$ . Los datos no se ajustan a una distribución normal.  $H_0$  = Los datos no son diferentes de una distribución normal.

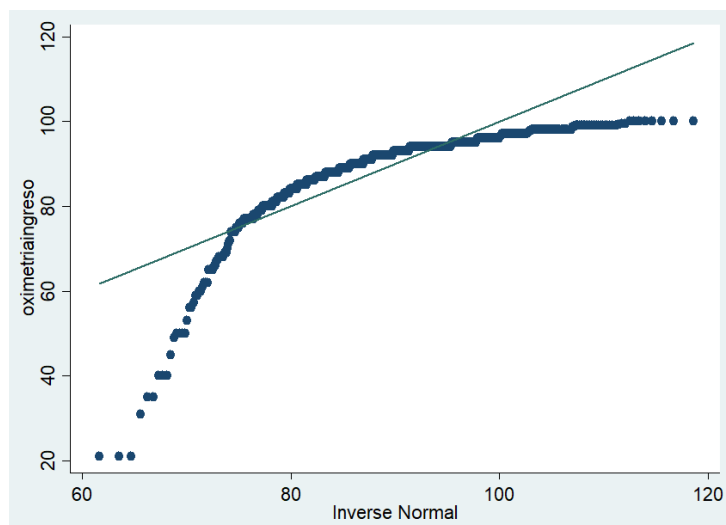
### 1.2.2. Saturación de oxígeno (%):

- Histograma:

Histograma de Oximetría



- Gráfico de normalidad:



- Prueba de normalidad de Shapiro Wilk:  $p < 0.01$ . Los datos no se ajustan a una distribución normal.  $H_0$  = Los datos no son diferentes de una distribución normal.



# RESCATE

## COVID-19



Boletín 9 - v1. marzo 9-2021

## REGISTRO DE CASOS PROBABLES O CONFIRMADOS DE COVID-19 BOLETÍN INFORMATIVO

### Actividades

**6 centros participantes**

#### Registro hospitalario

**2701** registros, **15231** formularios  
**2533** adultos, **168** niños  
**730** trabajadores de la salud

#### Seguimiento telefónico

**2419** llamadas, **1233** efectivas  
**1059** aceptan participación  
**1112** seguimientos

### Caracterización clínica

	1405 positivos		1296 negativos	
	Hospitalización (+) n=758	Hospitalización (-) n=647	Hospitalización (+) n=537	Hospitalización (-) n=759
Sexo hombre n (%)	453 (59.9)	277 (42.8)	294 (54.7)	257 (33.9)
Edad mediana (RIQ)	59 (45-73)	33 (25-46)	70 (49-82)	31 (25-41)
Antecedentes n (%)				
Hipertensión arterial	344 (45.4)	67 (10.4)	259 (48.2)	48 (8.3)
Diabetes	168 (22.2)	26 (4.0)	116 (21.6)	21 (2.8)
Uso de IECA/ARA II	133 (17.5)	38 (5.9)	89 (16.6)	25 (3.3)
Biomarcadores mediana (RIQ)				
Linfocitos 10 <sup>3</sup> cel/u	1.09 (0.74-1.57)	1.41 (1.02-1.97)	1.25 (0.81-1.97)	1.94 (1.46-2.70)
Ferritina ng/mL	819 (395-1564)	488 (250-971)	285 (93-674)	220 (97-686)
LDH U/L	402 (292-528)	287 (225-440)	261 (191-403)	203 (169-282)
Dímero D mg/L	0.97 (0.53-1.96)	0.61 (0.34-1.07)	1.63 (0.66-4.04)	0.43 (0.27-0.74)
Hallazgos en radiografía* n (%) e índice de oxigenación mediana (RIQ)				
Vidrio esmerilado	185 (24.8)	32 (5.8)	34 (6.8)	9 (1.3)
Infiltrado intersticial	143 (19.1)	21 (3.8)	61 (12.2)	14 (2.0)
PaO <sub>2</sub> /FiO <sub>2</sub>	226 (158-287)	298 (252-394)	261 (184-325)	317 (261-340)

Los casos positivos corresponden a una muestra consecutiva, los negativos a una muestra aleatoria representativa. A partir de agosto se considera la admisión a la UCI como criterio de estratificación. \*El 55% de los casos positivos y el 40% de los negativos tenían una radiografía disponible (en ambos grupos en >70% eran casos hospitalizados)

## Desenlaces clínicos

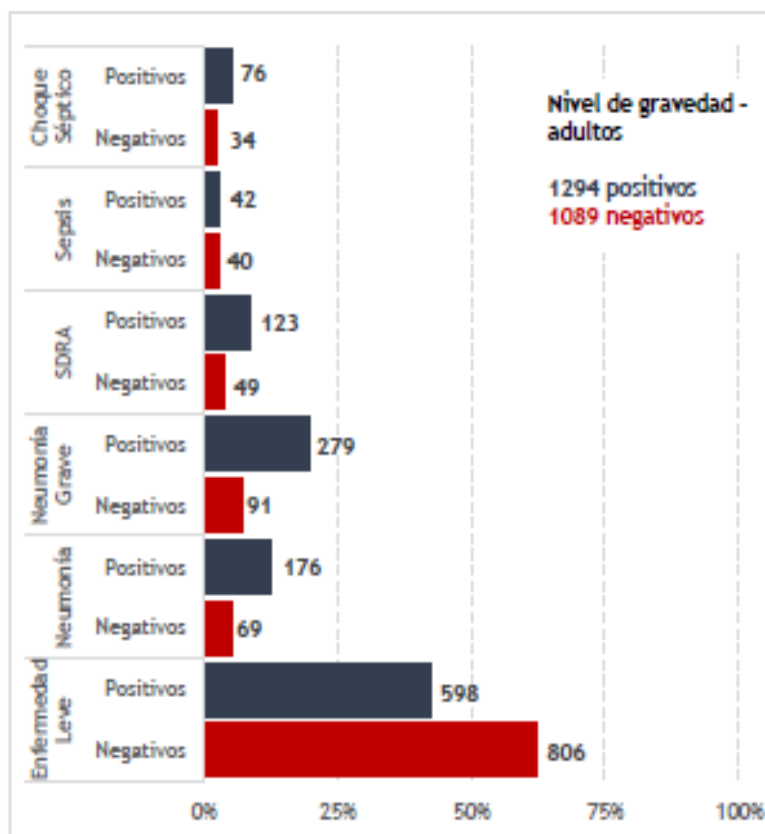
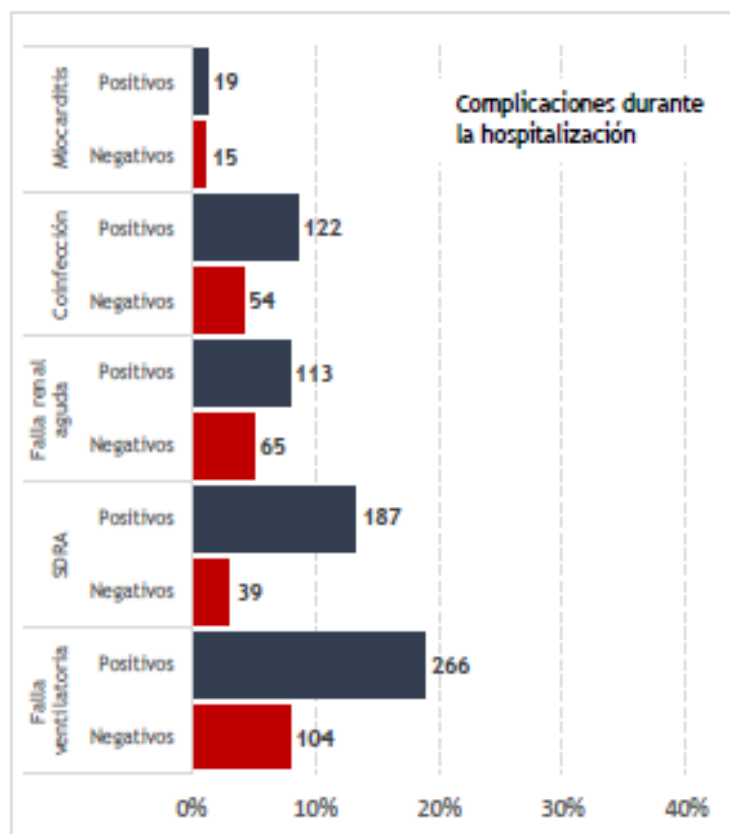
	1405 positivos	1296 negativos
Hospitalización n (%)	758 (54.0)	537 (41.4)
Días estancia*	8 (4-14)	7 (4-13)
Admisión a UCI n (%)	280 (36.9)	121 (22.5)
Días estancia*	9 (4-16)	5 (2-11)
VM*** n (%)	191 (25.2)	70 (13.0)
Días VM*	10 (5-16)	8 (2-16)

\*mediana (RIQ) \*\*No incluye VM \*\*\*VM= ventilación mecánica

	1405 positivos	1296 negativos
Mortalidad* n (%)	184 (13.2)	100 (7.8)
Falla multiorgánica	116 (8.3)	44 (3.4)
Insuficiencia Respiratoria Refractaria	105 (7.5)	40 (3.1)
Choque séptico	56 (4.0)	26 (2.0)
Otros**	25 (1.8)	50 (3.9)
Reingreso n (%)	90 (6.4)	110 (8.4)

\*Las causas de mortalidad no son mutuamente excluyentes.

\*\*ACV, TEP, choque cardiogénico, disintimiento de maniobras.



**Nota:** Este boletín proporciona la información para los centros e investigadores activos en el registro RESCATE COVID-19. Los datos presentados corresponden al periodo entre abril de 2020 y el 8 de marzo de 2021. Estos datos están sujetos a modificaciones.

### Contacto:

Centro de Investigaciones  
Fundación Cardíofantil-Instituto de Cardiología  
[fvillar@cardiofantil.org](mailto:fvillar@cardiofantil.org)

### Centro coordinador:



### Centros participantes:



Anexo 9:...nálisis de desempeño del modelo Zhang-China al retirar la variable PCR y puntos de corte para la toma de decisiones clínicas.

**Tabla 1A.** Asociación entre predictores del Modelo **Zhang-China** (sin proteína C reactiva) y mortalidad intrahospitalaria en una cohorte de validación externa

Predictor	Descripción			Análisis Bivariado (Regresión Logística Simple)			Análisis Multivariado ¶ (Regresión Logística Múltiple)			
	General n= 1342	Mortalidad		Probabilidad en el Promedio (%)	OR crudo (IC 95%)	p	B	EE	OR ajustado (IC 95%)	p
		No n = 1145	Si n = 171							
<b>Edad (años)</b>	47 (31 - 63)	43 (30 - 59)	72 (60 - 83)	7.3	1.06 (1.05-1.07)	<0.01*	0.0461075	0.0060654	1.05 (1.03 - 1.06)	<0.01*
<b>Sexo † (% hombres)</b>	694 (51.6)	587 (50.1)	106 (61.9)	10.0 ‡	1.62 (1.17 - 2.25)	<0.01*	0.0361441	0.2006138	1.04 (0.70 - 1.53)	0.82
<b>Rec. Neutrófilos (10<sup>3</sup>/L)</b>	5.7 (3.8 - 8.8)	5.5 (3.6 - 7.9)	8.1 (4.9 - 12.5)	17.8	1.02 (1.00 - 1.04)	0.01*	0.0253445	0.0101686	1.02 (1.00 - 1.04)	0.01*
<b>Rec. Linfocitos (10<sup>3</sup>/L)</b>	1.1 (0.8 - 1.7)	1.2 (0.85 - 1.73)	0.86 (0.58 - 1.31)	16.2	0.56 (0.42-0.73)	<0.01*	-0.3689177	0.1437314	0.69 (0.52 - 0.92)	0.01*
<b>Rec. Plaquetas (10<sup>3</sup>/L)</b>	232 (179 - 297)	236 (186 - 300)	209 (156 - 282)	17.5	0.99 (0.995-0.999)	<0.01*	-0.001591	0.0010176	0.99 (0.99 - 1.00)	0.13
<b>Creatinina (µmol/L)</b>	0.9 (0.7 - 1.2)	0.9 (0.7 - 1.1)	1.2 (0.87 - 1.8)	19.0	1.19 (1.09 - 1.31)	<0.01*	0.1371349	0.0494204	1.13 (1.02 - 1.25)	0.01*
<b>Bondad de ajuste de Hosmer-Lemeshow: p = 0.1741</b>										

Los valores de laboratorio corresponden a los medidos al ingreso a urgencias.

Para todas las variables continuas se reporta mediana y el rango intercuartílico.

B: Coeficiente (B1), EE: Error Estándar, Prob: Probabilidad.

\*Asociación con significancia estadística p<0.05

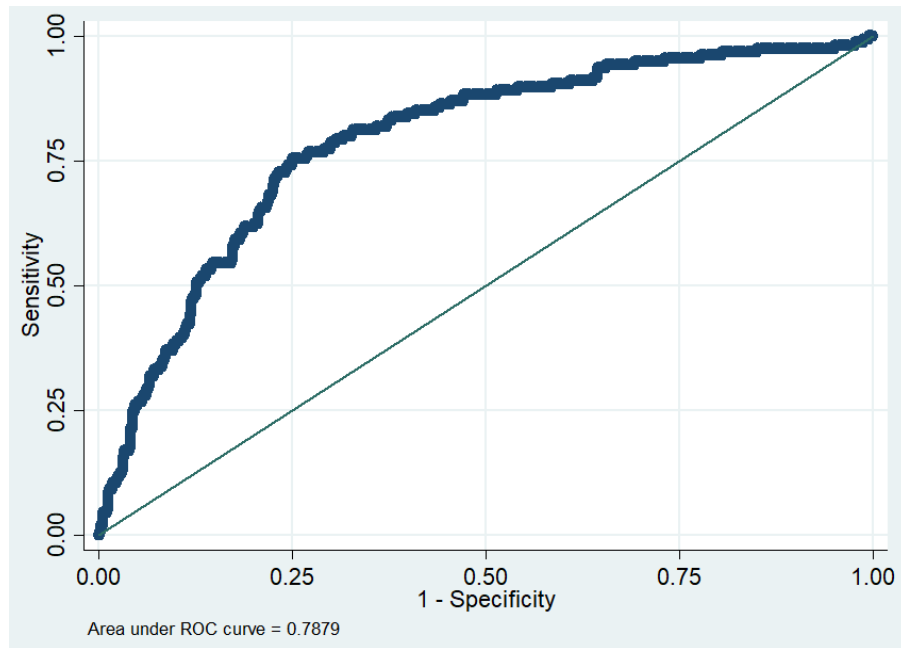
† Para sexo se reporta conteo y proporción en porcentaje (%)

‡ Probabilidad en mujeres

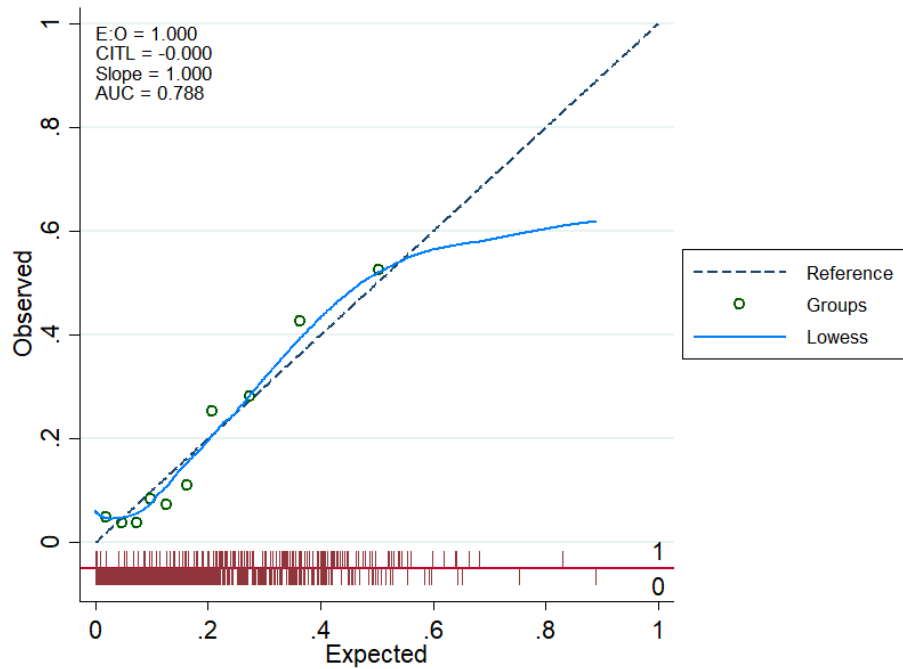
¶ Número de observaciones en el modelo: 823

**Bondad de ajuste de Hosmer-Lemeshow: p = 0.1741**

**Figura 1A.** Curva ROC. Capacidad de discriminación de la escala Zhang-China (sin proteína C reactiva) para predicción de mortalidad intrahospitalaria en la cohorte de validación externa.



**Gráfico 1A.** Gráfico de calibración modelo Zang-China (Sin PCR):



**Tabla 2A.** Medidas de desempeño de cada uno de los puntos de corte sobre las probabilidades predichos por el modelo **Zhang-China** sobre su capacidad de predecir mortalidad intrahospitalaria.

Puntos de corte de Probabilidad Obtenido con Herramienta en línea	Sensibilidad (%)	Especificidad (%)	Clasificación correcta (%)	LR+	LR-	VPP (%)	VPN (%)
0	100.00	0.00	13.11	1.0000	-	12,74%	
1	90.00	18.49	27.87	1.1042	0.5408	13,88%	92,68%
2	87.50	39.25	45.57	1.4402	0.3185	17,38%	95,56%
3	85.00	53.58	57.70	1.8313	0.2799	21,09%	96,07%
4	82.50	62.64	65.25	2.2083	0.2794	24,38%	96,08%
5	77.50	68.68	69.84	2.4744	0.3276	26,54%	95,44%
6	70.00	70.57	70.49	2.3782	0.4251	25,78%	94,16%
7	62.50	72.08	70.82	2.2382	0.5203	24,63%	92,94%
8	60.00	75.47	73.44	2.4462	0.5300	26,31%	92,82%
9	55.00	76.60	73.77	2.3508	0.5874	25,55%	92,10%
10	55.00	78.49	75.41	2.5570	0.5733	27,18%	92,28%
11	52.50	79.25	75.74	2.5295	0.5994	26,98%	91,95%
12	50.00	79.62	75.74	2.4537	0.6280	26,37%	91,60%
13	47.50	80.75	76.39	2.4681	0.6501	26,48%	91,33%
14	47.50	81.13	76.72	2.5175	0.6471	26,87%	91,37%
15	47.50%	82.64%	78.03%	2.7364	0.6353	28,55%	91,51%
16	47.50%	83.77%	79.02%	2.9273	0.6267	29,94%	91,62%
17	45.00%	84.53%	79.34%	2.9085	0.6507	29,81%	91,32%
18	42.50%	84.91%	79.34%	2.8156	0.6772	29,14%	91,00%
20	42.50%	86.04%	80.33%	3.0439	0.6683	30,77%	91,11%
22	42.50%	86.79%	80.98%	3.2179	0.6625	31,96%	91,18%
23	37.50%	87.17%	80.66%	2.9228	0.7170	29,91%	90,52%
26	37.50%	87.92%	81.31%	3.1055	0.7108	31,19%	90,60%
28	35.00%	88.68%	81.64%	3.0917	0.7330	31,10%	90,33%
29	35.00%	89.43%	82.30%	3.3125	0.7268	32,59%	90,41%
30	35.00%	89.81%	82.62%	3.4352	0.7237	33,40%	90,44%
33	32.50%	90.57%	82.95%	3.4450	0.7453	33,47%	90,19%
37	30.00%	90.94%	82.95%	3.3125	0.7697	32,59%	89,90%
39	30.00%	91.32%	83.28%	3.4565	0.7665	33,54%	89,93%
40	30.00%	91.70%	83.61%	3.6136	0.7634	34,54%	89,97%
41	27.50%	91.70%	83.28%	3.3125	0.7906	32,60%	89,65%
42	27.50%	92.08%	83.61%	3.4702	0.7874	33,64%	89,69%
43	27.50%	92.45%	83.93%	3.6437	0.7842	34,72%	89,73%
45	27.50%	92.83%	84.26%	3.8355	0.7810	35,90%	89,76%
46	27.50%	93.21%	84.59%	4.0486	0.7778	37,16%	89,80%
48	27.50%	93.96%	85.25%	4.5547	0.7716	39,93%	89,88%
49	25.00%	93.96%	84.92%	4.1406	0.7982	37,67%	89,56%
50	25.00%	94.34%	85.25%	4.4167	0.7950	39,21%	89,60%
53	22.50%	94.72%	85.25%	4.2589	0.8182	38,35%	89,33%
54	22.50%	95.47%	85.90%	4.9687	0.8118	42,03%	89,40%
55	22.50%	95.85%	86.23%	5.4204	0.8086	44,18%	89,44%
58	22.50%	96.23%	86.56%	5.9625	0.8054	46,56%	89,48%
59	22.50%	96.60%	86.89%	6.6250	0.8022	49,14%	89,51%
60	22.50%	96.98%	87.21%	7.4531	0.7991	52,10%	89,55%
62	22.50%	97.36%	87.54%	8.5179	0.7960	55,44%	89,59%
66	22.50%	97.74%	87.87%	9.9375	0.7930	59,24%	89,62%
67	22.50%	98.11%	88.20%	11.9250	0.7899	63,48%	89,66%
68	20.00%	98.11%	87.87%	10.6000	0.8154	60,71%	89,36%

<b>69</b>	17.50%	98.11%	87.54%	9.2750	0.8409	57,48%	89,07%
<b>76</b>	17.50%	98.49%	87.87%	11.5937	0.8376	62,85%	89,10%
<b>80</b>	17.50%	98.87%	88.20%	15.4583	0.8344	69,34%	89,14%
<b>82</b>	15.00%	98.87%	87.87%	13.2499	0.8597	65,96%	88,85%
<b>83</b>	10.00%	98.87%	87.21%	8.8333	0.9103	56,37%	88,27%
<b>86</b>	10.00%	99.25%	87.54%	13.2500	0.9068	66,06%	88,31%
<b>90</b>	10.00%	99.62%	87.87%	26.5001	0.9034	79,35%	88,35%
<b>92</b>	7.50%	99.62%	87.54%	19.8751	0.9285	74,24%	88,06%
<b>96</b>	7.50%	100.00%	87.87%		0.9250	100,00%	88,10%
<b>98</b>	5.00%	100.00%	87.54%		0.9500	100,00%	87,82%
<b>100</b>	2.50%	100.00%	87.21%		0.9750	100,00%	87,54%

**Anexo 10: Modelos de regresión logística múltiple obtenidos en STATA 14**

1. Modelo de regresión logística múltiple con los predictores de **Zhang-China** al aplicarlos en la cohorte RESCATE COVID-19:

```
. logistic estado edad sex neutrofilos linfocitos plaquetas creatinina protcreat

Logistic regression                               Number of obs   =       317
                                                    LR chi2(7)      =       43.14
                                                    Prob > chi2     =       0.0000
Log likelihood = -104.27236                       Pseudo R2      =       0.1714
```

estadoegresopte	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
edad	1.034325	.01134	3.08	0.002	1.012336	1.056791
sex	.9531114	.3610453	-0.13	0.899	.4536307	2.002557
neutrofilos	.9828169	.0222653	-0.77	0.444	.9401322	1.027439
linfocitos	1.017708	.0881874	0.20	0.839	.858745	1.206097
plaquetas	.9999623	.002044	-0.02	0.985	.9959642	1.003976
creatinina	1.196595	.1139901	1.88	0.060	.9927958	1.442229
protcreat	1.071523	.0187569	3.95	0.000	1.035384	1.108924
_cons	.0081011	.0079027	-4.94	0.000	.0011973	.0548146

2. Modelo de regresión logística múltiple con los predictores de **Zhang-China** al aplicarlos en la cohorte RESCATE COVID-19: Coeficientes:
3. **Prueba de bondad de ajuste de Hosmer.Lemeshow para el modelo de Zang-China:**

```

Logistic regression                               Number of obs   =       317
                                                  LR chi2(7)      =       43.14
                                                  Prob > chi2     =       0.0000
Log likelihood = -104.27236                    Pseudo R2      =       0.1714

```

estadoegresopte	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
edad	1.034325	.01134	3.08	0.002	1.012336 1.056791
sex	.9531114	.3610453	-0.13	0.899	.4536307 2.002557
neutrofilos	.9828169	.0222653	-0.77	0.444	.9401322 1.027439
linfocitos	1.017708	.0881874	0.20	0.839	.858745 1.206097
plaquetas	.9999623	.002044	-0.02	0.985	.9959642 1.003976
creatinina	1.196595	.1139901	1.88	0.060	.9927958 1.442229
protcreact	1.071523	.0187569	3.95	0.000	1.035384 1.108924
_cons	.0081011	.0079027	-4.94	0.000	.0011973 .0548146

```
. estat gof, group (10)
```

Logistic model for estadoegresopte, goodness-of-fit test

(Table collapsed on quantiles of estimated probabilities)

```

number of observations =       317
number of groups      =        10
Hosmer-Lemeshow chi2(8) =       23.55
Prob > chi2           =         0.0027

```

```

LR chi2(7)           =       43.14
Prob > chi2         =       0.0000
Pseudo R2           =       0.1714
Log likelihood = -104.27236

```

estadoegresopte	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
edad	.033749	.0109636	3.08	0.002	.0122607 .0552373
sex	-.0480235	.3788071	-0.13	0.899	-.7904718 .6944247
neutrofilos	-.0173325	.0226546	-0.77	0.444	-.0617347 .0270698
linfocitos	.0175533	.0866529	0.20	0.839	-.1522833 .1873898
plaquetas	-.0000377	.0020441	-0.02	0.985	-.004044 .0039685
creatinina	.1794799	.0952621	1.88	0.060	-.0072303 .3661902
protcreact	.0690812	.0175049	3.95	0.000	.0347723 .1033902
_cons	-4.815749	.9755028	-4.94	0.000	-6.7277 -2.903799



4. **Fórmula de regresión logística (logit)** con los predictores de **Zhang-China** al aplicarlos en la cohorte RESCATE COVID-19.

$$\text{Log}(P/1-P) = 0,0337 + (-0.48)*\text{sexo} + (-0.0173)*\text{neutrófilo} + 0.0175*\text{linfocitos} + (-0.000003)*\text{plaquetas} + 0.179*\text{creatinina} + 0.069*\text{pcr}$$

5. Modelo de regresión logística entre los **puntajes** obtenidos por la escala **Zhang-China** para cada paciente de la cohorte RESCATE COVID-19 y la probabilidad de mortalidad intrahospitalaria:

```
. logistic estado probzang
```

```
Logistic regression                Number of obs   =         305
                                   LR chi2(1)       =         23.78
                                   Prob > chi2       =         0.0000
Log likelihood = -106.6193         Pseudo R2      =         0.1003
```

estadoegresopte	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
probzang	1.032248	.0066393	4.93	0.000	1.019317	1.045343
_cons	.0879025	.0198562	-10.76	0.000	.0564578	.1368607

6. Prueba de bondad de ajuste de Hosmer Lemeshow para el modelo de regresión entre los puntajes de la escala Zhang-China y mortalidad:

```

Logistic regression          Number of obs   =       305
                             LR chi2(1)            =       23.78
                             Prob > chi2           =       0.0000
Log likelihood = -106.6193   Pseudo R2       =       0.1003

```

estadoegresopte	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
probzang	1.032248	.0066393	4.93	0.000	1.019317	1.045343
_cons	.0879025	.0198562	-10.76	0.000	.0564578	.1368607

```
. estat gof, group (10)
```

Logistic model for estadoegresopte, goodness-of-fit test

(Table collapsed on quantiles of estimated probabilities)  
(There are only 9 distinct quantiles because of ties)

```

number of observations =       305
number of groups      =         9
Hosmer-Lemeshow chi2(7) =      19.04
Prob > chi2           =         0.0081

```

7. Modelo de regresión logística entre los **niveles de riesgo** de la escala **Zhang-China** y mortalidad utilizando variables indicadoras:

. logistic estado i.ordinalzang

```

Logistic regression                Number of obs   =       305
                                   LR chi2(3)        =       21.84
                                   Prob > chi2         =       0.0001
Log likelihood = -107.59359        Pseudo R2      =       0.0921

```

estadoegresopte	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
ordinalzang						
1	1.113636	.6572627	0.18	0.855	.3502422	3.540938
2	3.004717	1.818286	1.82	0.069	.9177072	9.837914
3	12.25	8.597985	3.57	0.000	3.095265	48.48131
_cons	.0816327	.0424496	-4.82	0.000	.02946	.2262015

. estat gof, group (10)

8. Modelo de regresión logística múltiple con los predictores de **Xie-China** al aplicarlos en la cohorte RESCATE COVID-1
9. Prueba de bondad de ajuste de Hosmer Lemeshow para el modelo de regresión entre los

. logistic estado edad deshidrog linfocitos oximetriaingreso

```

Logistic regression                Number of obs   =       593
                                   LR chi2(4)        =       121.82
                                   Prob > chi2         =       0.0000
Log likelihood = -222.02489        Pseudo R2      =       0.2153

```

estadoegresopte	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
edad	1.054051	.0085753	6.47	0.000	1.037377	1.070993
deshidrog	1.002822	.0005164	5.47	0.000	1.001811	1.003835
linfocitos	.4976397	.1155615	-3.01	0.003	.3156819	.7844773
oximetriaingreso	.9785074	.0106672	-1.99	0.046	.9578219	.9996397
_cons	.0320493	.0384626	-2.87	0.004	.0030499	.3367841

Note: 1 failure and 0 successes completely determined.

puntajes de la escala Xie-China y mortalidad:

```
. logistic estado edad deshidrog linfocitos oximetriaingreso
```

```
Logistic regression                Number of obs   =       593
                                   LR chi2(4)         =       121.82
                                   Prob > chi2        =       0.0000
Log likelihood = -222.02489         Pseudo R2      =       0.2153
```

estadoegresopte	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
edad	1.054051	.0085753	6.47	0.000	1.037377	1.070993
deshidrog	1.002822	.0005164	5.47	0.000	1.001811	1.003835
linfocitos	.4976397	.1155615	-3.01	0.003	.3156819	.7844773
oximetriaingreso	.9785074	.0106672	-1.99	0.046	.9578219	.9996397
_cons	.0320493	.0384626	-2.87	0.004	.0030499	.3367841

Note: 1 failure and 0 successes completely determined.

```
. estat gof, group (10)
```

Logistic model for estadoegresopte, goodness-of-fit test

(Table collapsed on quantiles of estimated probabilities)

```
number of observations =       593
number of groups      =        10
Hosmer-Lemeshow chi2(8) =       10.54
Prob > chi2           =       0.2294
```

10. Modelo de regresión logística múltiple con los predictores de **Xie-China** al aplicarlos en la cohorte RESCATE COVID-19 (logit: Coeficientes):

```
. logit estado edad deshidrog linfocitos oximetriaingreso
```

```
Iteration 0:  log likelihood = -282.93426
Iteration 1:  log likelihood = -232.4652
Iteration 2:  log likelihood = -223.32171
Iteration 3:  log likelihood = -222.03449
Iteration 4:  log likelihood = -222.02489
Iteration 5:  log likelihood = -222.02489
```

```
Logistic regression                Number of obs   =          593
                                   LR chi2(4)         =          121.82
                                   Prob > chi2        =           0.0000
Log likelihood = -222.02489        Pseudo R2      =           0.2153
```

estadoegresopte	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
edad	.052641	.0081355	6.47	0.000	.0366956	.0685863
deshidrog	.0028183	.0005149	5.47	0.000	.0018091	.0038275
linfocitos	-.6978789	.2322192	-3.01	0.003	-1.15302	-.2427376
oximetriaingreso	-.0217269	.0109015	-1.99	0.046	-.0430934	-.0003604
_cons	-3.440479	1.200107	-2.87	0.004	-5.792645	-1.088313

Note: 1 failure and 0 successes completely determined.

11. **Fórmula de regresión logística (logit)** con los predictores de **Xie-China** al aplicarlos en la cohorte RESCATE COVID-19.

$$\text{Log}(P/1-P) = -3.440479 + 0.052641 * \text{edad} + 0.0028183 * \text{LDH} + (-0.6978789) * \text{linfocitos} + (-0.0217269) * \text{oximetría}$$

12. Modelo de regresión logística múltiple con los predictores de **Bello-México** al aplicarlos en la cohorte RESCATE COVID-19

```
. logistic estado Edad__40 Edad__65 Diabetes diabetes40cor Obesidad Neumonía Renal_Cronico EPOC Inmunosupresio
> n
```

```
Logistic regression                Number of obs    =    1,168
                                   LR chi2(9)         =    146.86
                                   Prob > chi2        =    0.0000
Log likelihood = -319.71055         Pseudo R2      =    0.1868
```

estadoegresopte	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
Edad__40	.8553405	.2653602	-0.50	0.614	.4656536 1.571141
Edad__65	3.050102	.7528647	4.52	0.000	1.880226 4.947875
Diabetes	1.015909	.9532193	0.02	0.987	.1615024 6.390436
diabetes40cor	4.306396	4.141176	1.52	0.129	.6539879 28.35686
Obesidad	2.314901	.6973178	2.79	0.005	1.282699 4.177726
Neumonía	4.196389	1.142142	5.27	0.000	2.461512 7.15401
Renal_Cronico	.9444023	.4582738	-0.12	0.906	.3648457 2.444584
EPOC	3.260622	1.064078	3.62	0.000	1.719957 6.18135
Inmunosupresion	.5281745	.5705402	-0.59	0.555	.0635769 4.387884
_cons	.0286327	.0082712	-12.30	0.000	.0162545 .0504374

13. Modelo de regresión logística múltiple (logit) con los predictores de **Bello-México** al aplicarlos en la cohorte RESCATE COVID-19 (coeficientes)

```
. logit estado Edad__40 Edad__65 Diabetes diabetes40cor Obesidad Neumonía Renal_Cronico EPOC Inmunosupresion
```

```
Iteration 0: log likelihood = -393.13967
Iteration 1: log likelihood = -345.81803
Iteration 2: log likelihood = -321.11858
Iteration 3: log likelihood = -319.71411
Iteration 4: log likelihood = -319.71055
Iteration 5: log likelihood = -319.71055
```

```
Logistic regression                                Number of obs    =    1,168
                                                    LR chi2(9)       =    146.86
                                                    Prob > chi2      =    0.0000
Log likelihood = -319.71055                       Pseudo R2        =    0.1868
```

estadoegresopte	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
Edad__40	-.1562556	.3102392	-0.50	0.614	-.7643134 .4518021
Edad__65	1.115175	.2468326	4.52	0.000	.631392 1.598958
Diabetes	.0157836	.9382922	0.02	0.987	-1.823235 1.854802
diabetes40cor	1.460101	.9616339	1.52	0.129	-.4246664 3.344869
Obesidad	.8393669	.3012301	2.79	0.005	.2489667 1.429767
Neumonía	1.434224	.2721727	5.27	0.000	.9007757 1.967673
Renal_Cronico	-.057203	.4852527	-0.12	0.906	-1.008281 .8938749
EPOC	1.181918	.326342	3.62	0.000	.5422995 1.821537
Inmunosupresion	-.6383286	1.080212	-0.59	0.555	-2.755504 1.478847
_cons	-3.553205	.2888741	-12.30	0.000	-4.119388 -2.987023

14. Cálculo de prueba de bondad de ajuste de Hosmer-Lemeshow para el modelo Bello-México:

```
. estat gof, group (10)
```

Logistic model for estadoegresopte, goodness-of-fit test

(Table collapsed on quantiles of estimated probabilities)

(There are only 8 distinct quantiles because of ties)

```
number of observations =    1168
number of groups       =      8
Hosmer-Lemeshow chi2(6) =    6.95
Prob > chi2            =    0.3254
```

15. Modelo de regresión logística múltiple de los **puntajes** de la escala **Bello-México** respecto a mortalidad intrahospitalaria en la cohorte RESCATE COVID-19:

```
. logistic estado puntajetotalcor
```

```
Logistic regression                Number of obs    =    1,171
                                   LR chi2(1)          =    231.66
                                   Prob > chi2         =    0.0000
Log likelihood = -277.64337        Pseudo R2       =    0.2944
```

estadoegresopte	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
puntajetotalcor	1.381957	.0422764	10.57	0.000	1.301531 1.467351
_cons	.0212671	.0056725	-14.44	0.000	.0126086 .0358714

16. Cálculo de prueba de bondad de ajuste de Hosmer-Lemeshow para el modelo Bello-México con los puntajes de la escala como variable predictora:

```
. estat gof, group (10)
```

```
Logistic model for estadoegresopte, goodness-of-fit test
```

```
(Table collapsed on quantiles of estimated probabilities)
(There are only 6 distinct quantiles because of ties)
```

```
number of observations =    1171
number of groups      =     6
Hosmer-Lemeshow chi2(4) =    13.37
Prob > chi2          =     0.0096
```

```
. estat gof, group (6)
```

```
Logistic model for estadoegresopte, goodness-of-fit test
```

```
(Table collapsed on quantiles of estimated probabilities)
(There are only 5 distinct quantiles because of ties)
```

```
number of observations =    1171
number of groups      =     5
Hosmer-Lemeshow chi2(3) =     3.23
Prob > chi2          =     0.3573
```



No es posible realizar la prueba tradicional de Hosmer-Lemeshow ya que no es posible formar deciles. Se realiza la prueba con el máximo número de grupos disponibles (6) con lo cual se obtiene un resultado de p 0.357

17. Modelo de regresión logística múltiple (variables indicadores) con los niveles de riesgo obtenidos con la escala de **Bello-México** en la cohorte RESCATE COVID-19:

```
. logistic estado i.ordinalmexcor
```

Logistic regression	Number of obs	=	1,171
	LR chi2(4)	=	244.86
	Prob > chi2	=	0.0000
Log likelihood = -271.0416	Pseudo R2	=	0.3112

estadoegresopte	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
ordinalmexcor						
Bajo riesgo	24.36508	19.02295	4.09	0.000	5.274668	112.5487
Riesgo Moderado	51.16666	53.29597	3.78	0.000	6.642977	394.1046
Riesgo Alto	54.26767	39.70268	5.46	0.000	12.93586	227.6603
Muy Alto Riesgo	231.8163	168.0223	7.51	0.000	55.99981	959.6248
_cons	.0032573	.002307	-8.09	0.000	.0008128	.0130536

18. Cálculo de prueba de bondad de ajuste de Hosmer-Lemeshow para la escala Bello-México con los niveles de riesgo como variable predictora.

Log likelihood = -271.0416 Pseudo R2 = 0.3112

estadoegresopte	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
ordinalmexcor						
Bajo riesgo	24.36508	19.02295	4.09	0.000	5.274668	112.5487
Riesgo Moderado	51.16666	53.29597	3.78	0.000	6.642977	394.1046
Riesgo Alto	54.26767	39.70268	5.46	0.000	12.93586	227.6603
Muy Alto Riesgo	231.8163	168.0223	7.51	0.000	55.99981	959.6248
_cons	.0032573	.002307	-8.09	0.000	.0008128	.0130536

. estat gof, group (10)

Logistic model for estadoegresopte, goodness-of-fit test

(Table collapsed on quantiles of estimated probabilities)  
(There are only 4 distinct quantiles because of ties)

number of observations = 1171  
number of groups = 4  
Hosmer-Lemeshow chi2(2) = 0.00  
Prob > chi2 = 1.0000

19. **Fórmula de regresión logística (logit)** con los predictores de **Bello-México** al aplicarlos en la cohorte RESCATE COVID-19.

$$\begin{aligned} \text{Log}(P/1-P) = & -3.553205 + -0.1562556 * \text{Edad} < 40 + 1.115175 * \text{Edad} \geq 65 + 0.0157836 * \text{Diabetes} + 1.460101 * \text{Diabetes} < 40 \\ & 0.8393669 * \text{Obesidad} + 1.434224 * \text{Neumonía} + (-0.057203) * \text{ERC} + 1.181918 * \text{EPOC} \\ & + (-0.6383286) * \text{Inmunosupresión} \end{aligned}$$

**Anexo 11:** Evaluación del desempeño de los modelos chinos en el grupo de pacientes con RT-PCTR para SARS-COV2 negativa de la cohorte RESCATE COVID 19.

Tabla 1S. Asociación entre predictores del modelo **Zang-China** y mortalidad intrahospitalaria en una cohorte de validación externa de pacientes con **PCR Negativa**:

Descripción		Análisis Bivariado (Regresión Logística Simple)					Análisis Multivariado (Regresión Logística Múltiple)			
Predictor	General n = 1258	Mortalidad		Riesgo en el Promedio (%)	OR crudo (IC 95%)	p	B	EE	OR ajustado (IC 95%)	p
		No n = 1145	Si n = 97							
<b>Edad (años)</b>	46.2 (23.8)	44.0 (22.8)	70.4 (21.1)	4.62	1.05 (1.03 – 1.06)	<0.01*	0.022996	0.006525	1.02 (1.01 – 1.03)	<0.01**
<b>Sexo (% hombres)</b>	542 (43.1)	479 (41.8)	56 (57.7)	5.7‡	1.89 (1.24 – 2.88)	<0.01*	0.075348	0.250970	1.07 (0.65 – 1.76)	0.764
<b>Rec. Neutrófilos (10<sup>3</sup>/L)</b>	8.7 (10.6)	8.4 (11.1)	9.9 (6.8)	13.8	1.00 (0.99 – 1.02)	0.25	0.017847	0.013408	1.01 (0.99 – 1.04)	0.183
<b>Rec. Linfocitos (10<sup>3</sup>/L)</b>	1.8 (1.3)	1.9 (1.4)	1.2 (0.9)	11.2	0.48 (0.35 - 0.65)	<0.01*	-0.420532	0.153945	0.65 (0.48 – 0.88)	<0.01**
<b>Rec. Plaquetas (10<sup>3</sup>/L)</b>	262.0 (111.6)	269 (107.9)	223 (130.1)	13.0	0.99 (0.993 – 0.998)	<0.01*	-0.002915	0.001238	0.997 (0.994 – 0.999)	0.01**
<b>Creatinina (μmol/L)</b>	1.46 (2.09)	1.3 (1.9)	2.1 (2.7)	15.4	1.13 (1.03 – 1.23)	<0.01*	0.095636	0.044934	1.10 (1.00 – 1.20)	0.03**
<b>PCR (mg/dL)</b>	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.

Los valores de laboratorio corresponden a los medidos al ingreso a urgencias.

Para todas las variables continuas se reporta media y desviación estándar.

Para sexo se reporta conteo y proporción en porcentaje (%)

B: Coeficiente (B1), EE: Error Estándar.

\*Asociación con significancia estadística p<0.05.

‡ Riesgo en mujeres

Número de observaciones: 550

Prueba de bondad de ajuste para el modelo de regresión logística múltiple de Hosmer-Lemeshow: p = 0.00752

**Tabla 2S. Valores perdidos de cada variable del modelo Zang-China en pacientes con RT-PCR para SARS-COV2 negativa en la cohorte RESCATE COVID-19**

Variable	Valores perdidos	Total	Porcentaje de pérdidas (%)
Muerte	16	1258	1.27
Edad	8	1258	0.64
Sexo (Hombres)	1	1258	0.08
Rec. Neutrófilos 10 <sup>3</sup> /L	582	1258	<b>46.26</b>
Rec. Linfocitos 10 <sup>3</sup> /L	582	1258	<b>46.26</b>
Rec. Plaquetas 10 <sup>3</sup> /L	585	1258	<b>46.50</b>
Creatinina (μmol/L)	688	1258	<b>54.69</b>

**Tabla 3S. Asociación entre predictores del modelo Xie-China y mortalidad intrahospitalaria en una cohorte de validación externa de pacientes con PCR negativa.**

Predictor	Descripción			Análisis Bivariado (Regresión Logística Simple)			Análisis Multivariado (Regresión Logística Múltiple)			
	General n= 1258	Mortalidad		Riesgo en el Promedio (%)	OR crudo (IC 95%)	p	B	EE	OR ajustado (IC 95%)	p
		No n = 1145	Si n = 97							
Edad (años)	46.2 (23.8)	44.0 (22.8)	70.4 (21.1)	4.62	1.05 (1.03 – 1.06)	<0.01*	0.0137025	0.0073403	1.01 (0.99 – 1.02)	0.062
LDH (U/L)	339.0 (373.6)	315 (388)	434 (307)	19.9	1.00 (0.99 – 1.00)	0.065	0.0005165	0.0003397	1.00 (0.99 - 1.00)	0.128
Rec. Linfocitos (10 <sup>3</sup> /L)	1.8 (1.3)	1.9 (1.4)	1.2 (0.9)	11.2	0.48 (0.35 - 0.65)	<0.01*	-0.7700531	0.2217164	0.46 (0.29 – 0.71)	<0.01**
Saturación de oxígeno (%)	91.9 (9.4)	92.4 (8.6)	86.0 (14.6)	7.7	0.96 (0.94 – 0.97)	<0.01*	-0.0200849	0.0128059	0.98 (0.11 – 1.005)	0.117

Los valores de laboratorio y saturación de oxígeno corresponden a los medidos al ingreso a urgencias.

Para todas las variables continuas se reporta la media y la desviación estándar.

\*Asociación con significancia estadística p<0.01

\*\* Asociación con significancia estadística p<0.05 tras ajustar efecto de covariables en el modelo.

B: Coeficiente (B1), EE: Error Estándar, LDH: Lactato deshidrogenasa.

Número de observaciones en el modelo: 312

Prueba de bondad de ajuste de Hosmer-Lemeshow:  $p = 0.017$

Tabla extra. Valores perdidos de cada variable del modelo Xie-China:

Variable	Valores perdidos	Total	Porcentaje de pérdidas
Muerte	16	1258	1.27
Edad	8	1258	0.64
Deshidrogenasa Láctica	915	1258	<b>72.73</b>
Rec. Linfocitos ( $10^3/L$ )	582	1258	<b>46.23</b>
Saturación de oxígeno	105	1258	<b>8.35</b>