



**Análisis de datos de la COVID-19 en Gipuzkoa.
Pruebas diagnósticas y modelos predictivos.**

AUTORA

M^a TERESA IGLESIAS GASPAR

Directores

José Ignacio Emparanza Knörr

Iratxe Urreta Barallobre

FEBRERO 2022

“A veces hay que pasar por caminos difíciles
para llegar a destinos maravillosos”

A mis princesas, Ainhoa y Nahia,
por ser el amor de todos mis segundos...
Estáis en todo y en todo momento.

AGRADECIMIENTOS

No imaginé que fuera tan difícil escribir estas líneas y lo es porque a todas las personas que quiero agradecerles el apoyo mostrado son demasiado importantes como para resumir en simples palabras tantos sentimientos.

En primer lugar, quiero agradecer a mis directores de tesis: José Ignacio Emparanza por todo el tiempo compartido, las enseñanzas (tanto en el trabajo como en la vida), tu infinita sabiduría y tu *inconmensurable* generosidad. Y a Iratxe Urreta, por ser mi apoyo desde que llegué, por estar a mi lado y ser más que una compañera.

A María José, por tus sabios consejos, tu cariño y por animarme a cada paso. Ha sido un regalo conocerte y voy a echar muchísimo de menos verte a diario. A Eukene, por encontrar siempre un hueco para ayudarme con las búsquedas. Por hacer las cosas fáciles y ser una excelente profesional además de gran persona. Y a Mari Mar, por las risas y los buenos momentos. Por ser el alma de tantas fiestas y por dedicar siempre minutos a nuestras charlas.

A todos los profesionales que han participado, directa o indirectamente, en esta tesis. Al personal de Urgencias, de Medicina Interna, de Infecciosas, de Farmacia, de la UCRI, de UNIRA y de Anestesia. Y A Xabi G. por su ayuda con la extracción de datos y por su infinita paciencia.

A mi familia de Madrid, por hacer que vuestra casa sea la mía y por estar siempre. Nunca os lo podré agradecer lo suficiente.

A mis padres, por ser dos seres humanos maravillosos. Si hubiera podido elegirlos, sin duda, habrías sido mi primera opción. Os agradezco muchísimo haberme enseñado a amar sin reservas y haber sido un ejemplo permanente.

A mis hermanos (Pedro, Arancha y Begoña), no os imagináis lo muchísimo que os quiero. Sois una fuente de inspiración constante y me siento privilegiada por teneros. Sois un regalo extraordinario.

Y a ti, Julen, por demostrarme que el amor no tiene medida. Por estar ahí siempre y por ser el mejor compañero de vida que se pueda tener. Te elijo cada día porque no imagino nadie mejor para compartir cada instante.

ÍNDICE DE CONTENIDO

Antecedentes y estado actual del tema	29
Comienzo y propagación del SARS-CoV-2	33
Modelos de predicción	34
o Conceptos previos sobre modelos de predicción	34
▪ Qué son y para qué sirven los modelos predictivos	34
▪ Clasificaciones y tipos de RPC	35
▪ Diseños de estudio para RPC	36
▪ Desventajas de las RPC	36
▪ Técnicas estadísticas para modelizar las RPC	37
▪ Decisiones previas a la creación de un modelo predictivo	38
▪ Etapas fundamentales en la creación de un modelo predictivo	48
▪ Medidas del rendimiento de los modelos de predicción	57
▪ Comparación de modelos predictivos	67
▪ Conclusiones	69
o Evaluación de riesgo de sesgos en modelos de predicción	70
▪ PROBAST	71
▪ CASPe	73
▪ JAMA	75
o Reportar estudios de modelos de predicción (TRIPOD)	76
o Modelos predictivos creados para la COVID-19	78
Pruebas diagnósticas de laboratorio para COVID-19	85
o Prueba RT-PCR	85
o Prueba rápida de antígenos PANBIO™	89
o Prueba de determinación de anticuerpos (Serología)	90
Scores creados en el HUD	94
o Score de Sospecha Clínica (SSC)	94
o Risk Score	96
Planteamiento de la tesis doctoral	97
Hipótesis y objetivos	103
Hipótesis	105
Objetivos	106

Material y métodos	109
Tipos de estudios.....	113
Ámbito	114
Búsqueda bibliográfica.....	114
o Para modelos predictivos de diagnóstico y de pronóstico.....	114
o Para la prueba rápida de antígenos PANBIO™.....	115
Sujetos de estudio	116
Recogida de datos.....	116
o Datos para modelos predictivos de diagnóstico y de pronóstico..	116
o Datos de pacientes del HUD.....	116
o Datos de los estudios de pruebas rápidas de antígenos PANBIO™	119
.....	
Criterios de inclusión y exclusión	120
o Criterios de inclusión	120
▪ Criterios de inclusión para los modelos predictivos de diagnóstico	120
y de pronóstico	
▪ Criterios de inclusión para los pacientes de la OSI-Donostialdea	120
▪ Criterios de inclusión para los estudios de pruebas rápidas de	120
antígenos (PANBIO™).....	
o Criterios de exclusión	121
▪ Criterios de exclusión para los modelos predictivos de diagnóstico	121
y de pronóstico	
▪ Criterios de exclusión para los pacientes de la OSI-	121
Donostialdea.....	
▪ Criterios de exclusión para los estudios de pruebas rápidas de	121
antígenos (PANBIO™).....	
Tamaño muestral.....	122
Variables de estudio	122
o Variables dependientes (outcome principales)	122
o Variables independientes.....	123
Análisis estadístico	131
o Modelos predictivos de pronóstico	131
o Pruebas diagnósticas de laboratorio	132
o Scores creados en el HUD	133
Consideraciones éticas.....	134
Resultados	135
Descripción de los pacientes de la OSI-Donostialdea	139

o	Características de pacientes ingresados en el HUD en la 1ª ola ...	141
o	Características de los pacientes COVID-19 del HUD en la 1ª ola ...	142
	Modelos predictivos de diagnóstico y de pronóstico	144
o	Modelos predictivos de diagnóstico.....	144
o	Modelos predictivos de pronóstico	146
▪	Resumen de los modelos predictivos incluidos en la RS	148
▪	Evaluación de riesgo de sesgos (PROBAST).....	147
▪	Evaluación del reporte de resultados (TRIPOD).....	155
▪	Selección del mejor modelo predictivo.....	159
▪	Segunda validación externa del score ANDC	160
▪	Comparativa de los pacientes de la 1ª y de la 2ª ola del HUD .	161
▪	Descripción y análisis del rendimiento del algoritmo de triaje modificado de urgencias tras la implementación del score ANDC en dicho algoritmo	166
	Caracterización de pruebas diagnósticas de laboratorio	178
o	Prueba RT-PCR	178
▪	Sensibilidad de la 1ª RT-PCR con los datos de pacientes de la 1ª ola del HUD.....	179
▪	Sensibilidad de la 2ª RT-PCR con los datos de pacientes de la 1ª ola del HUD.....	180
▪	Probabilidad de que un paciente sospechoso de COVID-19 se positivice (tras obtener dos RT-PCR negativas consecutivas con un máximo de 48h entre ambas) en los siguientes 7 días.....	180
o	Prueba rápida de antígenos PANBIO™ (Ag Rapid test Device)	182
▪	Resumen de los estudios incluidos en la RS	182
▪	Combinación de estudios de pacientes sintomáticos	184
▪	Combinación de estudios de pacientes asintomáticos.....	187
▪	Sensibilidad de la prueba rápida de antígenos Panbio™ (Ag Rapid Test Device) relativa a la RT-PCR, con datos de pacientes de la 1ª ola del HUD.....	190
o	Prueba de determinación de anticuerpos (Serología)	191
▪	Sensibilidad de la prueba de determinación de anticuerpos (Serología), relativa a la RT-PCR, con datos de pacientes de la 1ª ola del HUD.....	192
	Scores creados en el HUD	193
o	Score de sospecha clínica (SSC)	193

▪	Usos del SSC para discriminar entre enfermos COVID-19 y no enfermos COVID-19	193
▪	Usos del SSC para discriminar pacientes que se van a positivizar de los que no, de entre los que tienen RT-PCR negativa	195
▪	Estudio de la edad para mejorar la discriminación que hace el SSC entre pacientes que se van a positivizar de los que no.....	197
▪	Estudio de algunas de las variables que componen el SSC para mejorar la discriminación que hace el SSC entre pacientes que se van a positivizar y los que no	199
○	Risk score	201
Discusión	203
Modelos predictivos.....		207
○	Dificultades durante el desarrollo de la tesis	207
○	Uso de la herramienta PROBAST para la evaluación de riesgo de sesgos	209
▪	Tamaño muestral.....	210
▪	Selección de predictores	212
▪	Missings data	214
▪	Dicotomización de variables continuas	215
▪	Técnicas estadísticas	215
▪	Overfitting	216
▪	Test Hosmer-Lemeshow	217
▪	Fórmula completa de regresión o sistema de puntuación.....	218
▪	Aplicabilidad.....	219
○	Utilidad clínica.....	220
Pruebas diagnósticas de laboratorio		224
○	Prueba RT-PCR.....	224
○	Prueba rápida de antígenos Panbio™	225
○	Prueba de determinación de anticuerpos (Serología)	227
Scores propios creados en el HUD		228
○	Score de sospecha clínica (SSC).....	228
○	Risk score	229
Limitaciones.....		230
Acciones futuras		231
Conclusiones		233
Resumen.....		235
Conclusiones		237
Referencias bibliográficas		239
Anexos		255

ÍNDICE DE TABLAS

Tabla 1.	Jerarquías de evidencia en la derivación y la evaluación de una RPC	49
Tabla 2.	Algunas medidas de rendimiento de los modelos de predicción	66
Tabla 3.	PROBAST: Resumen de los 4 dominios para evaluación de riesgo y aplicabilidad.	72
Tabla 4.	JAMA Evidence. Users' Guides to the Medical Literature	75
Tabla 5.	Checklist TRIPOD: Desarrollo y Validación de un modelo de predicción	77
Tabla 6.	Resumen de la interpretación de los resultados de las pruebas moleculares y serológicas para el diagnóstico de la enfermedad COVID-19 causada por el virus SARS-CoV-2	91
Tabla 7.	Características que componen el SSC y su puntuación	95
Tabla 8.	Sospecha asignada según la puntuación obtenida en el SSC.....	95
Tabla 9.	Características que componen el Risk Score	96
Tabla 10.	Variables para modelos predictivos de pronóstico	123
Tabla 11.	Variables de ingreso del paciente	123
Tabla 12.	Variables sociodemográficas del paciente	124
Tabla 13.	Variables de analíticas del paciente	125
Tabla 14.	Variables de laboratorio	125
Tabla 15.	Variables de constantes del paciente	125
Tabla 16.	Variables recogidas en el formulario COVID-19 de Atención Primaria	126
Tabla 17.	Variables recogidas en el formulario COVID-19 de Atención Primaria (continuación).	127
Tabla 18.	Características de los pacientes que sufrieron variaciones, en cuanto a su recogida o no, en la primera ola	127
Tabla 19.	Variables recogidas en el formulario COVID-19 del HUD	128
Tabla 20.	Variables para la segunda validación externa del score ANDC	129
Tabla 21.	Variables para el análisis del uso y rendimiento del score ANDC en Urgencias.	129
Tabla 22.	Variables para el MA del test de antígenos Panbio™	130
Tabla 23.	Características seleccionadas de los pacientes ingresados en el HUD en la 1ª ola.	141
Tabla 24.	Características seleccionadas de los pacientes COVID-19 ingresados (HUD) en la 1ª ola	143

Tabla 25. RPC para pronóstico seleccionadas para ser validadas en nuestra población de estudio.....	148
Tabla 26. Resumen de las RPC validadas externamente en el HUD con datos de la 1ª ola.	153
Tabla 27. Prevalencias de las RPC validadas externamente y de los datos de la 1ª ola del HUD	154
Tabla 28. Evaluación del riesgo de sesgos con PROBAST en los estudios seleccionados.	155
Tabla 29. Características destacadas de los estudios seleccionados.....	156
Tabla 30. Cálculo del tamaño muestral mínimo requerido para los modelos predictivos seleccionados.....	158
Tabla 31. Probabilidad de morir en función del grupo de riesgo asignado según el ANDC score	160
Tabla 32. Edades medianas y rangos intercuartílicos para los pacientes de la 1ª y 2ª ola del HUD	161
Tabla 33. Fallecimientos de pacientes ingresados en la 1ª y la 2ª ola, desglosados por sexo.	163
Tabla 34. Letalidad de los pacientes ingresados en la 1ª y la 2ª ola, por grupos de edad .	165
Tabla 35. Características de pacientes (variables continuas) según score ANDC y destino al alta tras el triaje en urgencias.....	171
Tabla 36. Fallecidos según grupos de riesgo del score ANDC y destino al alta tras el triaje en urgencias	172
Tabla 37. Características de pacientes (variables categóricas) según score ANDC (riesgo alto) y destino al alta tras el triaje en urgencias.....	173
Tabla 38. Características de pacientes (variables categóricas) según score ANDC (riesgo moderado) y destino al alta tras el triaje en urgencias	174
Tabla 39. Características de pacientes (variables categóricas) según score ANDC (riesgo bajo) y destino al alta tras el triaje en urgencias	175
Tabla 40. Pacientes que retornan al HUD, según ANDC y destino al alta tras el triaje en urgencias cuando retornan (2ª visita).....	177
Tabla 41. Resultados de la prueba RT-PCR realizada en el HUD en la primera ola	178
Tabla 42. Resultados de la prueba RT-PCR realizada en el HUD en la primera ola, sin considerar los resultados dudosos o indeterminados.....	178
Tabla 43. Pacientes ingresados con formulario COVID-19 y con una RT-PCR en las 48h tras el ingreso	180
Tabla 44. Sensibilidad de la 2ª RT-PCR con los datos de la OSI-Donsotialdea (1ª Ola)...	180

Tabla 45. Estudios seleccionados para MA de la Se del test rápido de antígenos Panbio™.	183
Tabla 46. Se, Sp y prevalencia de los estudios seleccionados para MA del test rápido de antígenos Panbio™	184
Tabla 47. Subgrupo Osakidetza del HUD del estudio realizado por el ISCIII para conocer la Se de la prueba rápida de antígenos Panbio™.	190
Tabla 48. Datos de Se y Sp del Estudio del ISCIII, del Subgrupo Osakidetza y Abbott	191
Tabla 49. Resultados de la prueba de determinación de anticuerpos (Serología ICT) realizada en el HUD en la 1ª ola	191
Tabla 50. Resultados de la prueba de determinación de anticuerpos (Serología ICT) realizada en el HUD en la 1ª ola sin considerar las pruebas con resultado dudoso, no procede, muestra no recibida, no realizada o no valorable	191
Tabla 51. Pacientes con ambas pruebas (RT-PCR y Serología) en la 1ª ola	192
Tabla 52. Pacientes con ambas pruebas (RT-PCR y Serología) en la 1ª ola y se cumple que la RT-PCR es previa a la Serología y el tiempo transcurrido entre ambas pruebas es ≥ 15 días	192
Tabla 53. SSC inicial en relación al resultado de la RT-PCR	193
Tabla 54. SSC inicial en relación al resultado de la RT-PCR (simplificada)	194
Tabla 55. “Sospecha muy alta” vs “Sospecha alta, media o baja” en relación al resultado de la RT-PCR	194
Tabla 56. “Sospecha muy alta o alta” vs “Sospecha media o baja” en relación al resultado de la RT-PCR	194
Tabla 57. “Sospecha muy alta o alta o media” vs “Sospecha Baja” en relación al resultado de la RT-PCR	195
Tabla 58. SSC inicial en relación a la positivización del resultado de la RT-PCR	195
Tabla 59. SSC inicial en relación a la positivización del resultado de la RT-PCR (simplificada).	196
Tabla 60. “Sospecha muy alta” “Sospecha alta, media o baja” vs positivización del resultado de la RT-PCR	196
Tabla 61. “Sospecha muy alta o alta” “Sospecha media o baja” vs positivización del resultado de la RT-PCR	196
Tabla 62. “Sospecha muy alta, alta o media” “Sospecha baja” vs positivización del resultado de la RT-PCR	197
Tabla 63. Comparativa del SSC para discriminar la positivización de pacientes según grupos de edad (corte 75 años)	197

Tabla 64. Test exacto de Fisher para comparar las sensibilidades obtenidas para “SSC Muy alta o alta” vs “SSC Media o baja”, en pacientes que se positivizan, con el corte de edad en 75 años	198
Tabla 65. Comparativa del SSC para discriminar la positivización de pacientes según grupos de edad (corte 65 años).....	198
Tabla 66. Test exacto de Fisher para comparar las sensibilidades obtenidas para “SSC Muy alta o alta” vs “SSC Media o baja”, en pacientes que se positivizan, con el corte de edad en 65 años	199
Tabla 67. Fiebre o Tos o Disnea vs positivización de la RT-PCR.....	200
Tabla 68. Fiebre y (Tos o Disnea) vs positivización de la RT-PCR.....	200
Tabla 69. Fiebre vs positivización de la RT-PCR	200
Tabla 70. Disnea vs positivización de la RT-PCR.....	200
Tabla 71. Tos vs positivización de la RT-PCR.	201
Tabla 72. Resumen de la validación del Risk Score en nuestra población	201

ÍNDICE DE FIGURAS

Figura 1. Ejemplo de un gráfico de calibración	58
Figura 2. Ejemplo de gráfico de curva ROC con distintas AUC	61
Figura 3. Ejemplo de curva de decisión	65
Figura 4. Plantillas CASPe para realizar lectura crítica de artículos de Reglas de Predicción Clínica.....	74
Figura 5. Exactitud diagnóstica de la RT-PCR según los días transcurridos desde el comienzo de los síntomas.....	88
Figura 6. Interpretación de los resultados de la prueba de determinación de anticuerpos..	91
Figura 7. Evolución de los niveles de ARN e inmunoglobulinas (IgM e IgG) después de la infección de la enfermedad COVID-19 causada por el virus SARS-CoV-2	92
Figura 8. Algoritmo para resumir el desarrollo de la tesis doctoral.....	101
Figura 9. Diagrama de flujo de los pacientes considerados en este estudio (1ª ola).....	140
Figura 10. Diagrama de flujo para modelos predictivos de diagnóstico.....	145
Figura 11. Diagrama de flujo para modelos predictivos de pronóstico	147
Figura 12. Ingresos de pacientes COVID-19 positivos en la 1ª y 2ª ola, según edad y sexo.	161
Figura 13. Ingresos de pacientes COVID-19 positivos en la 1ª y 2ª ola según grupos de edad.	162
Figura 14. Curvas de supervivencia en pacientes ingresados en la 1ª y la 2ª ola, desglosado por edad.....	164
Figura 15. Algoritmo para el triaje de los pacientes en el servicio de urgencias del HUD (Dic 2020)	159
Figura 16. Informe de resultados de pruebas de laboratorio, con el cálculo del score ANDC, de un paciente atendido en el servicio de urgencias del HUD	168
Figura 17. Pruebas RT-PCR realizadas diariamente en la primera ola de la pandemia en el HUD.....	179
Figura 18. Forest plot para la Se del Panbio™ Ag test en pacientes sintomáticos.....	185
Figura 19. Forest plot para la Se del Panbio™ Ag test en pacientes sintomáticos sin 4 estudios que producen la mayor proporción de la heterogeneidad	186
Figura 20. Efecto de la duración de los síntomas en la sensibilidad	187
Figura 21. Forest plot para la Se del Panbio™ Ag test en pacientes asintomáticos	188
Figura 22. Forest plot para la Se del Panbio™ Ag test en pacientes asintomáticos sin 2 estudios que producen la mayor proporción de la heterogeneidad	189

TABLA DE ABREVIATURAS

ABREVIATURA	DEFINICIÓN
ACC	Análisis de Casos Completos
ACD	Análisis de Curvas de Decisión
ACP	Análisis de Componentes Principales
AIC	Akaike Information Criterion
ANDC	Age, Neutrophil, D dimer, C reactive protein
AUC	Area Under the Curve
BIC	Bayesian Information Criterion
BN	Beneficio Neto
CASPe	Critical Appraisal Skills Programme Español
CE	de Conformidad Europea
CEIC	Comité Ético de Investigación Clínica
CENTRAL	Cochrane Central Register of Controlled Trials
COVID-PRECISE	Precise Risk Estimation to optimise covid-19 Care for Infected or Suspected patients in diverse settings
CRA	Concern Regarding Applicability
CRB65	Confusion, Respiratory rate, Blood pressure, Age 65
CSS	Scoring System of Covid-19
C-Statistic	Índice de concordancia de Harrell
CURB65	Confusion, Urea, Respiratory rate, Blood pressure, Age 65
DPC	Datos de Población Celular
DS-CRB65	Disease, Peripheral blood saturation-Confusion, Respiratory rate, Blood pressure, Age 65
EPP	Eventos por variable predictora
EPV	Eventos por variable
EQUATOR	Enhancing the QUALity and Transparency Of health Research
GoF	Goodness of Fit
HEWS	Hamilton Early Warning Score
H-L	Hosmer-Lemeshow
HNC-LL	Hipertensión, Neutrófilos, proteína C reactiva, Linfocitos, Lactato deshidrogenasa
IM	Imputación Múltiple
ISARIC	International Severe Acute Respiratory and emerging Infections Consortium
LASSO	Least Absolute Shrinkage and Selection Operator

ABREVIATURA	DEFINICIÓN
MAR	Missing at random
MCAR	Missing completely at random
MCO	Mínimos Cuadrados Ordinarios
Me	Mediana
MEWS	Modified Early Warning Score
ML	Machine Learning
MNAR	Missing not at random
MuLBSTA	Multi-lobular infiltration, hypo-Lymphocytosis, Bacterial coinfection, Smoking history, hyper-Tension and Age
NEWS	National Early Warning Score
NRI	Net Reclassification Improvement
OAS	Oracle Analytic Server
OGP	Osabide Global Primaria
PC	Predictores Candidatos
PCT	ProCalciTonina
PDIA	Prueba Diagnóstica de Infección Activa
PF	Predictores Finales
PLS	Partial Least Square
PROBAST	Prediction model Risk of Bias Assesment Tool
PROGRESS	Prognosis research strategy partnership
REMS	Rapid Emergency Medicine Score
RF	Random Forest
RI	Rango Intercuartílico
RoB	Risk of Bias
ROC	Receiver Operating Characteristic
SEWS	Standardized Early Warning Score
SIC	Sistemas de Información Clínica
SSC	Score de Sospecha Clínica
SVM	Support Vector Machine
TRIPOD	Transparent reporting of multivariable prediction model
VCS	Volumen, Conductividad y Dispersión
VPN	Valor Predictivo Negativo
VPP	Valor Predictivo Positivo
WBC	White Blood Cell
-2LL	Prueba de razón de verosimilitudes Likelihood

ANTECEDENTES Y ESTADO ACTUAL DEL TEMA

ANTECEDENTES Y ESTADO ACTUAL DEL TEMA

- COMIENZO Y PROPAGACIÓN DEL SARS-CoV-2
- MODELOS DE PREDICCIÓN
 - Conceptos previos sobre modelos de predicción
 - Qué son y para qué sirven los modelos predictivos
 - Clasificaciones y tipos de RPC
 - Diseños de estudio para RPC
 - Desventajas de las RPC
 - Técnicas estadísticas para modelizar las RPC
 - Decisiones previas a la creación de un modelo predictivo
 - Etapas fundamentales en la creación de un modelo predictivo
 - Medidas del rendimiento de los modelos de predicción
 - Comparación de modelos predictivos
 - Conclusiones
 - Evaluación de riesgo de sesgos en modelos de predicción
 - Reportar estudios de modelos de predicción (TRIPOD)
 - Modelos predictivos creados para la COVID-19
- PRUEBAS DIAGNÓSTICAS DE LABORATORIO PARA COVID-19
 - Prueba RT-PCR
 - Prueba rápida de antígenos PANBIO™
 - Prueba de determinación de anticuerpos (Serología)
- SCORES CREADOS EN EL HUD
 - Score de Sospecha Clínica (SSC)
 - Risk Score

1. COMIENZO Y PROPAGACIÓN DEL SARS-CoV-2

El 31 de diciembre de 2019 se reporta en Wuhan, provincia de Hubei, el primer caso de síndrome respiratorio agudo de etiología desconocida. Las autoridades chinas identifican un nuevo coronavirus, coronavirus de tipo 2 causante del síndrome respiratorio agudo severo (SARS-CoV-2), que causa la enfermedad de la COVID-19. A partir de ese momento, el virus se propaga por todo el mundo provocando un gran número de contagios y dejando cifras demoledoras de fallecidos. Esta propagación mundial fue declarada a principios de 2020 emergencia de salud pública de importancia internacional y, el 11 de marzo, pandemia por la Organización Mundial de la Salud (OMS). A fecha de 21 de diciembre de 2021, hay un total de 274,2 millones de personas contagiadas y un total de 5,35 millones de fallecidos en el mundo⁽¹⁾.

El primer paciente registrado en España con COVID-19 se conoció el 31 de enero de 2020. Fue un paciente alemán ingresado en La Gomera que dio positivo en coronavirus. Su estado era "leve" y se contagió, presuntamente, al contactar en Alemania con un infectado. Nueve días después se detectó otro caso de coronavirus en La Palma. Pero no fue hasta el 24 de febrero cuando el virus saltó a la península, detectando los primeros casos en la Comunidad de Madrid, Cataluña y la Comunidad Valenciana⁽²⁾. En Euskadi, el protocolo contra el coronavirus se activó por primera vez a finales de febrero. El día 28 se confirmaron los dos primeros casos positivos, el de una persona en Gipuzkoa y el de una sanitaria en Álava. A partir de esta última, trabajadora del Hospital Universitario Araba-Txagorritxu, se contagió otra compañera. En poco tiempo, el hospital vitoriano pasó de ser el hospital de referencia para la lucha contra el virus a convertirse en el punto cero de los contagios y, Vitoria y Álava, en uno de los grandes focos⁽³⁾.

A 21 de diciembre de 2021, en España, se han contagiado más de 5,45 millones de personas y han fallecido 88.708 personas, según datos oficiales del Ministerio de Sanidad. De los cuales, 306.578 contagios y un total de 5.121 fallecidos corresponden a datos del País Vasco⁽¹⁾.

2. MODELOS DE PREDICCIÓN

2.1. CONCEPTOS PREVIOS SOBRE MODELOS DE PREDICCIÓN.

2.1.1. Qué son y para qué sirven los modelos de predicción.

Existen en la literatura diferentes términos: herramienta de predicción clínica⁽⁴⁾, algoritmo de predicción clínica⁽⁵⁾, modelo de predicción de riesgo⁽⁶⁾, modelo de riesgo⁽⁷⁾, regla de decisión clínica⁽⁸⁾, modelo pronóstico⁽⁹⁾, herramienta de decisión clínica⁽¹⁰⁾, índice de riesgo⁽¹¹⁾, score pronóstico⁽¹²⁾ o score de riesgo⁽¹³⁾ entre otras, que se usan indistintamente para referirse a este tipo de modelos predictivos. Hemingway et al⁽¹⁴⁾ definen una regla de predicción clínica (RPC) como una herramienta compuesta por un conjunto de variables (2 o más) de la historia clínica, exploración física y pruebas básicas complementarias⁽¹⁵⁾, que estima la probabilidad de un evento, sugiere un diagnóstico o predice una respuesta concreta a un tratamiento. Y Reilly et al⁽¹⁶⁾, indican que el elemento clave para pasar de un modelo de predicción a una RPC es la definición de un umbral de decisión o un punto de corte.

En esta tesis, los términos: *RPC* o *modelos predictivos*, se usan para referirnos a reglas, modelos de diagnóstico o modelos pronóstico derivados de análisis estadísticos multivariantes, que permiten predecir la probabilidad de una condición o resultado, con o sin el uso de un punto de corte clínico.

Algunas RPC han sido creadas basándose en la intuición de los clínicos que, debido a su experiencia profesional, pueden identificar qué elementos de los antecedentes, del examen físico y de las pruebas complementarias son necesarios para obtener un diagnóstico o un pronóstico adecuado, pero en general, el desarrollo de una RPC debe seguir un proceso estricto y complejo, de manera que otorgue un respaldo científico a dicha RPC y se justifique el uso generalizado de la misma en la práctica clínica.

2.1.2. Clasificaciones y tipos de RPC.

Existen diferentes clasificaciones de las RPC según diferentes autores⁽¹⁷⁻¹⁹⁾. Reilly et al⁽¹⁶⁾ establecen una clasificación de las RPC y diferencian entre:

- *Reglas de decisión*: pretenden influir directamente en la toma de decisiones, recomendando un curso de acción clínica específico al médico.
- *Reglas de predicción*: pretenden ayudar a los médicos a predecir el riesgo, sin proporcionar recomendaciones, con el supuesto de que las predicciones precisas conducirán a mejores decisiones.

En este estudio, nos vamos a centrar en las reglas de predicción. Por lo tanto, nuestro objetivo será determinar la probabilidad de un outcome específico con diferentes combinaciones de predictores en una población bien definida. Estos modelos utilizan datos de pacientes que permiten calcular una probabilidad predicha de enfermedad existente (modelos de diagnóstico) o resultado futuro (modelos de pronóstico).

En el caso concreto de la enfermedad de la COVID-19, nuestro interés se centra en estudiar:

- *Modelos predictivos de diagnóstico* que permitan predecir la presencia o la ausencia de la enfermedad en pacientes con síntomas compatibles con la enfermedad de la COVID-19.
- *Modelos predictivos de pronóstico* que permitan predecir qué pacientes, de los que están infectados por la enfermedad, tienen mayor riesgo de sufrir un peor desenlace (ingreso en UCI) o un desenlace fatal (éxito).

2.1.3. Diseños de estudio para RPC.

Para la creación de una RPC se hace necesario establecer un diseño de estudio óptimo. En el caso de una RPC de diagnóstico, el estudio adecuado es un estudio transversal, mientras que, para las RPC de pronóstico, el diseño preferido es un estudio de cohorte longitudinal⁽⁷⁾.

En general, los estudios de casos y controles son inapropiados^(6,9,20), ya que sobreestiman la capacidad nosológica del modelo (es decir, hacen creer que el modelo es mejor de lo que en realidad es)^(21,22). Sin embargo, se propone como alternativa realizar estudios de casos y controles anidados o de cohortes de casos^(20,23).

Aunque desgraciadamente la literatura está dominada por estudios retrospectivos, son preferibles los estudios de cohortes prospectivos. El motivo es que, con los estudios prospectivos, se puede conseguir que las mediciones de las variables predictivas y del outcome sean óptimas y así evitar sesgos de medición del evento.

2.1.4. Desventajas de las RPC.

No existen las RPC perfectas, George Box, el eminente estadístico británico dijo que: "todos los modelos de predicción son incorrectos, pero algunos de ellos son útiles". Esto hace que sea necesario lograr un equilibrio entre la capacidad de aplicar el modelo en amplias cohortes de pacientes similares y la optimización del rendimiento en la cohorte de desarrollo.

Como consecuencia de esto, sólo los modelos de predicción precisos, de alta calidad y generalizables deben implementarse en nuestros sistemas de salud para su uso en la práctica clínica habitual.

2.1.5. Técnicas estadísticas para modelizar las RPC.

Teniendo en cuenta la cantidad ingente de datos que se registran a diario en los sistemas de salud y a la enorme dificultad para manejarlos, se hace necesario el uso de técnicas estadísticas que permitan trabajar con un gran número de registros y con una amplia cantidad de predictores. Así como, programas estadísticos que posibiliten la explotación de datos en diferentes formatos y en los que se permita trabajar tanto con datos estructurados como con datos no estructurados.

Los modelos de regresión más habituales para desarrollar una RPC, en función del outcome que se quiere predecir, son:

- Outcome continuo: regresión lineal y regresión lineal efectos mixtos.
- Outcome binario: regresión logística y regresión logística efectos mixtos.
- Tiempo hasta el evento (outcome binario, pero teniendo en cuenta el momento temporal en el que éste se produce): modelos paramétricos de supervivencia o Weibull y riesgos proporcionales de Cox.

Así mismo, existen otras técnicas estadísticas que se engloban dentro del ámbito de la Inteligencia Artificial y que también se usan para modelizar las RPC: Support Vector Machine -SVM-, árboles de decisión, redes neuronales, análisis bayesiano, análisis discriminante, partición recursiva (divide la población de pacientes en grupos cada vez más pequeños, basándose en factores de riesgo discriminantes o redes neuronales)^(15,24), etc.

En el ámbito de la medicina, la técnica estadística más empleada para el desarrollo de una RPC ha sido el modelo de regresión logística, ya que es una de las técnicas con mejor capacidad para el análisis de datos en investigación clínica. Mediante regresión logística se seleccionan variables predictoras para construir un modelo (parsimonioso) que permita describir, explicar o predecir el outcome y evaluar la contribución de cada una de las variables predictoras (los coeficientes de regresión son mutuamente ajustados por los otros predictores y cada coeficiente representa la contribución de ese predictor a la probabilidad del outcome).

2.1.6. Decisiones previas a la creación de un modelo predictivo.

Siguiendo las recomendaciones del grupo PROGRESS⁽²⁵⁾ (prognosis research strategy partnership), antes de crear un modelo predictivo, debemos tomar numerosas decisiones que afectan al modelo y a las conclusiones de la investigación⁽²⁶⁾, como son:

2.1.6.1. Seleccionar los predictores candidatos clínicamente relevantes para entrar en el modelo.

En primer lugar, hay que seleccionar qué predictores deben ser incluidos en el modelo (selección pre-modelo). Se recomienda introducir en el modelo sólo los predictores que tengan una buena justificación teórica, y de ese modo, obtener un modelo estable, ya que Introducir variables sin base teórica, aunque sean predictivas (estadísticamente significativas), conduce a modelos poco generalizables y enmascaran el proceso de modelado.

La elección de predictores puede llevarse a cabo considerando una combinación de experiencia clínica, opinión de expertos, estudios cualitativos y de RS de la literatura con meta-análisis –MA-⁽²⁷⁻³¹⁾.

La información clínica, la exploración física y las pruebas complementarias deben ser clínicamente relevantes, precisas y reproducibles. Y para que los clínicos puedan aplicar la RPC sin dificultad es importante que dicha información esté disponible y sea fácilmente accesible en el momento de la toma de decisiones.

Hosmer, Lemeshow y Sturdivant (2013)⁽³²⁾ proponen realizar un análisis exploratorio de preselección de variables en base a un análisis univariado de cada predictor respecto al outcome que se desea medir y seleccionar a aquellos predictores que hayan obtenido un $p\text{-valor} \leq 0.20$ (si la muestra es pequeña, <100 sujetos, considerar los predictores que hayan obtenido un $p\text{-valor} < 0.3$).

El inconveniente de realizar un análisis univariado y establecer un umbral de significación de 0.05 (en lugar de emplear un umbral de $p \leq 0.2$) es, que:

- Si la muestra es pequeña, predictores relevantes pueden quedar fuera del modelo por ausencia de asociación estadísticamente significativa (debido quizás a la falta de potencia estadística, condicionada por el escaso tamaño muestral).

- Si la muestra es muy grande, puede dar lugar a problemas de sobreajuste debido a que se incluirán más predictores en el modelo (cualquier diferencia existente, por pequeña que sea, será estadísticamente significativa, aunque no tenga relevancia clínica).

Otra alternativa para seleccionar predictores es aplicar ciertas estrategias para reducir el número de predictores y seleccionar un conjunto de ellos⁽³³⁾, como son: subset selection, regularización (Ridge, Lasso y Elastic Net) y técnicas de agrupamiento de variables predictoras.

Subset selection⁽³³⁾

Se emplean mínimos cuadrados ordinarios (MCO) para ajustar un modelo lineal que contiene únicamente un subconjunto de predictores. Con esta técnica se realiza un proceso iterativo que elimina los predictores menos relevantes. Existen diferentes tipos de métodos subset selection: best subset selection, forward stepwise selection y backward stepwise selection. A continuación, explicamos cada uno de ellos.

Best subset selection

En este proceso, se exploran todos los posibles modelos que se pueden crear mediante combinación de los predictores disponibles. Si, por ejemplo, disponemos de k predictores:

- Se genera el modelo nulo (sin ningún predictor) "M0".
- Se generan todos los modelos con un único predictor y seleccionamos el que tiene el menor error. A este modelo lo denominamos "M1".
- Se repite este proceso para dos predictores y así sucesivamente hasta llegar al modelo que contiene a todos los predictores (Mk).
- De todos los modelos elegidos, para cada número de predictores, se selecciona el mejor modelo empleando una métrica que sirva para comparar los modelos.

Un inconveniente de este método es: que no es posible llevarlo a cabo cuando se dispone de más de 40 predictores, ya que hay que calcular 2^p modelos (donde p es el número total de predictores). Otro inconveniente es que, al generarse tantos modelos, puede haber problemas de sobreajuste⁽³¹⁾ (por simple azar se pueden obtener resultados significativos). Se recomienda usar este método si disponemos de un máximo de 10 predictores.

Forward stepwise selection

Es más eficiente que el *best subset selection* ya que no se evalúan todos los posibles modelos, sólo un subconjunto de ellos. Los pasos a seguir son:

- Se genera el modelo nulo (sin ningún predictor) "M0".
- Se generan todos los modelos con un único predictor y se selecciona el que tiene el menor error de entrenamiento, a este modelo lo denominamos "M1".
- Se repite este proceso para dos predictores, pero se parte del último modelo seleccionado, y así sucesivamente hasta llegar al modelo que contiene a todos los predictores (Mk).
- De entre todos los modelos seleccionados para cada n° de predictores, se selecciona el mejor empleando alguna métrica que permita comparar modelos.

Este método puede emplearse cuando el n° de predictores es mayor que el número de observaciones, suele proporcionar buenos modelos (no tiene por qué ser el mejor) y evita el sobreajuste.

Backward stepwise selection

Es similar al anterior en cuanto al concepto, pero se parte del modelo con todos los predictores posibles.

- En cada paso se generan todos los modelos que se pueden crear eliminando un único predictor y se selecciona el que tiene menor error.
- Se repite el paso anterior hasta que se llega al modelo nulo.
- De entre los mejores modelos seleccionados para cada n° de predictores (M0, M1, ..., Mk), se identifica el mejor empleando una métrica de comparación de modelos.

El inconveniente de este método es que se comienza con el modelo completo por lo que, si la regresión es por mínimos cuadrados, no se puede aplicar si el n° de observaciones es menor que el n° de predictores. La ventaja es, que permite evaluar cada variable en presencia de las otras.

Regularización o penalización (Shrinkage)⁽³³⁾

Con las estrategias de regularización se emplean todos los predictores para ajustar el modelo, pero se fuerza a que los coeficientes del modelo tiendan a cero. Esto ayuda a atenuar el efecto de la correlación entre predictores, a minimizar el riesgo de sobreajuste (overfitting), a reducir la varianza y a reducir la influencia en el modelo de los predictores menos relevantes. De este modo, en general, se consiguen modelos con mayor poder predictivo.

Estos métodos son adecuados cuando el n° de predictores es mayor que el n° de observaciones. Es muy importante estandarizar los predictores antes de derivar el modelo ya que la regularización actúa sobre la magnitud de los coeficientes del modelo y éstos deben estar en la misma escala.

Los modelos de regularización son: Ridge, Lasso y Elastic Net.

Ridge

Esta regularización se denomina "l2" y penaliza la suma de los coeficientes elevados al cuadrado. Reduce proporcionalmente el valor de los coeficientes del modelo (de todos a la vez), pero sin que éstos lleguen a tomar el valor cero. La regularización se aplica a través del parámetro λ (factor de contracción). Si $\lambda=0$, la penalización es nula (el resultado es equivalente al de un modelo lineal ajustado por MCO) y a medida que λ aumenta, el valor de los predictores será menor porque mayor es la penalización.

La desventaja es que incluye todos los predictores. A pesar de reducir la influencia de los predictores menos relacionados con la variable respuesta, no los elimina del modelo. Esto supone un problema para la interpretación del modelo.

La ventaja es que, empleando el valor adecuado de λ , el método es capaz de reducir la varianza (una varianza mayor indica que pequeños cambios en los datos tienen mucho impacto en el modelo resultante) sin apenas incrementar el sesgo en las estimaciones de los coeficientes del modelo, consiguiendo así un menor error total.

LASSO (least absolute shrinkage and selection operator)

La regularización de LASSO se denomina " l_1 " y penaliza la suma del valor absoluto de los coeficientes de regresión. Fuerza a que los coeficientes del modelo tiendan a cero. Los predictores menos relevantes toman el valor cero y son excluidos del modelo. Cuanto mayor es el valor de λ , mayor es la penalización y más predictores son excluidos.

No se puede saber de antemano cuál es el valor óptimo de λ . Una manera de obtenerlo es aplicando *cross-validation* y, dicho valor será el que minimice el error de la validación cruzada. En la práctica, está muy extendido seleccionar el valor de λ que se aleje menos de 1 desviación típica del óptimo.

Si hay una alta correlación entre predictores, Lasso tiende a seleccionar uno de ellos (al que le da todo el peso) y excluye al resto de predictores. Las predicciones de Lasso son muy inestables cuando hay predictores muy correlacionados, ya que pequeños cambios en los datos pueden producir grandes cambios en los predictores seleccionados.

Elastic Net

Es una regularización que combina las dos anteriores. El parámetro que indica cuánto influye cada una de las regularizaciones, l_1 y l_2 , es α (alpha). Si $\alpha=0$ se aplica ridge y si $\alpha=1$ se aplica Lasso. Frecuentemente se suele dar casi todo el peso a la penalización l_1 (para seleccionar los predictores) y un peso bastante más pequeño a l_2 (para estabilizar el modelo si hay predictores correlacionados).

Técnicas de agrupamiento⁽³³⁾

Además de los métodos *subselection* o de *regularización* existen otras técnicas, denominadas de *agrupamiento*, para reducir la dimensionalidad. Dichas técnicas consisten en crear un conjunto de componentes, a partir de combinaciones lineales o no lineales de las variables originales, y con esas componentes se ajusta el modelo. Es útil cuando hay un gran número de variables predictoras, están altamente correlacionados y existe multicolinealidad.

Hasta ahora, los métodos que hemos explicado permiten controlar la varianza y el sobreajuste y, para ello, emplean un nº reducido de predictores (bien porque se seleccionan un subconjunto de ellos o porque se fuerza a que tiendan a cero) y se emplean las variables originales, sin ser modificadas (salvo cuando las estandarizamos). Con las técnicas de agrupamiento los modelos generados tendrán un menor número de predictores que serán componentes que hemos creado a partir de las variables originales y tendrán casi la misma información que aportaban las variables originales.

Las dos técnicas de agrupamiento más conocidas son: Análisis de Componentes Principales y Partial Least Square.

Análisis de componentes principales (ACP)

Para la selección de los predictores se lleva a cabo un ACP. Se deben estandarizar los predictores antes de calcular las componentes para evitar que las variables que se midan en una escala mayor o las que tengan mayor varianza, tengan más peso. La selección del nº óptimo de componentes se puede llevar a cabo de diferentes maneras, una de ellas puede ser empleando una técnica de validación cruzada.

Partial Least Square (PLS)

Esta técnica es muy similar a la ACP ya que en ambas se obtienen las componentes con un ACP, pero en PLS se busca que, además de explicar la varianza observada a través de las componentes, éstas predigan el outcome lo mejor posible (es decir, se tiene en cuenta la relación del outcome con las combinaciones lineales seleccionadas para cada componente).

2.1.6.2. Elegir la estrategia para la selección de las variables importantes en el modelo final.

Una vez seleccionados los predictores, se introducen todos los que se han seleccionado en el paso anterior en un modelo multivariante y se emplea una técnica para la selección de los predictores relevantes. Esta selección de predictores relevantes es una selección intra-modelo, y permitirá seleccionar los predictores finales que conformarán el modelo predictivo.

Existen diferentes técnicas para llevar a cabo esta selección. Las técnicas más comunes son⁽³⁴⁾:

Selección hacia adelante (*Forward Stepwise Regression*).

- Las variables se introducen secuencialmente en el modelo.
- La primera variable que se introduce es la de mayor correlación (+ó-) con la variable dependiente (outcome)
- Dicha variable se introducirá en la ecuación sólo si cumple el criterio de entrada.
- A continuación, se considera la variable independiente cuya correlación parcial sea la mayor y que no esté en la ecuación.
- El procedimiento termina cuando ya no quedan variables que cumplan el criterio de entrada.

Eliminación hacia atrás (*Backward Stepwise Regression*).

- Se introducen todas las variables en la ecuación y después se van excluyendo una tras otra.
- En cada etapa se elimina la variable menos influyente según el contraste individual.

Pasos sucesivos (*Stepwise Regression*).

- Este método es una combinación de los procedimientos anteriores.
- En cada paso se introduce la variable independiente que no se encuentre ya en la ecuación y que cumpla el criterio de entrada.
- Las variables ya introducidas en la ecuación de regresión pueden ser eliminadas del modelo (i.e. hacia atrás).
- El método termina cuando ya no hay más variables candidatas a ser incluidas o eliminadas.

2.1.6.3. Evaluar la calidad de los datos y decidir qué hacer con los datos faltantes (missing values).

Hay 3 alternativas para trabajar con datos faltantes: omitir las variables que tengan valores faltantes, omitir individuos con datos faltantes o estimar (imputar) dichos valores. En este último caso, los datos faltantes serán reemplazados con valores predichos a partir de los datos presentes.

Los missing values se clasifican en tres categorías⁽³⁵⁾ y habrá que considerarlo antes de decidir realizar la imputación de datos:

- MCAR (missing completely at random): en este caso no hay diferencias entre los valores missing y los valores observados.
- MAR (missing at random): en este caso, la probabilidad de un valor perdido depende de los valores observados en las otras variables, pero no de los valores no observados.
- MNAR (missing not at random): indica que la probabilidad de un valor perdido depende de los valores no observados y de otros predictores no observados, condicionado a los datos observados.

2.1.6.4. Establecer el manejo de datos (creación de variables, colapsar categorías, ...).

2.1.6.5. Decidir cómo se van a considerar las variables continuas⁽³⁶⁻³⁹⁾.

Es habitual encontrar modelos predictivos formados por un conjunto de variables en los que una o varias de ellas se hayan categorizado, pero no se recomienda categorizar estas variables porque producen pérdida de información y pérdida de precisión.

2.1.6.6. Seleccionar las medidas de rendimiento del modelo o de precisión predictiva (ver apartado 2.1.8.).

2.1.6.7. Selección de outcome objetivos y relevantes para los pacientes.

2.1.6.8. Otras consideraciones incluyen:

- Decidir cómo se va a evaluar la solidez del modelo a observaciones influyentes y valores atípicos.
- Hay que estudiar la posible interacción entre predictores, decidir si ajustar el modelo final para evitar el sobreajuste (Shrinkage factor) y cómo hacerlo.
- Se debe plantear cómo se va a explorar la estabilidad (reproducibilidad) del modelo.

2.1.6.9. Calcular el tamaño muestral mínimo requerido.

Un aspecto fundamental previo a la derivación de un modelo predictivo es el tamaño muestral mínimo requerido. Esta última cuestión no viene recogida expresamente en las recomendaciones que establece el grupo PROGRESS, pero está tomando cada vez más relevancia y consideramos que es imprescindible incluirlo como una cuestión más a tener en cuenta.

No está claro qué constituye un tamaño de muestra suficientemente grande para el desarrollo de modelos⁽²⁹⁾ y existen varias “reglas generales” propuestas y debatidas⁽⁴⁰⁻⁴⁴⁾ sobre esta cuestión. Recientemente, Riley et al⁽⁴⁵⁾, han publicado una guía práctica para calcular el tamaño de muestra mínimo requerido para el desarrollo de modelos de predicción clínica, que se basa en artículos metodológicos publicados por ellos mismos en el último tiempo. Los autores realizan una propuesta tanto para outcome continuos como para outcome binarios o tiempo hasta el evento.

Históricamente se ha establecido que debemos disponer de al menos 10 eventos por variable (EPV) en el modelo de predicción que se quiere derivar. Aunque se refiere, más concretamente, al nº de eventos por variable predictora (EPP), ya que el nº de términos en el modelo incluye: variables independientes, sus interacciones y términos polinómicos. La regla de 10 EPV se remonta a 1990, cuando se realizaron diversos estudios de simulación⁽⁴⁶⁻⁴⁸⁾, pero estudios recientes han demostrado que este umbral no tiene base científica⁽⁴⁰⁾ y esto ha propiciado que se haya debatido la idoneidad de la misma. Algunos autores incluso han sugerido EPP más altas, de al menos 20^(40,44,49).

Está demostrado que el número de EPP requerido depende no solo del número de eventos en relación con el número de parámetros predictores candidatos, sino también del número total de participantes, de la proporción de resultados (incidencia) en la población de estudio y del rendimiento predictivo esperado del modelo⁽⁴⁰⁻⁴⁴⁾. Con esta idea, Riley et al proponen un enfoque que puede ser implementado en 4 pasos⁽⁴⁵⁾: B1 (estimar la proporción del outcome con suficiente precisión), B2 (seleccionar un error de predicción absoluto promedio pequeño), B3 (seleccionar un factor de contracción de 0.9) y B4 (Seleccionar un valor de optimismo pequeño, de 0.05, en el R^2 Nagelkerke aparente). Cada paso conduce a un cálculo del tamaño de la muestra y, en última instancia, el tamaño de muestra más grande identificado será el requerido.

El número razonable de participantes con el resultado (que da lugar a un bajo riesgo de sesgo) difiere según estemos derivando o validando un modelo.

- Para la derivación de una RPC: en términos generales, los estudios con EPP inferiores a 10 es probable que tengan sobreajuste, mientras que aquellos con EPP superiores a 20 tienen menos probabilidades de tenerlo. Cabe destacar que el tamaño de la muestra necesario para minimizar el sobreajuste depende del contexto, de la prevalencia de los resultados, del rendimiento general del modelo (R^2) y de las distribuciones de los predictores⁽⁵⁰⁻⁵²⁾. Por lo tanto, decidir si se utilizó un tamaño de muestra apropiado puede ser difícil, más aún cuando la EPP está entre 10 y 20.
- Para los estudios de validación, algunos autores proponen un mínimo de 100 eventos con y sin el outcome (obtenido mediante simulación Monte Carlo), pero idealmente 200 eventos^(53,54). Otros han sugerido un mínimo de 200 eventos y 200 no eventos para evaluar de manera fiable la calibración y producir gráficos de calibración útiles^(55,56).

Por lo tanto, cuanto menor sea el tamaño de muestra efectivo y menor el n° de EPP, mayor será el riesgo de que el modelo de predicción final haya incluido predictores espurios (modelo "sobreajustado") o no haya incluido predictores importantes (modelo "desajustado"). Es probable que el sobreajuste y el desajuste produzcan estimaciones sesgadas del rendimiento predictivo aparente del modelo^(32,57-61).

2.1.7. Etapas fundamentales en la creación de un modelo predictivo.

La mayoría de los autores establecen 3 etapas en la creación de una RPC^(6,16,26,30,62-65):

- Desarrollo del modelo (denominado *derivación del modelo*): en esta etapa se produce la identificación de los predictores relevantes, la asignación de los pesos relativos de cada predictor, la estimación del rendimiento predictivo del modelo a través de la calibración y la discriminación y la validación interna con sobreajuste si fuese necesario.
- Validación de dicho modelo (calibración y discriminación) en nuevos pacientes (*validación externa*). La validación puede ser estrecha (para estudiar la reproducibilidad de la RPC), también denominada temporal, o puede ser una validación amplia (para estudiar la aplicabilidad de la RPC), denominada geográfica.
- *Impacto clínico o utilidad clínica* que permite cuantificar si el uso del modelo en la toma de decisiones de la práctica clínica realmente mejora los outcome (resultados) del paciente. El impacto clínico puede valorarse de dos modos: estrecho o amplio.

Sin embargo, Stiell et Wells⁽⁸⁾ establecen 3 etapas más para la creación de una RPC: identificación de la necesidad de creación de la RPC, determinación de la RPC y difusión e implementación a largo plazo de una RPC.

En esta tesis, nos vamos a centrar en analizar las 3 etapas fundamentales que establecen la mayoría de los autores. Estas tres etapas corresponden a jerarquías crecientes de evidencia, como se describe en la Tabla 1^(16,30,62).

Tabla 1. Jerarquías de evidencia en la derivación y la evaluación de una RPC.

Niveles de evidencia	Definición y estándar de evaluación	Implicaciones para los clínicos
Nivel 1. Derivación de la RPC	Identificación de predictores mediante modelo multivariante; evaluación cegada de los resultados	Necesita validación y evaluación adicional antes de que se utilice clínicamente en la atención real del paciente.
Nivel 2. Validación estrecha	Validación de la RPC cuando se prueba de forma prospectiva en un entorno; evaluación cegada de los resultados.	Necesita validación en diversos entornos; puede utilizar la RPC con precaución en pacientes similar a la muestra de derivación.
Nivel 3. Validación amplia	Validación de la RPC en diversos entornos con un amplio espectro de pacientes y médicos.	Necesita análisis de impacto; puede utilizar las predicciones de la RPC con confianza en su precisión.
Nivel 4. Análisis de impacto estrecho usado para la toma de decisiones	Demostración prospectiva de que el uso de la RPC mejora las decisiones de los médicos en un entorno (calidad o rentabilidad de la atención al paciente)	Puede usarse con cautela para informar decisiones en entornos similares al estudiado.
Nivel 5. Análisis de impacto amplio usado para la toma de decisiones	Demostración prospectiva de que el uso de la RPC mejora las decisiones de los médicos para un amplio espectro de pacientes, en diversos entornos.	Puede usarse en diversos entornos con la confianza de que su uso beneficiará la calidad o eficacia de la atención al paciente.

Creada por Reilly y Evans (2016) y adaptada por Cowley et al (2019)

Una RPC que tras su derivación no ha sido validada externamente, corresponde al nivel más bajo de evidencia y no se recomienda para su uso en la práctica clínica, excepto en casos raros cuando se desarrolla una RPC para ser usada en un solo entorno.

Una RPC que haya sido validada externamente con éxito en un entorno o población similar a aquella de la que se derivó (validación "estrecha"), puede usarse con precaución en pacientes futuros similares⁽¹⁶⁾.

Del mismo modo, se propone que una RPC debe validarse externamente con éxito de manera consistente en múltiples contextos o poblaciones (validación "amplia"), antes de que los médicos puedan usar sus predicciones con confianza en pacientes futuros.

2.1.7.1. Derivación de un modelo predictivo.

En la etapa de derivación del modelo predictivo, se asignan los pesos relativos a cada predictor tras haber seleccionado los predictores finales del modelo. Posteriormente, se estima el rendimiento predictivo del modelo a través de la calibración y la discriminación. Y, por último, se lleva a cabo la validación interna.

En el caso de que se haya obtenido el modelo final mediante la regresión logística, a cada variable del modelo se le asigna un coeficiente que nos informa de su contribución individual al modelo predictivo. El peso de estos coeficientes puede ser fácilmente entendido si se exponencia su valor al número "e", ya que corresponde con el valor de la odds ratio (OR) ajustada, que relaciona la variable con el resultado de interés.

Podemos aplicar los coeficientes obtenidos mediante la regresión logística a una fórmula de regresión y obtener la probabilidad individualizada del diagnóstico o evento adaptada a un paciente concreto. Otra opción es que, a partir de los coeficientes, pueden asignarse pesos a cada variable y estos pesos se traducen en puntuaciones de una escala. La puntuación total obtenida para cada paciente en esa escala permitirá clasificar al paciente en un rango concreto y, a ese rango, le corresponderá una recomendación o predicción determinada⁽⁶⁶⁾.

Validación interna para evaluar sobreajuste

Una vez creado el modelo predictivo es muy importante aplicar alguna técnica de validación de dicho modelo para evaluar si se ha producido sobreajuste (sesgo). Un modelo sobreajustado es un modelo estadístico que contiene más parámetros de los que pueden justificar los datos. Este modelo puede fallar en ajustar datos adicionales o predecir observaciones futuras de manera fiable. Debemos tener en cuenta que, es más probable que se sobrestime el rendimiento de un modelo cuando el conjunto de datos de derivación es pequeño y utiliza una gran cantidad de predictores candidatos. Por lo tanto, independientemente de los enfoques utilizados en la etapa de derivación del desarrollo, se requiere validación interna para examinar y corregir la cantidad de sobreajuste u "optimismo" en el modelo y, por lo tanto, la estabilidad del mismo⁽⁶⁾.

Para analizar si se ha producido o no sobreajuste, se suelen usar submuestras aleatorias de los pacientes de la muestra original, o técnicas más complejas como son los métodos Jackknife y Bootstrap⁽²⁴⁾. Cabe destacar que, la validación interna no valida un modelo en sí mismo sino el proceso utilizado para ajustar el modelo^(31,67,68).

El optimismo se estima utilizando únicamente el conjunto de datos de derivación original. Hay varios métodos disponibles para este propósito, que presentamos a continuación:

Rendimiento aparente (apparent validation)

Consiste en evaluar el rendimiento del modelo empleando los mismos datos que han sido usados en el desarrollo del mismo. Este método producirá, muy probablemente, que la estimación del rendimiento del modelo sea optimista y presentará problemas de sobreajuste.

Bootstrap validation⁽⁶⁹⁾

Consiste en generar muestras del mismo tamaño que la muestra original, para ello, se seleccionan de manera aleatoria individuos de la muestra original, pudiendo repetirse dichos individuos en la misma muestra (muestreo con reemplazamiento). Esta muestra aleatoria es repetida cientos de veces (normalmente 1000) para generar nuevas muestras simuladas, que serán similares, pero no idénticas y que proporcionarán rendimientos variables. El promedio de estas diferencias indica el optimismo en las métricas de rendimiento del modelo que fue inicialmente derivado en el conjunto de datos de derivación^(6,31,68,70) y permite el ajuste del rendimiento general para aproximarse mejor al rendimiento esperado del modelo en nuevas muestras⁽⁶⁾. De ese modo, se asegura que los estadísticos obtenidos sean precisos. Bootstrapping también estima un factor de contracción uniforme para permitir el ajuste de los coeficientes de regresión estimados por sobreajuste^(31,68,70).

Esta técnica de validación interna es la más empleada, especialmente cuando disponemos de una muestra de validación pequeña y hay un gran nº de predictores. Permite identificar cuáles son los predictores más fiables.

En cada análisis de regresión realizado en las diferentes muestras bootstrap simuladas obtendremos modelos diferentes que incluirán predictores distintos, habrá predictores que no aparecerán en ninguna ocasión y otros,

que estarán muy frecuentemente. Si aparecen en el 50% de los modelos, los podemos considerar como fiables y los debemos incluir en nuestro modelo de regresión final⁽⁷²⁾. Sin embargo, ningún procedimiento de validación interna puede sustituir a la validación externa; la validación interna solo aborda la variabilidad del muestreo, mientras que la validación externa considera la variación en la población de pacientes⁽⁶¹⁾.

División de los datos (split sample)⁽⁷¹⁾

Es el método más simple de validación. Divide la muestra original en dos submuestras, con una de ellas se desarrolla el modelo y, con la otra, se valida el modelo⁽⁶³⁾ (es decir, se evalúa el rendimiento del modelo derivado). Esta técnica exige que la muestra original sea grande ya que la muestra de validación debe ser amplia. Sin embargo, la muestra de prueba generalmente comprende un tercio del conjunto de datos de derivación original y es probable que sea relativamente pequeña, lo que da como resultado estimaciones de rendimiento imprecisas^(73,74).

Los inconvenientes de esta técnica son, por un lado, que es ineficiente (ya que no utiliza todos los datos de la muestra para crear el modelo) y, por otro, que según hagamos las divisiones de la muestra original, obtendremos resultados diferentes (esto será más obvio cuanto menor sea la muestra).

Validación cruzada (cross validation)⁽⁷¹⁾

Con este método se divide la muestra original en dos partes iguales, se deriva el modelo en uno de ellos (training) y se valida en el otro (validation), después se repite el proceso intercambiando los grupos, el que era de validación pasa a ser el grupo de entrenamiento y viceversa. Así se emplean todos los datos en alguna ocasión para llevar a cabo la validación⁽⁶³⁾. El rendimiento se obtiene mediante el promedio de los dos rendimientos obtenidos en cada una de las validaciones.

Una extensión de este método es *k-folds cross validation*. En este caso, se divide la muestra original en k grupo iguales (folds). Se utilizan k-1 de los grupos para el desarrollo del modelo y el otro grupo para validación. El proceso se repite k veces, de manera que todos los grupos sean usados una vez para la validación. El rendimiento se obtiene como la media del rendimiento de los k rendimientos obtenidos.

2.1.7.2. Validación externa de un modelo predictivo.

La validación externa de un modelo predictivo consiste en validar el modelo derivado en un nuevo conjunto de pacientes.

Idealmente, un estudio de validación debería realizarse de forma prospectiva, mediante la inscripción de nuevos individuos en un estudio prediseñado específicamente, y la RPC debería aplicarse a todos los pacientes que cumplan los criterios de inclusión del estudio^(6,8). Sin embargo, los estudios de validación se suelen realizar de forma retrospectiva utilizando conjuntos de datos existentes.

Steyerberg establece dos tipos de validaciones externa: estrecha o amplia, pero Toll et al. (2008) plantea otra clasificación con 3 tipos de validaciones):

- Temporal: se emplean participantes de la misma institución, medidos de la misma manera, por los mismos investigadores, aunque en un momento posterior⁽⁷⁵⁾.
- Geográfica: se lleva a cabo en otras instituciones, hospitales, etc., realizada por diferentes investigadores, utilizando mediciones de predictores y variables de resultado diferentes en diferentes lugares y con pacientes con características diferentes según la ubicación.
- Dominio: se lleva a cabo con participantes muy diferentes a los que se emplearon para la derivación del modelo. Por ej: derivada en AP y validada en especializada o en diferentes grupos de edad (derivada en adultos y validada en niños).

Otros tipos de validación externa son: validación metodológica que se refiere a las pruebas que utilizan datos recopilados a través de diferentes métodos, la validación del espectro que se refiere a las pruebas en pacientes con diferente gravedad de la enfermedad o prevalencia del resultado de interés y validación totalmente independiente que se refiere a las pruebas realizadas por investigadores independientes en diferentes sitios^(31,61).

Para poder validar un modelo predictivo es necesario que se informen adecuadamente los métodos empleados y el manejo de los datos, así como todos los predictores candidatos y los que finalmente han sido incluidos en el modelo final. Es necesario que se reporten todos los coeficientes de regresión (incluida la constante) para permitir que otros investigadores puedan calcular las predicciones de riesgo para sus propios pacientes⁽²⁶⁾.

Hay que tener en cuenta que el rendimiento predictivo o la precisión de un modelo pueden verse afectados negativamente por elecciones metodológicas deficientes, debilidades en los datos o porque la muestra de pacientes de la cohorte de validación difiera significativamente de la muestra de la cohorte de desarrollo del modelo, o por otras causas, como son: que se hayan realizado diferentes mediciones de las variables, que se hayan realizado diferentes mediciones del evento de interés o que se hayan aplicado diferentes tratamientos⁽⁷⁶⁾. Por lo tanto, el requisito crítico de un modelo multivariable es la transportabilidad o la validez externa, es decir, la confirmación de que el modelo funciona como se esperaba en pacientes nuevos, pero similares, de otros lugares⁽²⁶⁾.

Es habitual que el rendimiento del modelo predictivo sea inferior en la muestra de validación externa que en la muestra que sirvió para el desarrollo de dicho modelo (incluso tras la validación interna y el ajuste)^(77,78). Existe la tentación de rechazar la RPC y derivar una completamente nueva en el conjunto de datos de validación, a menudo considerablemente más pequeño^(78,79). Este enfoque conduce a una pérdida de información científica que había sido capturada en el estudio de derivación y propiciará que haya varias RPC desarrolladas para la misma situación clínica, dejando a los médicos en un dilema sobre cuál usar^(65,78).

La alternativa recomendada para este caso es actualizar, ajustar o recalibrar la RPC, utilizando los datos de validación. De este modo, combinamos así la información capturada en la RPC original con la información de nuevos pacientes y mejoramos la generalización.

Métodos para actualizar, ajustar o recalibrar las RPC son⁽⁸⁰⁾:

- Si la prevalencia del outcome, en el estudio de validación, es diferente a la del estudio de derivación: hará que la calibración en la muestra de validación se vea afectada. Se puede mejorar ajustando el riesgo basal (la constante del modelo, denominada

intercept) del modelo original a los pacientes de la muestra de validación⁽⁷⁹⁾.

- Si la RPC está sobreajustada o mal ajustada: la calibración puede mejorarse ajustando simultáneamente todos los coeficientes de regresión.
- Si hay problemas de discriminación: se puede mejorar volviendo a estimar los coeficientes de regresión individuales o se pueden agregar predictores adicionales^(79,81).

Las RPC que se han actualizado, reajustado o recalibrado deben validarse externamente, al igual que las RPC cuando son derivados⁽⁷⁸⁾.

2.1.7.3. Utilidad clínica o impacto clínico.

Una vez que se ha desarrollado una RPC y se ha validado externamente de manera adecuada, dicha RPC permite predecir, de forma más o menos precisa, un resultado. Esto no justifica que se emplee la regla para adoptar decisiones clínicas. Hay que llevar a cabo, por lo tanto, un paso más, que consiste en transformar una RPC en una regla de decisión, es decir, evaluar cómo repercute en la atención de los pacientes el uso de la RPC. Este proceso se denomina *Utilidad Clínica o Impacto Clínico* de la RPC.

Reilly et al⁽¹⁶⁾, realizan una interesante revisión sobre el impacto clínico en las RPC e indican que esta última fase es superada por muy pocos modelos y que se debe llevar a cabo aplicando el modelo en la práctica clínica y evaluando su impacto en las medidas de resultados clínicos como: duración de la enfermedad, nº de complicaciones, supervivencia, costes, etc.

Los estudios de impacto clínico evalúan el efecto de usar un modelo para guiar la atención al paciente en comparación con no usar dicho modelo. Utilizan un diseño comparativo para estudiar el efecto del modelo en la toma de decisiones clínicas, los resultados de los pacientes o los costes de la atención. Existen tres diseños para alcanzar este objetivo: antes-después, presencia-ausencia y ensayo clínico aleatorizado (ECA). Sin lugar a dudas, el ECA es la mejor opción, ya que permite controlar los posibles sesgos.

Se debe esperar que la RPC confirme su capacidad en un ensayo clínico. El ECA debería evaluar si los pacientes tratados según lo que sugiere esa RPC se "curan mejor" que cuando no se usa la RPC (es decir, cuando se emplea la intuición o la experiencia de cada profesional). También se debe tener en cuenta si la regla es fácil de usar, si su empleo es coste-efectivo y si su aplicación no impone nuevos riesgos a los pacientes⁽⁸²⁾.

Hay diferentes aspectos que influyen en el impacto clínico, cabe destacar, el espectro de pacientes en el que se aplicará el modelo predictivo desarrollado. Si el modelo es menos eficiente en el grupo de pacientes que más incertidumbre generan, su rendimiento clínico se verá limitado. La valoración del impacto clínico va a requerir una ponderación de los costes y beneficios del modelo predictivo, tanto desde la perspectiva del médico, como de la del paciente, el sistema sanitario o la sociedad.

Existe una clara falta de orientación para el diseño, la realización y la presentación de informes de estudios de análisis de impacto de las RPC. Con este fin, Wallace et al⁽³⁰⁾ formularon un marco iterativo de cuatro fases:

- La fase 1: implica determinar si la RPC está lista para el análisis de impacto, es decir, si se ha derivado rigurosamente y se ha validado ampliamente de acuerdo con estándares metodológicos predefinidos.
- La fase 2: incluye evaluar la aceptabilidad de la RPC e identificar las barreras potenciales para su aceptación e implementación, así como evaluar la viabilidad de realizar un estudio de impacto.
- La Fase 3: consiste en probar formalmente el impacto de la RPC usando un diseño de estudio comparativo.
- La fase 4: implica la difusión e implementación a largo plazo de la RPC.

El impacto clínico de una RPC debe evaluarse en términos de *seguridad* (proporción de pacientes que tienen el outcome de interés y recibieron la intervención adecuada) y de *eficacia* (proporción de pacientes sin el outcome de interés y quienes no recibieron la intervención). La sensibilidad (Se) y la especificidad (Sp) de una RPC solo será la misma que su seguridad y eficacia si los médicos siguen exactamente la lógica y las recomendaciones de la RCP⁽¹⁶⁾.

2.1.8. Medidas del rendimiento de los modelos de predicción.

El rendimiento predictivo de un modelo se evalúa típicamente en términos de rendimiento general, calibración y discriminación y, a veces (sobre todo en estudios de modelos de diagnóstico) de clasificación⁽²⁹⁾.

2.1.8.1. Medidas de rendimiento respecto a las predicciones⁽⁸³⁾.

Rendimiento general

- *Brier score*: las medidas tradicionales para outcomes binarios y de tiempo hasta el evento incluyen el Brier score (similar al R^2 obtenido para outcome continuo^(18,84)).

La puntuación de Brier es una regla de puntuación cuadrática^(18,85), donde se calculan las diferencias cuadráticas entre los resultados binarios reales Y y las predicciones p : $(Y - p)^2$. Se puede escribir: $Y * (1 - p)^2 + (1 - Y) * p^2$.

La puntuación de Brier para un modelo puede oscilar entre 0 (modelo perfecto) y 0.25 (modelo no informativo) con una incidencia del outcome del 50%. Se puede escalar Brier por su puntaje máximo bajo un modelo no informativo: $Brier_{scaled} = 1 - (Brier / Brier_{max})$ donde $Brier_{max} = \text{mean}(p) * (1 - \text{mean}(p))$, para que oscile entre 0% y 100%.

- *R^2 ajustado*: es la medida de rendimiento general más habitual cuando se trabaja con outcome continuos.

Representa el porcentaje de varianza explicada por el modelo respecto del total de varianza observada. Toma valores entre 0 y 1. Cuanto mayor sea el valor, mejor será el modelo.

Calibración

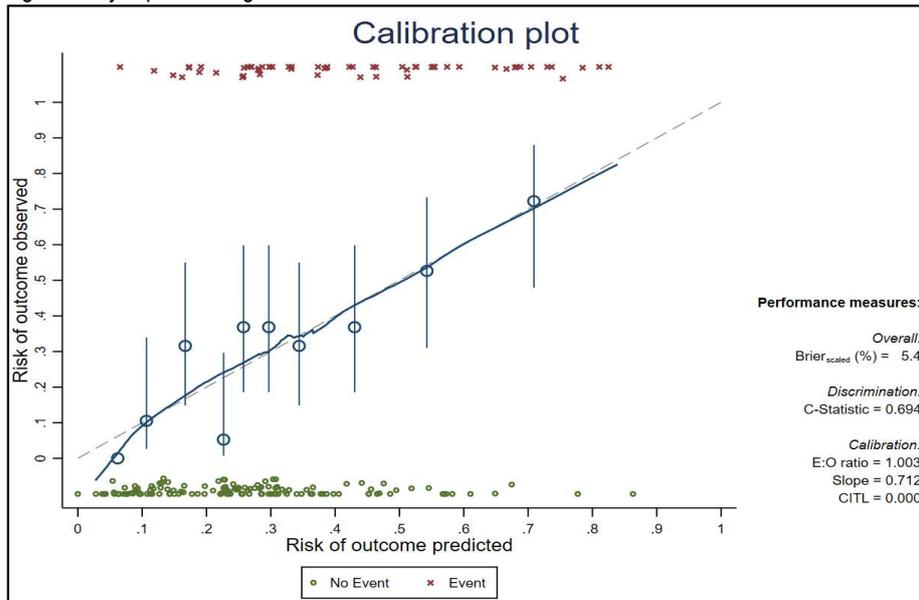
Para llevar a cabo la calibración, se pueden emplear diferentes técnicas:

- *Gráfico de calibración⁽⁸³⁾*: es la técnica más recomendada y permite evaluar, mediante un gráfico, cuán bien clasifica el modelo a los pacientes en las categorías de riesgo, es decir, analizar si las proporciones observadas de eventos y las probabilidades predichas

coinciden en todo el rango de probabilidades (un enfoque común es utilizar 7 o más grupos de riesgo de igual tamaño⁽⁸⁶⁾).

Si las proporciones observadas de eventos y las probabilidades predichas coinciden, en todo el rango de probabilidades, la gráfica muestra una línea de 45° (es decir, la pendiente es 1 y el punto de intersección es 0). Con este gráfico podemos ver la dirección y magnitud de la mala calibración, si la hubiera, a través del rango de probabilidades (Figura 1).

Figura 1. Ejemplo de un gráfico de calibración.



Creado a partir de los datos "lbw.dta" (Hosmer & Lemeshow data) de Stata.

- *Test de bondad de ajuste Hosmer-Lemeshow (Godness of Fit –GoF-)*⁽³²⁾: el test de Hosmer-Lemeshow compara los valores esperados por el modelo con los valores realmente observados. Ambas distribuciones, esperada y observada, se contrastan mediante una prueba de χ^2 . La hipótesis nula del test es que no hay diferencias entre los valores observados y los valores pronosticados (el rechazo a este test indicaría que el modelo no está bien ajustado).

El estadístico se puede calcular cuando la mayor parte de las frecuencias esperadas son superiores a 5 y ninguna de ellas es menor de 1.

Se recomienda acompañar el test de bondad de ajuste de Hosmer-Lemeshow (Goodness of Fit –GoF-) del gráfico de calibración, ya que el test tiene un poder limitado para evaluar una mala calibración. Puede carecer de poder estadístico para identificar el sobreajuste^(26,31) y es sensible al número de grupos y al tamaño de la muestra: el test suele ser no significativo para conjuntos de datos pequeños y, casi siempre, es significativo para conjuntos de datos grandes⁽⁸⁷⁾.

El inconveniente es, como en todos los test frecuentistas, que pequeñas diferencias serán estadísticamente significativas cuando trabajemos con tamaños muestrales grandes.

- *Pendiente de calibración (calibration slope)*: El predictor lineal viene definido como la suma en la escala logarítmica de los coeficientes de regresión (estimados en la muestra de desarrollo) multiplicado por el valor del individuo en el predictor correspondiente. La pendiente por definición es uno en la muestra de desarrollo.

En un conjunto de validación o en las muestras bootstrap, cuando aplicamos esta técnica, una calibración perfecta mostraría una pendiente de 1, pero generalmente debido al sobreajuste la pendiente de calibración es inferior. Esto indica que los coeficientes que fueron estimados en la muestra de desarrollo deben ser estrechados mediante *Shrinkage*. Es por eso, que la pendiente de calibración también es conocida como *Shrinkage factor* (factor de contracción)^(88,89).

La calibración debe llevarse a cabo tanto en el modelo de derivación como en el modelo de validación.

- La "calibración interna" se refiere a la concordancia entre las probabilidades pronosticadas y las frecuencias de los resultados observados en el conjunto de datos de derivación, donde una mala calibración puede indicar una falta de ajuste del modelo o una especificación incorrecta del modelo⁽⁹⁰⁾.
- La "calibración externa" se refiere a la concordancia entre las probabilidades pronosticadas y las frecuencias de los resultados observados en nuevos conjuntos de datos externos a aquel del que se derivó el modelo, donde una mala calibración puede indicar un modelo sobreajustado⁽⁹⁰⁾.

Discriminación⁽⁸³⁾

Es la capacidad que tiene el modelo para diferenciar entre pacientes que presentan o no la enfermedad o presentan o no un determinado evento. En nuestro caso, enfermos con COVID-19 vs no enfermos (para modelos diagnósticos) o éxitos o ingresos en UCI vs no éxitos o no ingreso en UCI (para modelos pronósticos).

Diremos que, un modelo discrimina perfectamente cuando la probabilidad predicha para todos los individuos que tienen el evento de interés sea mayor que la de todos los individuos que no la tienen.

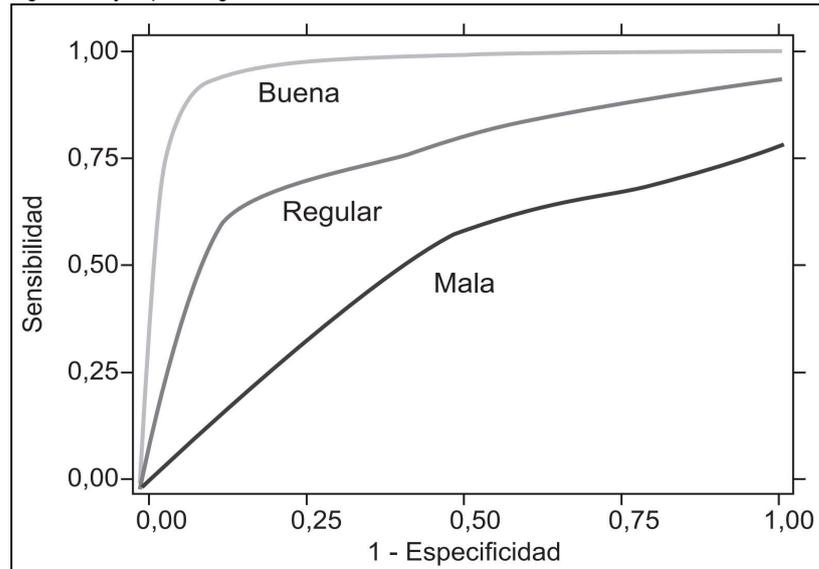
- *Curva ROC (receiver operating characteristic) o AUC (área under the curve):* La capacidad discriminativa de los modelos de predicción se evalúa calculando *la curva ROC* y obteniendo el AUC (Figura 2).

La curva ROC es una representación gráfica de la relación existente entre los verdaderos positivos (Se) y los falsos positivos (1-Sp) para los distintos puntos de corte. La Se es la probabilidad que tiene una prueba diagnóstica de dar resultados positivos en los sujetos enfermos (identificar enfermos entre los enfermos). La Sp es la probabilidad de dar resultados negativos entre los sujetos que no están enfermos (identificar no enfermos entre los no enfermos).

El AUC representa los diagnósticos correctos (verdaderos positivos y verdaderos negativos) y el área que queda por encima son los diagnósticos incorrectos (falsos positivos y falsos negativos). Cuanto más se aproxima la curva a la esquina superior-izquierda del gráfico, mayor será el AUC y mayor la validez de la RPC. Si el AUC vale 1, la discriminación es perfecta, y si vale 0.5 indica que el modelo discrimina igual que si tirásemos una moneda al aire para decidir si el individuo tendrá o no el evento de interés.

El AUC es el índice habitual en las publicaciones, pero tiene el inconveniente de tomar el valor máximo cuando se incluyen todos los predictores. Así mismo, facilita la elección de puntos de corte de los criterios diagnósticos de una prueba y permite comparar pruebas diagnósticas.

Figura 2. Ejemplo de gráfico de curva ROC con distintas AUC.



Fuente: Rev. Chilena de Cirugía. Vol 62 - Nº 3, junio 2010; pág. 301-308

La curva ROC tiene en cuenta todos los valores de corte consecutivos para definir un grupo de riesgo elevado frente a un grupo de riesgo bajo. Pero podemos estar interesados en seleccionar el punto de corte óptimo. Hay diversas maneras de determinar dicho valor⁽⁸³⁾:

- *Índice de Youden* (enfoque basado en los datos): el índice es máximo en el ángulo superior izquierdo de la curva ROC, por lo tanto, podríamos buscar el valor de corte de dicho punto. En ese punto se maximiza la sensibilidad y la especificidad. Se calcula haciendo: $(Se + Sp - 1)$.
- *Valor de corte de análisis de decisión* (enfoque basado en los análisis de decisión): se mide la utilidad o satisfacción relativa de la consecuencia de una clasificación verdadera o falsa (una clasificación falsa positiva implica un sobret ratamiento). Se calcula haciendo el cociente D/B , denominado *probabilidad* o *valor de corte*, donde D es el daño (mala clasificación) y B es el beneficio (buena clasificación).

Definir un punto de corte alto dará como resultado una buena especificidad y pocos falsos positivos, pero poca sensibilidad y muchos falsos negativos. Una prueba con una alta especificidad es útil para confirmar una enfermedad si una persona da positivo. Esto se debe a que rara vez diagnostica erróneamente, como enfermos, a quienes no tienen la afección de interés⁽⁹¹⁾.

Definir un punto de corte bajo dará como resultado una buena sensibilidad y pocos falsos negativos, pero una baja especificidad y muchos falsos positivos. Una prueba con una alta sensibilidad es útil para descartar una enfermedad si una persona da negativo. Esto se debe a que rara vez diagnostica erróneamente, como sanos, a quienes padecen la afección de interés⁽⁹¹⁾.

- *Índice de concordancia de Harrell (C-Statistic o C-index):* Se calcula tomando todos los posibles pares de sujetos donde uno experimenta el evento y el otro no. Dicho estadístico representa la proporción de todos los pares en los cuales la probabilidad predicha para el individuo que experimentó el evento es mayor que la predicha para el individuo que no experimentó el evento. El valor del C-Statistic para modelos pronósticos varía, generalmente, entre 0.6 y 0.85 aunque para modelos diagnósticos dichos valores pueden ser más altos⁽⁶⁴⁾.

Se establece que un modelo predictivo es bueno cuando el $AUC \geq 0,8$ y se considera que dicho modelo es excelente cuando $AUC \geq 0,9$ ⁽⁹²⁾. Estos mismos valores son aplicables para el C-Statistic, ya que, en regresión logística, este estadístico es equivalente al AUC⁽⁹³⁾.

- *D de Somers:* su interpretación es similar a la del AUC, pero toma valores entre -1 y 1. Se calcula haciendo: $2 \times (AUC - 0.5)$. Cuanto más próximo esté de la unidad, mejor discrimina el modelo. Este índice tiene la ventaja de no verse afectado por la prevalencia del evento.

2.1.8.2. Medidas de rendimiento respecto a las clasificaciones⁽⁸³⁾.

Hemos vistos una serie de medidas (en calibración y discriminación) para evaluar el rendimiento de las predicciones, pero podemos estar interesados en evaluar la calidad de las clasificaciones. Algunas de las medidas de rendimiento que nos permiten esta evaluación son:

- *Curvas ROC utilizando un solo valor de corte:* en lugar de usar todos los puntos de corte en las curvas ROC, podemos construir curvas ROC, pero utilizando un solo valor de corte basado en los datos o basado en un análisis de decisión.
- *Reclasificación:* podemos medir el valor incremental de un nuevo marcador valorando los cambios que se producen en la clasificación del riesgo cuando se consideran las probabilidades predichas del marcador en el modelo predictivo⁽⁹⁴⁾.

Cook propone hacer una tabla para mostrar cuántos sujetos son reclasificados al añadir un nuevo predictor al modelo. Pero dicho cambio no es suficiente para valorar la mejoría en la estratificación del riesgo, debe ser adecuado. Para ello, Cook⁽⁹⁵⁾ propone aplicar el test de reclasificación para llevar a cabo la evaluación de la reclasificación.

- *Índice de reclasificación neta (NRI):* Pencina⁽⁹⁶⁾ propone no tener en cuenta la reclasificación en todos los pacientes, sino centrarnos en la reclasificación en el sentido correcto, es decir, una clasificación en un riesgo superior en los pacientes con el outcome y en un riesgo inferior en los pacientes sin el outcome. De manera que, cualquier movimiento "hacia arriba" en las categorías de sujetos con el resultado implica una clasificación mejorada y cualquier "movimiento hacia abajo" indica una reclasificación peor. La interpretación es opuesta para los sujetos sin el resultado.

Para ello, compara la proporción de predicciones correctas con el modelo B, una vez descontadas las incorrectas y le resta la proporción de predicciones correctas con modelo A, una vez descontadas las incorrectas.

$$NRI = \sum_E \frac{V}{N_E} - \sum_{NE} \frac{V}{N_{NE}}$$

donde E significa tener el evento, NE significa no tener el evento, N es el tamaño del grupo y V predicciones correctas menos las incorrectas.

Muchos investigadores advierten contra el uso de esta métrica, ya que los valores positivos pueden surgir simplemente debido a modelos mal ajustados^(7,97-101). Por tanto, no se puede recomendar la estadística NRI de manera generalizada⁽¹⁰²⁾.

- *Beneficio Neto (BN)*^(103,104): que expresa el valor relativo de los beneficios y daños asociados con el uso del modelo para determinar las decisiones clínicas. Beneficio Neto es: $BN = (VP - w FP) / N$ donde VP es el número de clasificaciones positivas verdaderas, FP es el número de clasificaciones falsas positivas, N es el número total de individuos y w es igual al cociente D/B donde D es el daño y B el beneficio.

Si $w=1$, FP y VP se ponderan por igual, esto implica una probabilidad de 1:1 para el cociente D/B , que representa que consideramos a los falsos positivos iguales que a los positivos verdaderos.

- *Análisis de curvas de decisión (ACD)*: el valor de corte para la aplicación clínica de un modelo de predicción no suele definirse de manera precisa. La ponderación relativa de daños y beneficios puede no ser conocida, por dicho motivo Vickers y Elkin⁽¹⁰⁵⁾ proponen dar una gama de valores de corte y calcular el BN para dichos valores de corte. El resultado obtenido se representa en una curva denominada *curva de decisión* (Figura 3).

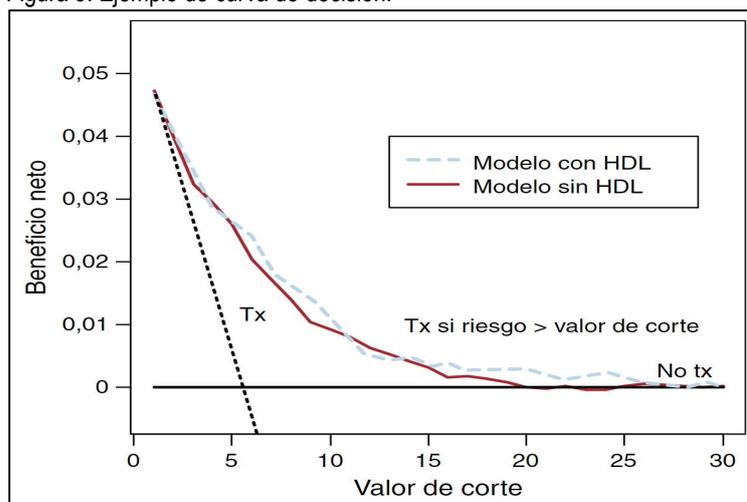
Este método incorpora el coste de clasificación errónea y, por lo tanto, indican la utilidad clínica de las RCP⁽¹⁰⁶⁾. Y El beneficio neto permite determinar si basar las decisiones clínicas en un modelo sería más beneficioso que perjudicial⁽¹⁰⁴⁾. Esto contrasta con las medidas tradicionales como la sensibilidad, la especificidad o el AUC, que son abstracciones estadísticas que no informan directamente sobre el valor clínico.

No se ha podido especificar, cuán alta debe ser la sensibilidad, la especificidad o el AUC para justificar el uso clínico de un determinado modelo. Tampoco qué grado de error de calibración indicaría que no se debe utilizar un modelo de predicción, o qué modelo se debe elegir, entre dos modelos dados, cuando uno presenta mejor calibración y otro mejor discriminación⁽¹⁰⁴⁾.

El análisis de decisiones intenta abordar estos problemas incorporando las consecuencias clínicas del uso de un modelo, de una prueba o de un marcador. Se calcula el BN especificando un tipo de cambio entre diferentes criterios de valoración médicos. De modo que se puede obtener del siguiente modo: $\text{Beneficio neto} = \text{Beneficio} - (\text{Daño} \times \text{Tasa de cambio})$ donde la tasa de cambio es el n° de pacientes a los que se les realiza una prueba para prevenir 1 evento del outcome que se quiere medir (por ej: n° de hombres que está dispuesto a biopsiar el médico para prevenir un cáncer de próstata).

Como resumen, Steyerberg et al⁽¹⁸⁾, sugieren que se reporten siempre medidas de discriminación y de calibración en los estudios de modelos predictivos, tanto de derivación como de validación. Y establecen que se deberían reportar las curvas de decisión para aquellos modelos predictivos que vayan a ser usados en la toma de decisiones clínicas.

Figura 3. Ejemplo de curva de decisión.



Fuente: Rev Esp Cardiol. 2011;64(9):788-794

Mostramos un resumen de las medidas de rendimiento explicados anteriormente y cuáles de ellas son empleadas para medir la utilidad clínica (Tabla 2).

Tabla 2. Algunas medidas de rendimiento de los modelos de predicción.

Tipo de evaluación	Medidas	Características
Evaluación de las predicciones Discriminación	AUC o C-Statistic	Probabilidad de clasificación correcta para un par de pacientes con y sin el resultado evaluado.
	Calibración Valor de la intersección (a/b=1) Pendiente de recalibración (b)	Calibración general Media del efecto de los factores predictivos en el resultado
Evaluación de las clasificaciones Clasificación	Índice de Youden	(Se + Sp - 1)
	Utilidad clínica BN y DCA	Fracción neta de los VP ganados mediante la toma de decisiones basadas en las predicciones para un único umbral (BN) o para una gama de umbrales (DCA)
Evaluación del valor incremental de un marcador Reclasificación	NRI	Fracción neta de reclasificaciones en el sentido correcto mediante la toma de decisiones basadas en las predicciones realizadas con un marcador en comparación con las decisiones tomadas sin el marcador
	Utilidad clínica Diferencia en BN y DCA	Fracción neta de VP ganados con la toma de decisiones basadas en predicciones realizadas con un marcador en comparación con las decisiones tomadas sin el marcador para un único umbral (BN) o en una gama de umbrales (DCA)
	NRI ponderado	Ponderaciones según las consecuencias de las decisiones (BN y NRI ponderado)

Fuente: Rev Esp Cardiol. 2011; 64(9):788-794

2.1.9. Comparación de modelos predictivos.

Para comparar modelos predictivos, primero, se deben tener en cuenta dos factores⁽¹⁰⁷⁾:

1. El diseño del estudio:

- *Diseño de datos pareados*: cuando se construyen dos o más curvas empíricas sobre la base de pruebas realizadas en los mismos individuos, el análisis estadístico de las diferencias entre curvas debe tener en cuenta la naturaleza correlacionada de los datos.
- *Diseño de grupos independientes*: si las curvas se han realizado sobre pruebas aplicadas a diferentes individuos. (Ej: si se comparan dos pruebas diferentes que han sido creadas para lo mismo).

2. La distribución del outcome:

- Enfoque paramétrico^(108,109) (o semiparamétrico): asumimos una distribución paramétrica (binormal) para el outcome tanto del grupo de pacientes con la condición estudiada como para el grupo sin dicha condición.
- Enfoque no paramétrico^(110,111): no se hacen supuestos de distribución de los outcomes para los individuos con la condición y sin la condición.

En base a esto, se pueden aplicar diferentes tests estadísticos para calcular el error estándar del AUC y realizar las comparaciones de las AUC:

1. Paramétricos:

- Método binormal: fue desarrollado por Dorfman y Alf Jr. (1969), McClish (1989) e implementado y desarrollado por Metz et al (1998).

2. No paramétricos:

- DeLong⁽¹¹¹⁾: es el método más empleado y suele ser el recomendado ya que es un método robusto. Se basa en la teoría sobre el estadístico

U generalizado para generar una matriz de covarianza estimada. Incorpora las dependencias causadas por el hecho de que los modelos generalmente se generan en los mismos pacientes y, por lo tanto, están "emparejados".

- Test bivalente para comparar dos AUC, con ajuste de Bonferroni para comparaciones múltiples de pares de AUC.
- Hanley & McNeil⁽¹¹²⁾ (aproximación exponencial): desarrollaron una aproximación exponencial del error estándar del AUC basada en los componentes originales de la variabilidad muestral de las estadísticas de Wilcoxon. Tiene un atractivo obvio cuando los resultados de las pruebas se registran en una escala continua y se puede aplicar incluso para tamaños de muestra pequeños.
- Propuesta de test χ^2 para comparación de 2 o más AUC correlacionadas⁽¹¹³⁾: es fácil de calcular, de entender y de aplicar. Si se rechaza la hipótesis nula de igualdad de las AUC, es necesario realizar más análisis para determinar cuál de las AUC o prueba de diagnóstico ha contribuido al rechazo de la hipótesis nula. Esto no es posible cuando se usa el método de DeLong, pero sí con éste.

En general, para comparar 2 o más modelos de predicción, se suele emplear la diferencia entre los valores AUC obtenidos para diferentes modelos, o el índice C-Statistic equivalente⁽¹¹¹⁾. Sin embargo, este enfoque es inapropiado en el caso de modelos anidados que se ajustan al mismo conjunto de datos⁽¹¹⁴⁾.

Por lo tanto, para seleccionar el mejor modelo, se deben evaluar las ventajas e inconvenientes de los predictores incluidos en cada uno de los mejores modelos:

- Importancia (conceptual y clínica) de los predictores.
- Coherencia teórica del signo de los coeficientes del modelo.
- Calidad (fiabilidad y precisión) de las medidas de dichas variables.
- Las variables objetivas serán preferibles a las subjetivas.
- Bajo Coste.
- Simplicidad, estandarización, reproducibilidad y disponibilidad de los predictores⁽¹¹⁵⁾.

2.1.10. Conclusiones.

Podemos concluir que, para que un modelo predictivo se considere útil:

- La puntuación de riesgo debe ser clínicamente creíble y precisa (bien calibrada y con buena capacidad discriminativa).
- Reproducible y generalizable a otros pacientes (validarse externamente).
- Demostrar eficacia clínica (es decir, proporcionar información adicional y útil a los médicos, que mejore la toma de decisiones y, por tanto, los resultados del paciente⁽⁹⁾).
- Que no imponga nuevos riesgos a los pacientes.

Los modelos predictivos pueden ser una función compleja de los predictores, pero se recomienda, de cara a la validación y de cara a la implementación en la práctica clínica, que el modelo sea sencillo (principio de parsimonia), fácil de usar (que los predictores de la RPC sean accesibles) e interpretable.

2.2. EVALUACIÓN DE RIESGO DE SESGOS EN MODELOS DE PREDICCIÓN.

El objetivo de una RPC es maximizar la precisión predictiva y la utilidad clínica, para ello es vital que estas RPC se desarrollen, validen y evalúen rigurosamente. Sin embargo, numerosas RS han demostrado deficiencias en la calidad metodológica y la presentación de informes de los estudios de predicción, lo que limita la utilidad de la RPC en la práctica⁽¹¹⁶⁻¹¹⁸⁾.

Es importante evaluar los posibles sesgos en los modelos de predicción para poder valorar si la RPC es fiable y si se puede aplicar en la práctica clínica. El sesgo ocurre cuando los resultados de un estudio no representan la verdad debido a limitaciones inherentes en el diseño o en la realización del estudio⁽¹¹⁹⁾.

Se define el sesgo como la presencia de un error sistemático en un estudio que conduce a resultados distorsionados o defectuosos y obstaculiza la validez interna del estudio⁽⁸⁷⁾.

En el desarrollo y validación de un modelo de predicción, existen características conocidas que hacen que un estudio presente riesgo de sesgos (Risk of Bias –RoB-), pero hay evidencia empírica limitada que muestre las fuentes más importantes de sesgo. Una manera de detectar el sesgo en las estimaciones del rendimiento predictivo de un modelo, puede ser pensar en cómo se habría diseñado, realizado y analizado un estudio de modelo de predicción hipotético que fuese metodológicamente sólido.

Se indica que hay RoB cuando las deficiencias en el diseño, la realización o el análisis del estudio podrían llevar a estimaciones sistemáticamente distorsionadas del rendimiento predictivo de un modelo⁽⁸⁷⁾.

Existen diferentes herramientas para evaluar el RoB y la aplicabilidad de los estudios de modelos de predicción de diagnóstico y pronóstico. Vamos a describir algunas de ellas:

2.2.1. PROBAST (Prediction model risk of bias assessment tool).

PROBAST es la primera herramienta diseñada rigurosa y específicamente para evaluar el RoB y la aplicabilidad de estudios que desarrollan, validan o actualizan modelos de predicción diagnóstico o pronóstico para predicciones individualizadas. Y es también la más completa de las que se encuentran disponibles actualmente⁽⁸⁷⁾.

Aunque PROBAST fue diseñado para la evaluación de estudios primarios incluidos en una RS, puede usarse de manera más general en la evaluación crítica de estudios de modelos de predicción y para actualizar o ampliar modelos predictivos cuando estemos interesados en añadir un nuevo predictor a un modelo existente.

La herramienta PROBAST fue desarrollada por un grupo directivo a través de técnicas de consenso, en un procedimiento Delphi, en el que participaron 38 expertos en la materia y fue perfeccionada mediante pruebas piloto. Para la creación de PROBAST se tuvieron en cuenta las herramientas de RoB y las pautas de presentación de informes existentes. Se estableció una estructura similar a la utilizada en otras herramientas RoB: ROB 2.0⁽¹²⁰⁾, ROBINS-I⁽¹²¹⁾, QUADAS-2 (evaluación de la calidad de los estudios de precisión diagnóstica 2)⁽¹²²⁾ y ROBIS⁽¹²³⁾. Y, finalmente, se definieron 4 dominios: participantes, predictores, resultado y análisis, con un total de 20 preguntas relacionadas con el diseño, la realización o el análisis del estudio⁽¹²⁴⁾.

PROBAST evalúa posibles fuentes de sesgos y problemas de aplicabilidad relacionadas con⁽⁸⁷⁾: las fuentes de datos utilizadas y cómo se seleccionaron los individuos para la participación en el estudio (*Dominio 1*); la definición y medición de los predictores. Considerando que los predictores son variables evaluadas por su asociación con el resultado de interés y se combinan para formar el modelo de predicción (*Dominio 2*); la definición y medición del resultado (*Dominio 3*) y las consideraciones estadísticas clave que son evaluadas para detectar si se abordaron correctamente (*Dominio 4*).

Mostramos a continuación (Tabla 3) un resumen de los 4 dominios y sus correspondientes preguntas así como posibles problemas respecto a la aplicabilidad (doi.org/10.7326/M18-1376)

Tabla 3. PROBAST: Resumen de los 4 dominios para evaluación de riesgo y aplicabilidad.

1. Participantes	2. Predictores	3. Outcome	4. Análisis
Preguntas de señalización			
1.1. ¿Fueron apropiadas las fuentes de datos empleadas, por ej: cohortes, ECA, o datos de estudios de casos y controles anidados?	2.1. ¿Fueron los predictores definidos y evaluados de manera similar en todos los participantes?	3.1. ¿Se determinó adecuadamente el resultado?	4.1. ¿Hubo un número razonable de participantes con el outcome?
1.2. ¿Fueron todas las inclusiones y exclusiones de pacientes apropiadas?	2.2. ¿Las evaluaciones de los predictores fueron hechas sin conocer el outcome?	3.2. ¿Se usó una definición del outcome preespecificada o estándar?	4.2. ¿Se manejaron los predictores continuos y categóricos de manera adecuada?
		3.3. ¿Fueron los predictores excluidos de la definición del outcome?	4.3. ¿Fueron todos los participantes reclutados incluidos en el análisis?
		3.4. ¿El outcome fue determinado y definido de manera similar para todos los participantes?	4.4. ¿Se manejaron adecuadamente los missing data?
		3.5. ¿Fue el outcome determinado sin conocimiento de información de los predictores?	4.5. ¿Se evitó que la selección de predictores se realizara basándose en análisis univariante?
		3.6. ¿Fue el intervalo de tiempo entre la evaluación de los predictores y la determinación del outcome apropiada?	4.6. ¿Se contabilizaron adecuadamente las complejidades de los datos (por ejemplo, censura, riesgo competitivo, muestreo de participantes de control)?
			4.7. ¿Se evaluaron adecuadamente las medidas de rendimiento relevantes del modelo?
			4.8. ¿Se tuvieron en cuenta el sobreajuste, el desajuste y el optimismo del modelo en el rendimiento del modelo?
			4.9. ¿Los predictores y sus pesos asignados en el modelo final se corresponden con los resultados del análisis multivariable reportado?
ROB			
Selección de participantes	Evaluación de sus predictores	Outcome y su determinación	Análisis
Aplicabilidad			
Los participantes o el entorno incluidos no coinciden con la pregunta de revisión	La definición, evaluación o tiempo de los predictores no coincide con las preguntas de revisión	Su definición, el momento o determinación no coincide con la pregunta de revisión	

Respuestas para las cuestiones: sí, probablemente sí, no, probablemente no y no información.

Respuestas para RoB y Aplicabilidad: bajo, alto o incierto (?)

2.2.2. CASPe (*Critical Appraisal Skills Programme Español*)

Existen otras herramientas que permiten realizar una evaluación crítica de los estudios, aunque no permiten evaluar el RoB tal y como lo hace la herramienta PROBAST. Dichas herramientas son las plantillas creadas por CASPe⁽¹²⁵⁾.

CASPe es una organización sin ánimo de lucro creada en 1998 y asociada a CASP (1992). Se organiza en red con múltiples nodos distribuidos por todo el territorio nacional y una sede coordinadora en Alicante.

Se basa en la colaboración entre personas (en su mayoría clínicos o epidemiólogos clínicos), es independiente y esa independencia le permite criticar la evidencia sin conflictos de interés.

Las tareas fundamentales que se desarrollan en CASPe son las de:

- Enseñar "lectura crítica de la evidencia clínica". Está orientada a personas con diferentes perfiles (*clínicos, investigadores, gestores, y pacientes*) de instituciones u organizaciones del sistema de salud, mediante talleres presenciales o/y virtuales. Asimismo, se enseñan habilidades para la docencia de la lectura crítica en sus entrenamientos de entrenadores.
- Producir materiales docentes específicos: libros, instrumentos de lectura (plantillas), calculadoras, Apps, etc.
- Desarrollar investigación sobre diferentes áreas.

La red CASP se creó para enseñar a evaluar la bondad de las evidencias publicadas, ya que no parece sensato creer ciegamente todo lo que se publica. Parece más adecuado, hacer una lectura crítica de esa evidencia publicada. Para ello, han desarrollado diferentes plantillas, calculadoras y un glosario de términos en función del tipo de estudio llevado a cabo,

A continuación, mostramos la primera página de la plantilla empleada para leer críticamente la evidencia clínica publicada sobre una RPC (Figura 4). En dicha plantilla se indica qué preguntas debemos hacernos a la hora de valorar si una RPC ha sido o no bien desarrollada o validada. Se puede descargar dicha plantilla de manera gratuita en la siguiente dirección: <https://www.redcaspe.org>.

Figura 4. Plantillas CASPe para realizar lectura crítica de artículos de Reglas de Predicción Clínica.



PROGRAMA DE LECTURA CRÍTICA CASPe
Leyendo críticamente la evidencia clínica

11 preguntas para entender las Reglas de Predicción Clínica

Comentarios generales

- Hay tres aspectos generales a tener en cuenta cuando se hace la lectura crítica de un artículo:
 - ¿Son válidos sus resultados?*
 - ¿Cuáles son los resultados?*
 - ¿Son aplicables a tu medio?*

Las 11 preguntas de las próximas páginas están diseñadas para ayudarte a pensar sistemáticamente sobre estos aspectos.

- Las tres primeras preguntas son “de eliminación” y se pueden responder rápidamente. Sólo si la respuesta a estas tres preguntas es afirmativa, entonces merecerá la pena continuar con las restantes.
- Puede haber cierto grado de solapamiento entre algunas de las preguntas. En la mayoría de ellas se te pide que respondas “sí”, “no” o “no sé”.
- En *itálica* y debajo de las preguntas encontrarás una serie de pistas para contestarlas. Están pensadas para recordarte por qué la pregunta es importante. ¡En los pequeños grupos no suele haber tiempo para responder a todo con detalle!

Estas 11 preguntas están inspiradas en:

- *McGinn TG, Guyatt GH, Wyer PC, Naylor CD, Stiell IG, Richardson WS. Users' guides to the medical literature: XXII: how to use articles about clinical decision rules. Evidence-Based Medicine Working Group. JAMA 2000; 284:79-84.*

El marco conceptual necesario para la interpretación y el uso de estos instrumentos puede encontrarse en la referencia de abajo o/y puede aprenderse en los talleres de CASPe:

Juan B Cabello por CASPe. Lectura crítica de la evidencia clínica. Barcelona: Elsevier; 2015. (ISBN 978-84-9022-447-2)

Juan B Cabello por CASPe. Lectura crítica de la evidencia clínica. Barcelona: Elsevier; 2015. (ISBN 978-84-9022-447-2).

2.2.3. JAMA Evidence (Using Evidence to Improve Care).

Existen guías para el usuario que pueden ayudar a obtener conocimiento sobre RPC, y de ese modo identificar posibles sesgos, por ejemplo: la proporcionada por la revista JAMA (Journal of the American Medical Association), que en 1992, inició la serie de artículos «Users' Guides to the Medical Literature», cuya iniciativa fue liderada por el Dr. Guyatt y epidemiólogos clínicos de universidades norteamericanas y europeas. Desde entonces hasta ahora, se han ido ampliando los temas tratados y se ha desarrollado y elaborado material para facilitar la aplicación de las guías. Este material corresponde a la 3ª edición y se denomina *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice*.

Ofrecen herramientas de aprendizaje (learning tools) y calculadoras, así como material adicional para las guías de usuario. Entre la información proporcionada se trata el tema de las RPC, en el capítulo 19.4 (Tabla 4).

Tabla 4. JAMA Evidence. Users' Guides to the Medical Literature.

+ The Foundations
+ Therapy
+ Harm (Observational Studies)
– Diagnosis
Chapter 16: The Process of Diagnosis
Chapter 17: Differential Diagnosis
Chapter 18: Diagnostic Tests
Chapter 18.1: How to Read Articles That Use Machine Learning
Chapter 19.1: Spectrum Bias
Chapter 19.2: Examples of Likelihood Ratios
Chapter 19.3: Measuring Agreement Beyond Chance
Chapter 19.4: Clinical Prediction Rules
– Prognosis
Chapter 20: Prognosis
Chapter 21.1: How to Use an Article About Genetic Association
Chapter 21.2: Discrimination and Calibration of Clinical Prediction Models
+ Summarizing the Evidence
+ Moving From Evidence to Action

2.3. REPORTAR ESTUDIOS DE MODELOS DE PREDICCIÓN (TRIPOD).

Es fundamental llevar a cabo un reporte adecuado de los estudios que se desean publicar, por ello, la guía TRIPOD (transparent reporting of multivariable prediction model for individual prognosis or diagnosis) establece un conjunto de recomendaciones, basadas en la evidencia, para informar de forma estándar los estudios de modelos de predicción en ciencias biomédicas. En el documento de explicación y elaboración de TRIPOD se describen aspectos de buenas prácticas para tales estudios, y también se destacan algunos enfoques inapropiados que deben evitarse⁽²⁹⁾.

Estas recomendaciones han sido elaboradas dentro de la Red EQUATOR⁽¹²⁶⁾, (Enhancing the QUALity and Transparency Of health Research) que trabaja para mejorar la confiabilidad y el valor de la literatura en investigaciones médicas estimulando informes completos, precisos, transparentes y exactos de las investigaciones en salud. La Red EQUATOR está dirigida por un grupo directivo internacional que reúne a los principales expertos en metodología de investigación en salud, estadísticas, informes y trabajo editorial, en una iniciativa global que cubre todas las áreas de la investigación en salud y todas las naciones, e involucra activamente a todas las partes interesadas.

TRIPOD es una herramienta simple y estructurada que Incluye 22 ítems (Tabla 5) que deben ser explicados, con el objetivo de mejorar el reporte en los estudios de desarrollo, validación o actualización de modelos de predicción, ya sean con fines diagnósticos o pronósticos y que dicho reporte sea lo más transparente posible. TRIPOD no pretende implantar una estructura estandarizada de informes, sino que los autores se aseguren de abordar todos los elementos de la lista de verificación en algún lugar de su artículo con suficiente detalle y claridad.

La guía ha sido desarrollada para apoyar a los autores en la redacción de informes de modelos de predicción, ayudar a los editores y revisores (peer review) a revisar este tipo de manuscritos enviados para su publicación y ayudar a los lectores a evaluar críticamente los informes publicados⁽¹²⁷⁾.

Los documentos oficiales de TRIPOD se encuentran en <https://www.tripod-statement.org/resources/>.

Tabla 5. Checklist TRIPOD: Desarrollo y Validación de un modelo de predicción.

Section/Topic	Item	Checklist Item	Page
Title and abstract			
Title	1	D:V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.
Abstract	2	D:V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.
Introduction			
Background and objectives	3a	D:V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.
	3b	D:V	Specify the objectives, including whether the study describes the development or validation of the model or both.
Methods			
Source of data	4a	D:V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.
	4b	D:V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.
Participants	5a	D:V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.
	5b	D:V	Describe eligibility criteria for participants.
	5c	D:V	Give details of treatments received, if relevant.
Outcome	6a	D:V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.
	6b	D:V	Report any actions to blind assessment of the outcome to be predicted.
Predictors	7a	D:V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.
	7b	D:V	Report any actions to blind assessment of predictors for the outcome and other predictors.
Sample size	8	D:V	Explain how the study size was arrived at.
Missing data	9	D:V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.
	10c	V	For validation, describe how the predictions were calculated.
	10d	D:V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.
Risk groups	11	D:V	Provide details on how risk groups were created, if done.
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.
Results			
Participants	13a	D:V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.
	13b	D:V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).
Model development	14a	D	Specify the number of participants and outcome events in each analysis.
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.
Model specification	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).
	15b	D	Explain how to use the prediction model.
Model performance	16	D:V	Report performance measures (with CIs) for the prediction model.
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).
Discussion			
Limitations	18	D:V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.
	19b	D:V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.
Implications	20	D:V	Discuss the potential clinical use of the model and implications for future research.
Other information			
Supplementary information	21	D:V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.
Funding	22	D:V	Give the source of funding and the role of the funders for the present study.

Fuente: www.tripod-statement.org

2.4. MODELOS PREDICTIVOS CREADOS PARA LA COVID-19.

Desde que comenzó la pandemia hasta el momento actual, se han realizado una cantidad ingente de publicaciones relacionadas con modelos predictivos para la infección por SARS-CoV-2, tanto de diagnóstico como de pronóstico. Debido al gran volumen de publicaciones, resultaba difícil poder evaluar metodológicamente y críticamente todos los estudios publicados, pero ha habido un equipo de profesionales que ha llevado a cabo esta tarea, es el grupo COVID-PRECISE (Precise Risk Estimation to optimise covid-19 Care for Infected or Suspected patients in diverse settings)

El grupo COVID-PRECISE⁽¹²⁸⁾ que trabaja en colaboración con el grupo Cochrane de Métodos Pronósticos, ha tenido y sigue teniendo un papel destacado en lo referente a investigaciones sobre modelos de predicción relacionados con la COVID-19. Es un gran consorcio formado por un amplio número de expertos a nivel internacional, y desde que comenzó la pandemia, ha estado trabajando para revisar sistemáticamente y evaluar críticamente todos los modelos de predicción disponibles actualmente para la COVID-19. Han ido realizando publicaciones de una *Living Systematic Review*⁽¹²⁹⁾, que fue aceptada para su publicación el 31 de marzo de 2020 con el título "Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal" y, a fecha de 31 de diciembre de 2021 se ha actualizado en 2 ocasiones: el 4 de mayo de 2020 y el 12 de enero de 2021 (última versión).

En dicha RS se han analizado aquellos modelos que permiten predecir el riesgo de infección o la admisión en el hospital por la enfermedad, modelos para predecir la presencia de infección por SARS-CoV-2 en pacientes con sospecha de infección, y modelos para predecir el pronóstico o el curso de la infección en los pacientes con COVID-19. En este proyecto también se incluyen modelos y estudios de validación externa.

Los autores (Wynants et al⁽¹³⁰⁾), identifican 169 estudios que describen 232 modelos de predicción. 7 modelos para identificar personas a riesgo en la población general. 118 modelos diagnósticos (de los cuáles, 75 se basan en pruebas de imagen y 10 se diseñan para diagnosticar severidad de la enfermedad) y 107 modelos pronósticos para predecir diferentes riesgos: de mortalidad (39 estudios), progresión de la enfermedad o ingreso en UCI (28 estudios), ventilación, intubación o estancia en el hospital (medida en días de ingreso).

De los 232 modelos de predicción que cumplieron los criterios de inclusión, 208 eran modelos nuevos, creados para COVID-19, y los 24 restantes fueron validaciones externas de modelos existentes. Algunos estudios investigan múltiples modelos y algunos modelos fueron validados en más de un estudio. 72 modelos (35%) son válidos como herramienta para el uso en la práctica clínica. 22 fueron presentados mediante una calculadora web, 12 como un score de puntuación, 11 modelos como nomogramas, 8 en un software, 5 mediante árboles de decisión, 3 como puntuación de un gráfico y los 6 restantes en otros formatos utilizables.

Los autores evalúan el RoB de los modelos incluidos en la RS mediante la herramienta PROBAST (herramienta para evaluar la calidad, el riesgo de sesgo y aplicabilidad de los modelos predictivos) e indican que todos los modelos analizados tienen alto riesgo de sesgos, debido a que: los controles no eran representativos de la población objetivo, el test empleado para determinar el resultado no fue el mismo entre participantes, los predictores no fueron claramente definidos o fueron indirectos y/o subjetivos, los estudios estaban mal informados -fueron opacos o ambiguos- y a que el 94% de los estudios tuvieron un alto riesgo de sesgo en el dominio de análisis. Por todo ello, los autores indican que las predicciones podrían ser poco fiables y que el desempeño predictivo de los modelos, cuando se usen en la práctica clínica, probablemente será menor que el reportado.

Así mismo, se concluye que hay 2 modelos que, en general, son de buena calidad (están basados en grandes conjuntos de datos y tienen un bajo riesgo de sesgo en la mayoría de los dominios, pero con una calificación general de riesgo de sesgo poco clara, debido al dominio de análisis). De los dos modelos, uno de ellos es para diagnóstico y, el otro, para pronóstico.

- El modelo diagnóstico propuesto por Jehi et al^(131,132) se construyó en Ohio (Florida) empleando un modelo logístico multivariable para predecir el resultado de la prueba de hisopado nasofaríngeo COVID-19 en función de los datos de sexo, edad, raza y características socioeconómicas y sugieren un papel modificador potencial de ciertas vacunas y medicamentos comunes.

Se consideró una muestra de 11.672 pacientes para la creación del modelo y una muestra de 2.295 pacientes para la validación externa del mismo.

El modelo ofrece buenos resultados tanto en la cohorte de desarrollo con un C-Statistic corregido por bootstrap (1000 remuestreos) del 0.86 (IC95%: 0.85-0.87), como en la cohorte de validación con C-Statistic=0.83 (IC95%: 0.82-0.86) para predecir el riesgo individualizado de una prueba COVID-19 positiva.

- El score propuesto para pronóstico es el score de Mortalidad 4C (Coronavirus Clinical Characterisation Consortium) que fue creado por el consorcio ISARIC4C⁽¹³³⁾ (International Severe Acute Respiratory and emerging Infections Consortium), que es un consorcio de médicos y científicos del Reino Unido comprometidos en responder preguntas urgentes sobre COVID-19 de manera rápida, abierta y para el beneficio de todos. Fue probado en 260 hospitales de Inglaterra, Escocia y Gales y mostró una alta discriminación para mortalidad.

Este modelo predictivo está formado por 8 variables: edad, sexo, frecuencia respiratoria, número de comorbilidades, saturación de oxígeno, nivel de conciencia, urea y proteína C reactiva (PCR). El AUC para la cohorte de derivación (n=35.463) fue 0.786 (IC95%: 0.78-0.79) y para la cohorte de validación (n=22.361) de 0.77 (IC95%: 0.76-0.78) con una excelente calibración.

En conclusión, Wynants et al, indican que estos dos estudios propuestos pueden ser estudios prometedores pero que deben ser validados en múltiples cohortes antes de ser implementados en la práctica clínica, ya que:

- El modelo diagnóstico propuesto por Jehi⁽¹³²⁾ tuvo una gran cantidad de datos faltantes, y no queda claro si la imputación de datos (con el valor mediano) afectó a los resultados. Existen inexplicables discrepancias entre el nomograma, la regresión lineal y la calculadora web.
- El modelo pronóstico propuesto por Knight⁽¹³³⁾ para mortalidad hospitalaria, 4C score, plantea dudas porque no queda claro cómo se pasa de las puntuaciones del score a los riesgos de mortalidad previstos que se utilizaron para evaluar la calibración.

Otra red internacional que ha tenido un papel destacado en esta pandemia ha sido la Colaboración Cochrane⁽¹³⁴⁾. Es una red, con sede en Reino Unido, sin ánimo de lucro que reúne a un grupo de investigadores de

ciencias de la salud con más de 30.000 voluntarios y colaboradores (investigadores, profesionales sanitarios, pacientes, cuidadores y ciudadanos) en más de 130 países del mundo. Este grupo, recopila y resume la mejor evidencia de la investigación aplicando un riguroso y sistemático proceso de revisión de las intervenciones en salud con el objetivo de proporcionar información de alta calidad para la toma de decisiones en salud. Los resultados de estas RS se publican en la Biblioteca Cochrane Cochrane Library⁽¹³⁵⁾.

En el caso de la COVID-19, desde el comienzo de la pandemia, la Cochrane ha estado trabajando en diferentes aspectos para intentar ayudar a los clínicos a responder a ciertas cuestiones como, por ejemplo: qué síntomas o signos de la exploración clínica inicial permiten hacer un diagnóstico, lo más exacto posible, de la enfermedad. No ha sido tarea fácil ya que uno de los problemas de esta enfermedad es, que el diagnóstico formal de la enfermedad requiere de pruebas diagnósticas de laboratorio (ej: RT-PCR) y de pruebas de diagnóstico por imagen (ej: tomografía axial computerizada –TAC-), lo que conlleva mucho tiempo. La cuestión se agravó durante el comienzo de la pandemia porque no siempre se disponía del tiempo necesario para obtener el resultado de las pruebas o no se disponía de los recursos necesarios y se hacía indispensable encontrar otras formas de detectar la enfermedad.

Con este objetivo, el grupo Cochrane realizó una RS⁽¹³⁶⁾ en la que se analizaban 16 estudios. En ellos, se evalúan 27 signos y síntomas separadamente pero no se evaluó la combinación de ambos. Ningún estudio incluyó niños y sólo 1 estudio se centró específicamente en adultos más mayores.

Los signos y síntomas individuales incluidos en esta revisión tenían propiedades diagnósticas deficientes ya que la mayoría de los participantes ya habían sido ingresados o habían acudido al hospital. Esto hace que los hallazgos encontrados sean menos aplicables a las personas que acuden a AP, ya que pueden tener síntomas más leves y una menor prevalencia de la enfermedad.

La revisión concluye que ni la presencia ni la ausencia de signos o síntomas son lo suficientemente exactas como para confirmar o descartar la enfermedad.

Además de la aportación realizada por estos dos grandes grupos de trabajo (COVID-PRECISE y grupo Cochrane), desde que comenzó la pandemia hasta el momento actual, se ha desarrollado una importante labor investigadora y se han publicado una gran cantidad de estudios de modelos predictivos creados expresamente para la COVID-19, tanto para diagnóstico como para pronóstico. En esta tesis, nuestro interés se centra en los modelos que permiten predecir la presencia de infección por SARS-CoV-2 en pacientes con sospecha de infección, y modelos para predecir el pronóstico o el curso de la infección (ingreso en UCI o desenlace fatal) en los pacientes con la enfermedad.

Para la derivación de los modelos predictivos se ha considerado la información clínica, la exploración física y las pruebas complementarias. Los estudios de modelos predictivos diagnóstico se han focalizado, en su mayoría, en los signos y síntomas que refieren los pacientes, en las pruebas de imagen (radiografía de tórax, ultrasonidos o TAC) y/o en los parámetros analíticos (biomarcadores) como métodos para el diagnóstico de la COVID-19.

Los síntomas más frecuentes en estos pacientes, según la Organización Mundial de la Salud (OMS)⁽¹³⁷⁾, han sido: fiebre, tos no productiva y fatiga. La revista *The New England Journal of Medicine*⁽¹³⁸⁾ publicó que los síntomas febriles estaban presentes en el 89% de los pacientes, la tos en el 68% y la fatiga en el 38%. Hay otros síntomas que también son referidos en un porcentaje considerable de pacientes (40%)⁽¹³⁹⁾, como son: pérdida del gusto (ageusia) y del olfato (anosmia) y otros, que debido a que eran síntomas inespecíficos hacían complicado que se pudiese diferenciar a los pacientes afectados por la enfermedad de la COVID-19 de aquellos con otras infecciones.

Las pruebas de imagen se han empleado para el screening y para la monitorización de pacientes, siendo el TAC la prueba que mejores resultados ha proporcionado para la detección de la enfermedad, con una sensibilidad del 97-98% frente a la radiografía de tórax que obtiene una $Se=69%$ ⁽¹⁴⁰⁾. El inconveniente de la tomografía es que puede no ser concluyente en las primeras fases de la enfermedad⁽¹⁴¹⁾. Lu et al.⁽¹⁴²⁾, indican que la prueba de ultrasonido (cuando se usa el TAC como Gold-Standard) obtiene unas sensibilidades del 69%, 78% y del 100% para lesiones medias, moderadas y severas del pulmón respectivamente.

Algunos autores se han centrado en el análisis diferencial de glóbulos blancos (White blood cell –WBC-), que se basa en la tecnología "VCS". WBC se caracteriza e identifica por su volumen "V" (medido por impedancia directa), su conductividad "C" (analizada por conductividad en corriente de radiofrecuencia) y su dispersión "S" de un rayo láser. El interés por estudiar los parámetros VCS, llamados DPC (Datos de Población Celular), es porque se sabe que se modifican en caso de enfermedades infecciosas⁽¹⁴³⁻¹⁴⁶⁾ y se ha analizado si estas variaciones podrían constituir un elemento de ayuda para el diagnóstico de COVID-19.

Así mismo, se ha analizado si los datos de la historia clínica (diabetes, obesidad, hipertensión, etc) del paciente, recogidos de rutina en la práctica habitual, permitían la identificación de los pacientes infectados por la enfermedad y posibilitaban su tratamiento y aislamiento tempranos.

Cabe destacar que para crear un modelo predictivo de diagnóstico es necesario tener una prueba diagnóstica de referencia (Gold-Standard), en este caso, se ha empleado la RT-PCR (reverse transcription polymerase chain reaction). Sin embargo, en los modelos predictivos de pronóstico, no es necesario disponer de ninguna prueba de referencia.

En el caso de los modelos predictivos pronósticos, la mayoría de los estudios publicados han centrado su atención en desarrollar modelos para predecir un peor pronóstico del paciente (ingreso en UCI) o un desenlace fatal (muerte). Algunos de los estudios han derivado modelos para predecir únicamente el riesgo de muerte en pacientes críticos, pero son muy pocos los estudios que se han centrado en ello. En algunos casos, se han derivado modelos que luego se han validado en pacientes críticos, pero no se crearon expresamente para este tipo de pacientes.

También se han validado modelos predictivos que habían sido creados con anterioridad a la pandemia en pacientes COVID-19. Se ha evaluado la capacidad predictiva de dichos modelos en estos pacientes y se ha estudiado si se podían emplear en la práctica clínica.

Modelos predictivos no creados para la COVID-19.

Los clínicos suelen usar scores pronósticos para ayudarse en la toma de decisiones en su práctica clínica diaria. Generalmente, estos scores permiten predecir qué pacientes tienen más riesgo de tener un mal pronóstico. En esta pandemia, dada la gravedad de la situación, los clínicos se han basado en scores ya creados y los han aplicado en pacientes afectados por la enfermedad del SARS-CoV-2.

Los scores más validados han sido:

- NEWS2⁽¹⁴⁷⁾ (National Early Warning Score 2)
- SEWS⁽¹⁴⁸⁾ (Standardized Early Warning Score)
- HEWS⁽¹⁴⁹⁾ (Hamilton Early Warning Score)
- MEWS⁽¹⁵⁰⁾ (Modified Early Warning Score)
- MuLBSTA⁽¹⁵¹⁾ (Multi-lobular infiltration, hypo-Lymphocytosis, Bacterial coinfection, Smoking history, hyper-Tension and Age)
- CURB65⁽¹⁵²⁾ (Confusion, Urea, Respiratory rate, Blood pressure, Age 65)
- SOFA⁽¹⁵³⁾ (Sequential – sepsis related- Organ Failure Assessment)
- qSOFA⁽¹⁵⁴⁾ (quick Sequential – sepsis related- Organ Failure Assessment)

En el estudio de Knight et al⁽¹³³⁾, los autores compararon el score 4C para predecir mortalidad en pacientes COVID-19 con 15 scores más, entre los que se encontraban scores que no fueron creados para el SARS-CoV-2, como son: SOFA, qSOFA, NEWS, CRB65 (Confusion, Respiratory rate, Blood pressure, Age 65), DS-CRB65 (Disease, Peripheral blood saturation-Confusion, Respiratory rate, Blood pressure, Age 65) y CURB65. Ninguno de ellos obtuvo un AUC superior a 0.77.

Gupta⁽¹⁵⁵⁾ valida 22 modelos pronósticos, de los cuales, 12 predicen deterioro clínico y 10 sirven para predecir mortalidad. De los 22 modelos, 17 fueron creados específicamente para COVID y los otros 5 (NEWS, MEWS, REMS (Rapid Emergency Medicine Score), qSOFA y CURB65) no fueron creados para COVID, pero se emplearon en los servicios de urgencias para predecir el pronóstico en este tipo de pacientes. En el estudio se concluye que NEWS2 para predecir deterioro a las 24h es el que tiene mayor área bajo la curva (AUC=0.78; IC95%: 0.73-0.83). Gidari et al⁽¹⁵⁶⁾, validaron también el NEWS2 en pacientes COVID-19 ingresados en la unidad de críticos, pero en este caso, obtuvieron un AUC=0.90 (IC95%: 0.82-0.97).

3. PRUEBAS DIAGNÓSTICAS DE LABORATORIO PARA COVID-19

Los modelos predictivos son ecuaciones matemáticas que se operan como combinación de diferentes componentes denominados variables o predictores. Estos predictores son: signos, síntomas, pruebas de imagen, pruebas diagnósticas, etc.

En este apartado, nos hemos enfocado en las pruebas diagnósticas de laboratorio. Dichas pruebas están pensadas para ayudar a los clínicos a identificar qué pacientes presentan la enfermedad y cuáles no. La RT-PCR se considera el método de referencia para el diagnóstico de la infección por SARS-CoV-2. Sin embargo, la gran demanda de pruebas de RT-PCR debido a la propagación mundial del virus ha llegado a poner de relieve las limitaciones de este tipo de diagnóstico. Además, su sensibilidad ha sido un tema de debate, ya que algunos estudios han reportado hasta un 20% de resultados falsos negativos para este tipo de prueba⁽¹⁵⁷⁾.

Hasta el momento, disponemos de dos tipos de pruebas para la detección de infección activa por coronavirus SARS CoV-2, una prueba rápida de detección de antígenos (*Antigen Rapid Diagnostic test, Ag-RDT*) y una prueba para la detección de ARN (Ácido RiboNucleico) viral mediante la RT-PCR. La realización de una u otra, depende del ámbito de realización, de la disponibilidad y de los días de evolución de los síntomas⁽¹⁵⁸⁾. Existe otra prueba, denominada *prueba de determinación de anticuerpos (Serología)*, que permite detectar infección pasada. Mediante esta prueba se detecta si el cuerpo humano ha producido anticuerpos contra el virus.

Cabe destacar que ninguna de las pruebas diagnósticas de laboratorio tiene una exactitud del 100%.

3.1. PRUEBA RT-PCR.

Como en tantas otras crisis de salud pública, relacionadas con enfermedades infecciosas, se está utilizando la prueba de la RT-PCR para detectar si una persona está infectada o no con coronavirus. Mediante la RT-PCR se localiza y amplifica un fragmento de material genético que, en el caso del coronavirus, es una molécula de ARN.

Un sistema habitual de RT-PCR suele constar de 40 ciclos. Tras cada ciclo de RT-PCR se duplica el número de copias que había en la muestra original, por lo que el número final de copias al final del proceso es variable. Los sistemas de RT-PCR en tiempo real, permiten detectar con certeza una diana positiva cuando la amplificación se produce antes del ciclo 35.

En nuestro hospital, la detección del genoma de coronavirus se realiza con 3 RT-PCR simultaneas dirigidas a tres genes distintos del virus. El resultado se considera negativo cuando no se produce la amplificación de ninguno de los tres genes. El resultado positivo requiere la amplificación de las tres dianas antes de llegar al ciclo 35. En ocasiones, se produce la amplificación aislada de uno o dos genes, o amplificaciones débiles (por encima del ciclo 35). Estos resultados son interpretados de forma individual con ayuda de otros datos, requiriendo en muchas ocasiones la repetición de la determinación con una nueva muestra.

La prueba RT-PCR detecta la infección desde prácticamente el primer momento. Permite confirmar que la infección se ha curado cuando ya no hay virus en el organismo (en cuyo caso, la prueba pasa de ser positiva a negativa). Tanto para RT-PCR como para pruebas rápidas de detección de antígenos, las muestras recomendadas para el diagnóstico de infección activa de SARS-CoV-2, según indicaciones del fabricante de la prueba diagnóstica de infección activa (PDIA), son del tracto respiratorio superior: denominadas nasofaríngeas (de elección en casos con infección grave y/o evolucionada, ya que se ha demostrado que, las muestras obtenidas de las regiones más profundas de la nasofaringe, tienen mayor concentración viral o mayor nivel de anticuerpos), o las denominadas orofaríngeas.

La RT-PCR con exudado nasofaríngeo es, por el momento, la prueba más fiable y sensible, por lo tanto, Gold-Standard para el diagnóstico de la COVID-19 en la población general. Pero tiene la desventaja de ser técnicamente compleja, por lo que se precisa de un período de tiempo de aproximadamente 24h hasta obtener el resultado.

No sirve como prueba generalizada en la población para conocer hasta qué punto la infección se ha extendido entre personas que no son conscientes de haberla pasado (porque no tuvieron síntomas o porque fueron leves), ya que, si en ese momento ya se han curado, la PCR será negativa. Para poder conocer la prevalencia en una determinada

población, se debe realizar la prueba de determinación de anticuerpos (Serología).

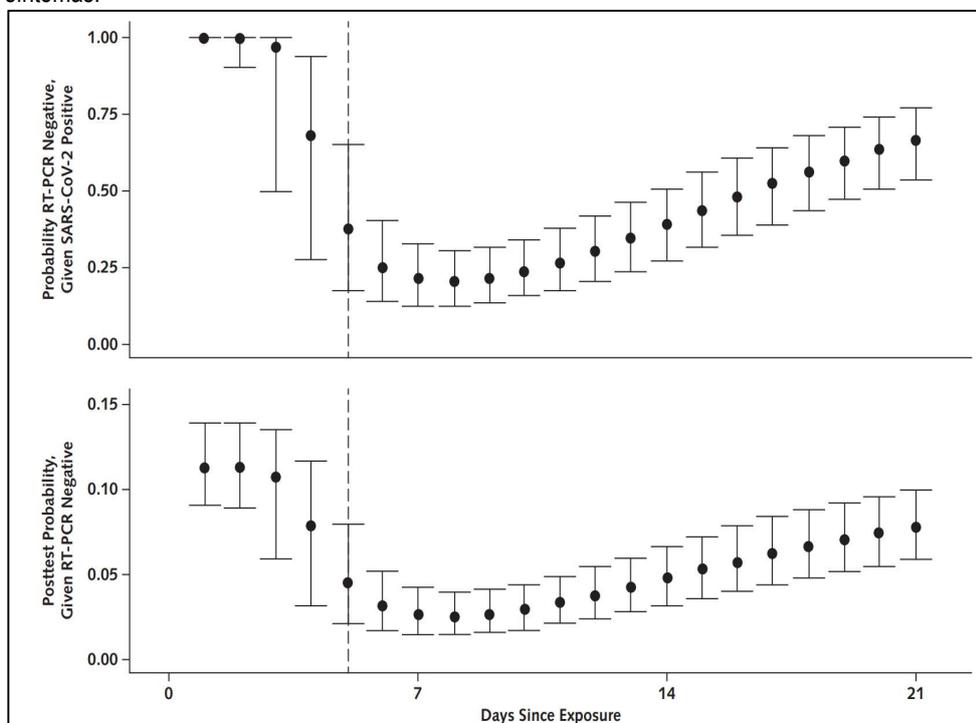
Mediante la técnica de RT-PCR se ha observado que los infectados presentan en su mayoría una alta carga viral (entre 10⁴ y 10⁸ copias de genoma/ml por muestra nasofaríngea o de saliva) durante los primeros días tras el inicio de la sintomatología. En pacientes que tienen un curso leve de infección, el pico de la carga viral en muestras nasales y orofaríngeas ocurre durante los primeros 5-6 días tras el inicio de síntomas y prácticamente desaparece al décimo día. Pero en algunos pacientes se detecta virus más allá del décimo día y la carga viral es del orden de 100-1.000 veces menor, lo cual sugiere una baja capacidad de transmisión en estos días⁽¹⁵⁹⁻¹⁶²⁾. Además, a partir del día 6º tras el inicio de síntomas, se ha podido demostrar, la ausencia de virus infectivo (no se detecta crecimiento del virus en cultivos) con cargas virales por debajo de 10⁵ copias por torunda⁽¹⁶²⁾. La comunidad científica internacional ha tenido en cuenta estos datos como parte de la evidencia para conocer los periodos de infecciosidad de la enfermedad, y se han considerado claves para el control de la misma^(163,164).

En su estudio, Beatriz Böger⁽¹⁶⁵⁾, realiza una RS con MA, en la que se incluyeron 16 estudios para diagnóstico de COVID-19: en 6 estudios se evalúa el TAC de tórax con Se=91.9% (IC95%: 89.9-93.7), en 5 estudios los tests inmunológicos –IgM (Inmunoglobulina M) Se=84.5% (IC95%: 82.2-86.6) e IgG (Inmunoglobulina G) Se=91.6% (IC95%: 86-95.4)- y, en 7 estudios, la RT-PCR. En estos 7 ensayos, se indica que la exactitud diagnóstica de la RT-PCR varía según el sitio de toma de la muestra: en hisopos rectales la Se=24.1% (IC95%:16.7-33), en orina la Se=0% (IC95%: 0-3.7), en plasma la Se=7.3% (IC95%: 4.1-11.7), en esputo (entendiendo que nos referimos al lavado bronquial) la Se=97.2% (IC95%: 90.3-99.7), en saliva la Se=62.3% (IC95%: 54.5-69.6), en nasofaríngea y orofaríngea la Se=73.3% (IC95%: 68.1-78).

Concluye que las muestras de heces/hisopos rectales, la orina y el plasma son las menos sensibles, mientras que el esputo es la de mayor sensibilidad para detectar el virus (aunque sólo puede realizarse a pacientes que hayan sido previamente intubados). Finalmente, se indica que es altamente recomendable combinar diferentes tests diagnósticos para mejorar la sensibilidad y la especificidad de la misma.

Por otro lado, Kim⁽¹⁶⁶⁾ indica que la sensibilidad del TAC de tórax es del 94% (IC95%: 91-96; I²=95%) y la de la RT-PCR es de 89% (IC95%: 81-94; I²=90%), aunque concluye que el TAC de tórax no es adecuada en regiones de baja prevalencia debido a la alta tasa de falsos positivos. Yang et al⁽¹⁶⁷⁾ indican que la sensibilidad de la RT-PCR esputo es del 90%, la nasofaríngea del 73% y la orofaríngea del 60% si se toma la muestra entre los 7 primeros días desde el inicio de los síntomas. Y Kucirka et al⁽¹⁵⁷⁾ concluyen que la sensibilidad de la RT-PCR es máxima (aprox. 80%) 3 días después del inicio de los síntomas y, a partir de ese día comienza a decrecer la sensibilidad llegando a un porcentaje de falsos negativos del 21% (IC95%: 13-31) en el día 9 (4 días después del inicio de los síntomas) hasta un 66% (IC95%: 54-77) en el día 21, es decir, indican que la exactitud diagnóstica de la prueba varía con el tiempo, como puede verse en el gráfico que mostramos a continuación (Figura 5).

Figura 5. Exactitud diagnóstica de la RT-PCR según los días transcurridos desde el comienzo de los síntomas.



Fuente: Variation in False-Negative Rate of Reverse Transcriptase Polymerase Chain Reaction–Based SARS-CoV-2 Tests by Time Since Exposure. *Ann Intern Med.* 2020 May 13;M20-1495.

3.2. PRUEBA RÁPIDA DE ANTÍGENOS PANBIO™ (AG RAPID TEST DEVICE).

El diagnóstico de la enfermedad se ha basado mayoritariamente en la positividad de la RT-PCR y/o en el TAC de tórax, que requieren mucho tiempo y pueden retrasar el diagnóstico. Por este motivo y, por el coste que conlleva cada prueba (entre otras razones), se hace necesario obtener otras formas de detectar la enfermedad. Con el transcurso del tiempo han ido surgiendo otras pruebas de laboratorio, por ejemplo: las pruebas rápidas de antígenos.

Se comercializan diferentes tipos de test de antígenos a nivel europeo, que han sido acordados por el Comité de Seguridad Sanitaria, que en el punto 11 de la Recomendación del Consejo de 21 de enero de 2021 pide a los Estados miembros que, sin perjuicio de la Directiva 98/79/CE (de Conformidad Europea), acuerden y mantengan una lista común y actualizada de pruebas rápidas de antígenos COVID-19 que se consideren apropiadas para su uso en el contexto de las situaciones descritas en el punto 6 de medidas sanitarias.

Abbott anunció el 9/9/2020 que había recibido la certificación de la CE para su dispositivo Panbio™ COVID-19 Ag Rapid Test Device. Nosotros hemos centrado nuestro estudio en analizar la prueba de antígeno de la empresa Abbott por ser el elegido para su uso en España y, por lo tanto, en nuestro hospital.

En esta prueba se busca la presencia de algunas de las proteínas (antígenos leucocitarios humanos) del virus SARS CoV-2 mediante una reacción química. Es una técnica rápida que proporciona el resultado en 15 minutos. Se necesita, como para la RT-PCR, una muestra nasofaríngea.

El rendimiento diagnóstico de esta prueba parece indicar que es mayor para detectar pacientes infectados por la enfermedad en el grupo de pacientes sintomáticos que en el grupo de pacientes asintomáticos, por lo que, la empresa ABBOTT, recomienda su uso en este tipo de pacientes.

3.3. PRUEBA DE DETERMINACIÓN DE ANTICUERPOS (SEROLOGÍA).

Las pruebas que detectan los anticuerpos que el organismo produce frente al virus, son llamadas "pruebas serológicas" o "serología". No detectan la presencia del virus, aunque cuando son positivas, es porque hay una reacción inmune contra éste y sí confirman que el organismo de la persona ha sido infectado por el virus. Estas pruebas permiten saber si una persona ha estado expuesta al virus y la prevalencia de la enfermedad.

En estas pruebas, lo que se busca es la presencia en la sangre de anticuerpos contra el virus. El sistema inmunitario (nuestras defensas) produce estos anticuerpos (inmunoglobulinas) para combatir y eliminar el virus:

- **Inmunoglobulina M (IgM):** la detección de los anticuerpos IgM tiende a indicar una exposición reciente a la COVID-19. Es la primera línea de defensa del organismo contra las infecciones. Aparece pronto, coincidiendo con la enfermedad, pero dura poco tiempo.
- **Inmunoglobulina G (IgG):** la detección de los anticuerpos IgG indica una etapa posterior de la infección, por lo que proporciona información sobre la etapa de la infección. Aparece después que la IgM y perdura más en el tiempo, más allá de la curación. Teóricamente, protegen al paciente frente a reinfecciones por el SARS-CoV-2.

No hay certezas sobre cuánto dura la IgG en la sangre ni sobre el grado de protección que proporciona. Tampoco conocemos si los enfermos están protegidos por la inmunidad celular (es otro componente de las defensas del organismo contra las infecciones) una vez que han perdido sus anticuerpos.

Si el análisis de sangre demuestra la presencia de anticuerpos, el paciente ha estado infectado por el SARS-CoV-2. Aunque se recomienda que esta información sea interpretada por profesionales sanitarios. Un resumen de dicha interpretación se muestra en la tabla siguiente (Tabla 6).

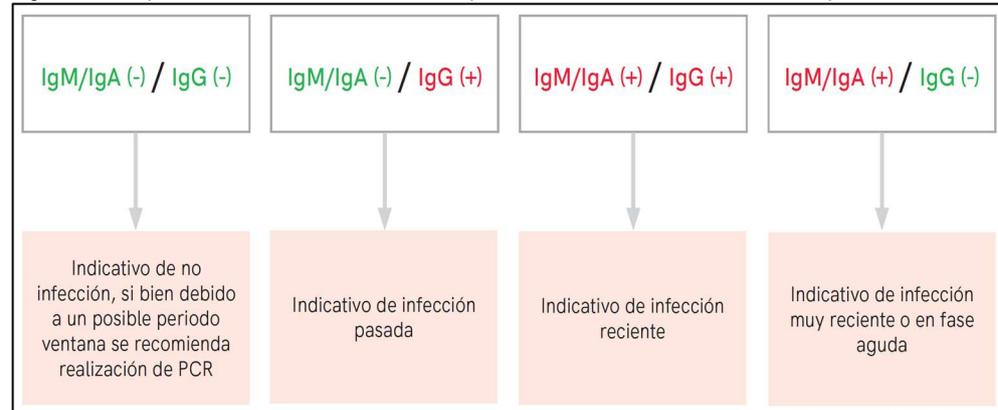
Tabla 6. Resumen de la interpretación de los resultados de las pruebas moleculares y serológicas para el diagnóstico de la enfermedad COVID-19 causada por el virus SARS-CoV-2.

RT-PCR	IgM/A	IgG	Significado Clínico
-	-	-	Negativo
+	-	-	Periodo ventana
+	+	-	Estadio temprano de la infección
+	+	+	Fase activa de la infección
+	-	+	Fase final de la infección
-	+	-	Estadio temprano con falso negativo (confirmar RT-PCR)
-	-	+	Infección pasada y curada
-	+	+	Enfermedad en evolución (confirmar RT-PCR)

Cuadro basado en los datos de Sanidad. (Fuente: MicroBIO)

En base a esto, podemos resumir del siguiente modo la interpretación de los resultados obtenidos al realizar la prueba de determinación de anticuerpos (Figura 6).

Figura 6. Interpretación de los resultados de la prueba de determinación de anticuerpos.

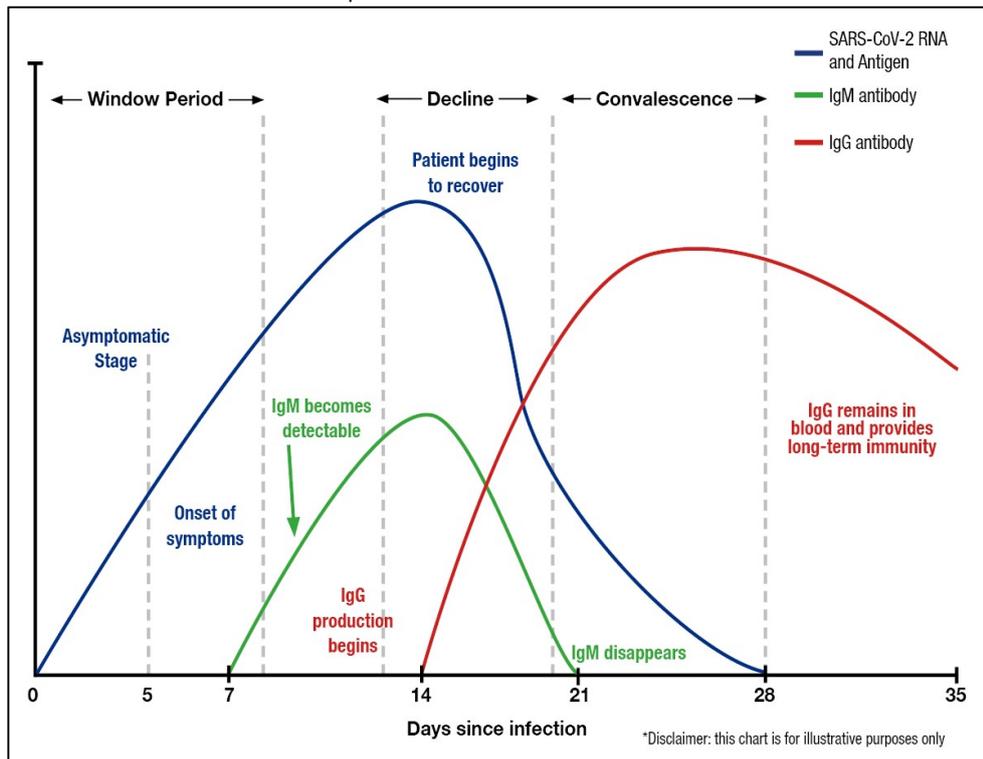


Cuadro creado por *Laboratorios Echevarne*.

El problema es que estas pruebas pueden ser negativas en los primeros días de la infección, cuando el organismo aún no ha producido anticuerpos o su nivel aún es muy bajo. Por lo general, aunque puede haber variaciones, la prueba es positiva a los 7 días de la infección y continúa siendo positiva una vez que la infección está curada. En ausencia de un resultado de la prueba RT-PCR, la serología (por sí sola) no permite concluir que la infección esté curada.

Tal y como vemos en la Figura 7, una vez que una persona se ha contagiado con el virus, comienza un período asintomático que puede durar varios días. Con el inicio de los síntomas podemos detectar el ARN mediante la prueba RT-PCR, que inicia su aumento. Posteriormente desciende a medida que el sujeto comienza a recuperarse y desaparece a los 20-21 días. Sobre los días 6 o 7 comienza la respuesta de anticuerpos, siendo la IgM la primera. A los 14-17 días aparece la IgG, que se eleva notablemente, manteniéndose en forma de meseta unos 35 días, aproximadamente⁽¹⁶⁸⁾.

Figura 7. Evolución de los niveles de ARN e inmunoglobulinas (IgM e IgG) después de la infección de la enfermedad COVID-19 causada por el virus SARS-CoV-2.



Fuente: Analytical and Bioanalytical Chemistry (2021) 413:2311–2330

No se recomienda la realización de ningún tipo de pruebas serológicas para el diagnóstico de infección activa ni en personas con síntomas ni en asintomáticos⁽¹⁶⁹⁾.

Se siguen desarrollando nuevos tests con el fin de obtener resultados fiables y que eviten poner en riesgo la salud de los profesionales sanitarios al mismo tiempo que permitan un diagnóstico en el menor tiempo posible, por ej: test de saliva. Se espera que con el paso del tiempo surjan nuevos tests que proporcionen buenos resultados o que puedan surgir nuevos modelos diagnósticos que nos ayuden de manera rápida y eficaz a identificar a los pacientes que presentan la enfermedad.

Los test clínicos que se lanzan al mercado deben ser suficientemente sensibles (es decir, reconocer muy bien a los infectados y tener un porcentaje lo más reducido posible de "falsos negativos") y suficientemente específicos (reconocer bien a los no infectados y tener un porcentaje reducido de "falsos positivos"). En cualquier caso, los expertos advierten de que todas las pruebas deben ser interpretadas en el contexto clínico de cada persona.

4. SCORES PROPIOS CREADOS EN EL HUD

La unidad de Medicina Interna (UMI) junto con otras unidades del HUD, ha creado dos scores que se han tenido en cuenta durante la pandemia, uno de ellos para diagnóstico clínico (creado expresamente para COVID-19) y otro para pronóstico (fue creado en el 2019 con la ayuda de las unidades de Epidemiología Clínica e Investigación del HUD y del Hospital Ramón y Cajal). El primero, se denomina Score de Sospecha Clínica (SSC) y, el segundo, Risk Score. Explicamos, a continuación, cada uno de ellos.

4.1. SCORE DE SOSPECHA CLÍNICA (SSC)

Al comienzo de la pandemia había mucha incertidumbre sobre cómo diagnosticar la enfermedad del SARS-CoV-2. No se disponía de modelos predictivos para el diagnóstico de la enfermedad y la prueba diagnóstica más fiable era la RT-PCR. Uno de los problemas al que nos enfrentábamos era que los tiempos de espera hasta obtener un resultado con la RT-PCR eran demasiado altos (entre 24 y 48h) y el número de pruebas que se realizaban diariamente en nuestro hospital (700 pruebas/día de media) no permitía abarcar la demanda del momento.

Durante ese tiempo, se establecieron 3 tipos de zonas en el hospital: zonas limpias (en donde se ubicaban los pacientes libres de COVID), zonas sucias (para ubicar a pacientes con COVID) y zonas grises (para aquellos pacientes que no tenían un diagnóstico COVID/No COVID confirmado).

Cuando los hospitales comenzaron a llenarse de pacientes con signos y síntomas compatibles con la enfermedad, se puso de manifiesto que era necesario poder ubicar a los pacientes en zonas limpias o sucias lo antes posible para poder descongestionar las urgencias y las zonas grises pero el tiempo hasta obtener el diagnóstico de los pacientes (mediante un resultado positivo de la RT-PCR) era demasiado alto, lo que complicaba mucho poder ubicar a esos pacientes en un breve espacio de tiempo.

Es así como surge la idea, entre los clínicos del HUD, de crear un score que les permitiera diagnosticar qué pacientes presentaban la enfermedad del SARS-CoV-2 y cuáles no.

Por dicho motivo, consideraron que los síntomas y signos que presentaban los pacientes podían ayudar a identificar qué pacientes estaban enfermos y diferenciarlos de aquellos que no presentaban la enfermedad. Se seleccionaron algunos de ellos para crear un Score de Sospecha Clínica (SSC): si el paciente presentaba fiebre, tos seca o expectoración blanquecina, si el resultado de la RT tórax era normal, unilateral o multilobar y si el paciente había tenido contacto epidemiológico. A cada uno de los síntomas y signos considerados se les asignó una puntuación, tal y como se muestra en la tabla 7.

Tabla 7. Características que componen el SSC y su puntuación.

ASPECTOS CONSIDERADOS	CARACTERÍSTICAS	PUNTUACIÓN
Contacto epidemiológico	<ul style="list-style-type: none"> - Contacto estrecho con una persona COVID positivo en los últimos 14 días. - Personal de riesgo (sanitario, policía, personal de prisiones, etc) que trabaje en un centro con brote conocido. - Residente en institución con brote conocido. 	1 pto
Temperatura	<ul style="list-style-type: none"> - Fiebre > 37.5°, con foco respiratorio - Fiebre > 37.5°, sin otro foco aparente 	1 pto
Criterios clínicos	Tos seca / anosmia / Mialgias	1 pto
Hallazgos radiológicos	<ul style="list-style-type: none"> - Afectación unilobar - Afectación multilobar 	1 pto 2 ptos

En base a estas características se asigna una puntuación total a cada paciente y dicha puntuación se asocia a un grupo de sospecha de tener o no tener la enfermedad (Tabla 8).

Tabla 8. Sospecha asignada según la puntuación obtenida en el SSC.

Puntuación	Sospecha
1 pto	Baja
2 ptos	Moderada
3 Ptos	Alta
4 – 5 ptos	Muy alta

4.2. RISK SCORE

La UMI, emplea en su práctica clínica habitual el score de riesgo "Risk Score" que es un score de alerta temprana que calcula el riesgo que tienen los pacientes que están ingresados en dicha unidad de presentar eventos graves, ingreso en UCI o fallecer en el futuro inmediato (24h siguientes). Este score lleva tiempo empleándose para priorizar la asistencia de los pacientes con mayor riesgo y se basa en la información clínica básica de las constantes vitales que se recogen de los pacientes rutinariamente en cada turno de enfermería.

Las variables que conforman el Risk Score son: frecuencia cardíaca (FC), frecuencia respiratoria (FR), presión arterial sistólica (PAS) y nivel de conciencia. Se considera que el paciente puede presentar un evento grave si se cumple que un paciente es positivo en alguna de las 4 variables mencionadas. En la tabla 9 indicamos qué condiciones se establecen para que se consideren positivas cada una de las variables.

Tabla 9. Características que componen el Risk Score.

VARIABLES	CONDICIÓN PARA ESTAR A RIESGO DE SUFRIR UN EVENTO GRAVE, INGRESO EN UCI O FALLECIMIENTO EN LAS PRÓXIMAS 24H
Presión arterial sistólica (PAS)	$PAS \leq 90$ o $PAS \geq 200$
Frecuencia cardíaca (FC)	$FC \leq 40$ o $FC \geq 120$
Frecuencia respiratoria	$FR \leq 8$ O $FR \geq 30$
Nivel de conciencia	Cualquier estado diferente a estar "Alerta"

Podemos clasificar a los pacientes ingresados en dos grupos:

- Paciente con la RPC (Risk Score) negativa: que representa a un paciente con las constantes vitales normales y sin riesgo de padecer evento grave durante las 24h siguientes a la toma de constantes.
- Paciente con la RPC (Risk Score) positiva: que representa a un paciente con al menos una de las constantes alterada y, por lo tanto, a riesgo de sufrir un evento grave.

PLANTEAMIENTO DE LA TESIS DOCTORAL

DESCRIPCIÓN DE LOS TEMAS TRATADOS EN LA TESIS DOCTORAL

Existen dos vías para la obtención del conocimiento, primaria y secundaria. La primaria se obtiene mediante la explotación directa de los datos que, en nuestro caso, se obtienen de los sistemas de información clínica (SIC) disponibles en el Servicio Vasco de Salud. La secundaria se obtiene mediante la búsqueda y revisión de la literatura científica disponible.

En esta tesis doctoral se van a emplear ambas vías para dar respuesta a cuestiones relacionadas con el SARS-CoV-2. Dichas cuestiones han sido planteadas por los clínicos del Hospital Universitario Donostia –HUD– (San Sebastián) durante la pandemia.

En un primer apartado, señalado con color amarillo en la Figura 8, se han realizado dos revisiones sistemáticas (RS) de los estudios publicados de modelos predictivos para COVID-19 durante el período (1/7/2020-5/9/2020). Una de las RS se ha realizado para identificar modelos predictivos de diagnóstico y, la otra RS, para identificar modelos predictivos de pronóstico. Entre los estudios identificados se han incluido aquellos modelos que cumplieran ciertos criterios, con el objetivo de intentar responder a cuestiones clínicas de vital importancia en los momentos de mayor incertidumbre.

Posteriormente, se han extraído datos de la primera ola (16/03/2020-17/05/2020) de los sistemas de información clínica de los pacientes de la organización de servicios integrados de Donostia (OSI-Donostialdea). Y, con estos datos, se han validado externamente los modelos predictivos de los estudios incluidos en las RS realizadas.

El mejor modelo predictivo obtenido (score ANDC) ha sido validado externamente, por segunda vez, con los pacientes COVID-19 de la segunda ola (1/8/2020-15/11/2020) del HUD. Y se ha realizado un análisis comparativo de algunas de las características de los pacientes de la primera y de la segunda ola para detectar posibles diferencias entre ambas poblaciones.

El score ANDC se ha implementado en la práctica clínica de las urgencias de nuestro hospital, como una característica más dentro del algoritmo de triaje. Para analizar su rendimiento, se reclutaron a todos los pacientes que acudieron a las urgencias del HUD durante el período comprendido entre (1/1/2021-28/2/2021) y se realizó un análisis descriptivo de dicha implementación.

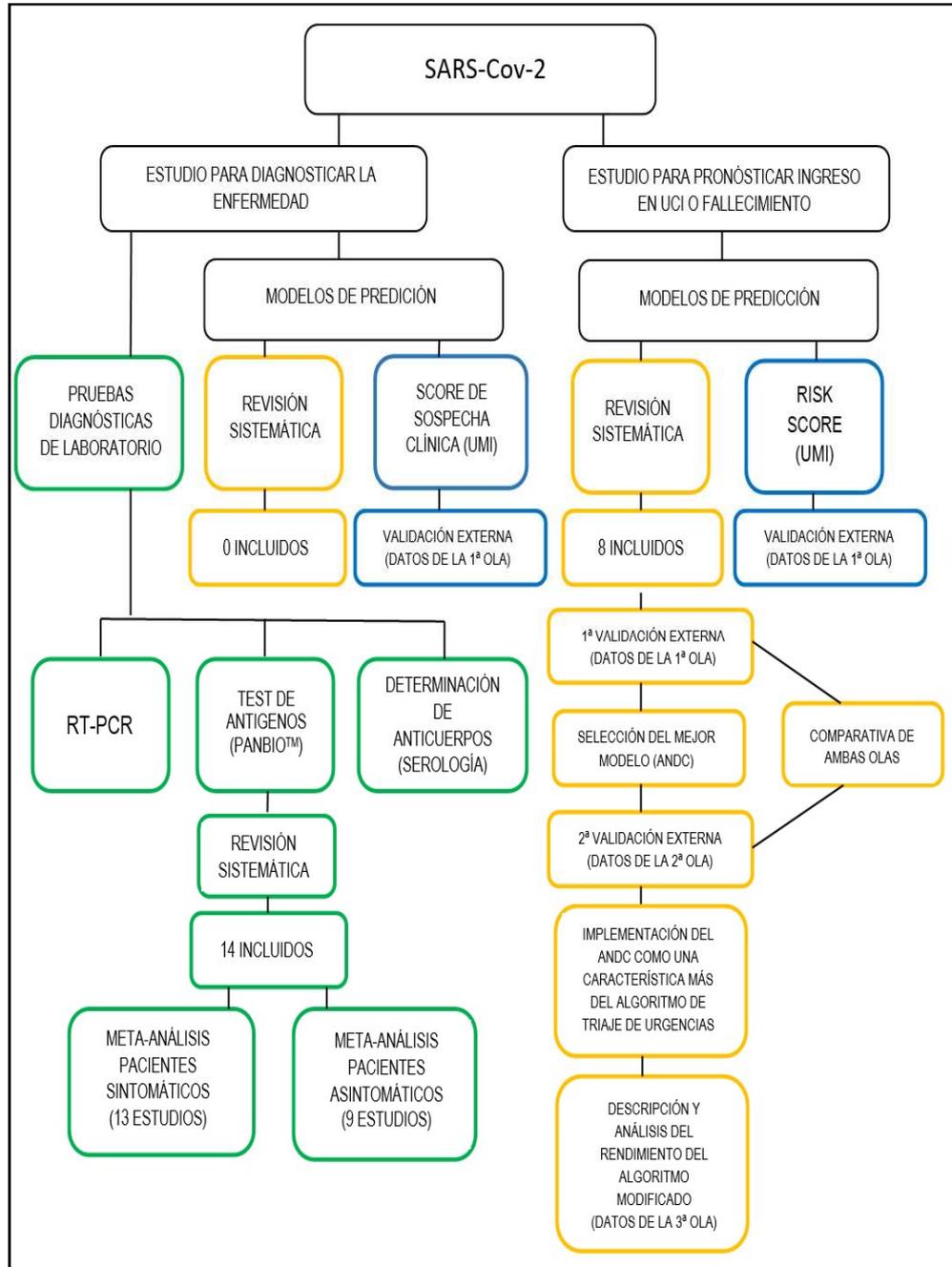
En un segundo apartado, señalado con color verde en la Figura 8, se ha analizado el rendimiento diagnóstico de las pruebas diagnósticas de laboratorio que han demostrado mayor efectividad hasta el momento, con los datos extraídos de nuestros SIC.

Por un lado, se han analizado las pruebas que detectan infección actual: prueba de la RT-PCR (reverse transcription polymerase chain reaction) y prueba rápida de antígenos Panbio™ y, por otro lado, la prueba que detecta infección pasada, es decir, la determinación de anticuerpos (denominada comúnmente *Serología*).

Hemos realizamos una RS, con meta-análisis (MA), de todos los estudios publicados hasta el 1 de diciembre de 2020 de la prueba rápida de antígenos Panbio™. Y se han llevado a cabo los análisis considerando, por un lado, pacientes sintomáticos y, por el otro, pacientes asintomáticos con el objetivo de evitar posible heterogeneidad entre estudios.

Y, por último, **en un tercer apartado** (señalado con color azul en la Figura 8), se han analizado dos scores de riesgo creados en nuestro hospital. Un Score de Sospecha Clínica –SSC- (para diagnóstico) y el Risk Score (para pronóstico).

Figura 8. Algoritmo para resumir el desarrollo de la tesis doctoral.



◇ Primer apartado ◇ Segundo apartado ◇ Tercer apartado

HIPÓTESIS Y OBJETIVOS

HIPÓTESIS

Una revisión sistemática de la literatura nos permitirá identificar modelos predictivos de diagnóstico y de pronóstico y esperamos que las características de los pacientes de la OSI-Donostialdea de Gipuzkoa sean similares, en términos generales, a las características de los pacientes que se han considerado para derivar los modelos predictivos identificados mediante las revisiones sistemáticas. Por lo que, esperamos obtener resultados similares a los proporcionados en los estudios cuando éstos sean validados externamente en nuestro entorno. Así mismo, se espera que el rendimiento predictivo de los modelos validados sea un poco menor al rendimiento que los autores reportan en sus estudios.

Entre todos los modelos validados externamente, esperamos obtener alguno con buen rendimiento predictivo y que pueda ser implementado en la práctica clínica de nuestro hospital.

Creemos que las características de los pacientes de la primera ola serán similares a las características de los pacientes de la segunda ola. Y se espera que no haya diferencias en la mortalidad y la letalidad entre ambas olas.

Es esperable que los valores de Se y Sp reportados en otros estudios para las pruebas diagnósticas de laboratorio: prueba de la RT-PCR, la prueba rápida de antígenos Panbio™ y la determinación de anticuerpos (Serología), sean similares cuando se evalúen dichas pruebas con los pacientes de nuestro entorno (OSI-Donostialdea).

Creemos que la prueba rápida de antígenos Panbio™ aportará resultados similares a los reportados por la empresa ABBOTT cuando se combinen, mediante MA, todos los estudios realizados hasta la fecha consultada. Cabe esperar que exista poca heterogeneidad entre estudios cuando se combinen los estudios mediante MA considerando, por un lado, los pacientes sintomáticos y, por el otro, los pacientes asintomáticos.

Pensamos que los dos scores de riesgo creados en nuestro hospital pueden ayudar: el Risk Score a predecir eventos graves (ingreso en UCI o muerte) en las 24h siguientes al ingreso y, el SSC, a discriminar a los enfermos de los no enfermos respectivamente y a predecir quiénes se van a positivizar. Creemos que la edad o los síntomas que presentan los pacientes pueden ayudar a mejorar esta última predicción.

OBJETIVOS

MODELOS PREDICTIVOS

1. Identificar, mediante una RS, qué modelos predictivos diagnósticos tienen características nosológicas suficientes para ser empleados en la práctica clínica ante sospecha de la enfermedad de la COVID-19.
2. Identificar mediante una RS qué modelos predictivos pronósticos tienen capacidad pronóstica suficiente para ser empleados en la práctica clínica en pacientes que presentan la enfermedad.
3. Validar externamente, en nuestro entorno y con los pacientes de la OSI-Donostialdea (primera ola), los modelos predictivos incluidos en las RS. Y elegir el mejor modelo predictivo.
4. Validar externamente, con datos de nuestros pacientes (segunda ola), el mejor modelo predictivo (score ANDC).
5. Comparar las características (edad, sexo y letalidad) de los pacientes de la primera y de la segunda ola.
6. Describir el uso del score ANDC y analizar su rendimiento tras la implantación del mismo en nuestro HUD.

PRUEBAS DIAGNÓSTICAS DE LABORATORIO

7. Caracterizar las pruebas diagnósticas de laboratorio: las pruebas de detección de infección activa (prueba RT-PCR y prueba rápida de antígenos Panbio™), y por otro, la prueba de detección de infección pasada (prueba de determinación de anticuerpos –Serología-) con los pacientes de la OSI-Donostialdea.
8. Realizar investigación de síntesis (RS) con MA para la prueba diagnóstica del test rápido de antígenos Panbio™, mediante 2 MA separados, uno para el grupo de pacientes sintomáticos y otro para el grupo de pacientes asintomáticos.

SCORES CREADOS EN EL HUD

9. Evaluar si el SSC discrimina entre enfermo y no enfermo COVID-19 o si permite discriminar pacientes que se van a positivizar de los que no.
10. Analizar si la variable edad o los síntomas más frecuentes: fiebre, tos y/o disnea (de manera conjunta o separada) ayudan a mejorar la discriminación entre pacientes que se van a positivizar de los que no.
11. Evaluar si el Risk Score permite predecir qué pacientes van a presentar eventos graves (ingreso en UCI o muerte) en las 24h siguientes tras al ingreso.

MATERIAL Y MÉTODOS

MATERIAL Y MÉTODOS

- TIPOS DE ESTUDIOS
- ÁMBITO
- BÚSQUEDA BIBLIOGRÁFICA
 - Para modelos predictivos de diagnóstico y de pronóstico
 - Para la prueba rápida de antígenos PANBIO™
- SUJETOS DE ESTUDIO
- RECOGIDA DE DATOS
 - Datos para los modelos predictivos de diagnóstico y de pronóstico
 - Datos de pacientes del HUD
 - Datos de los estudios de pruebas rápidas de antígenos PANBIO™
- CRITERIOS DE INCLUSIÓN Y EXCLUSIÓN
- TAMAÑO MUESTRAL
- VARIABLES DE ESTUDIO
 - Variables dependientes (outcome principales)
 - Variables independientes
- ANÁLISIS ESTADÍSTICO
 - Modelos predictivos de pronóstico
 - Pruebas diagnósticas de laboratorio
 - Scores creados en el HUD
- CONSIDERACIONES ÉTICAS

1. TIPOS DE ESTUDIOS

En la realización de esta tesis se han definido diversos tipos de estudios para dar respuesta a los objetivos planteados.

- RS de todas las RPC de diagnóstico y de pronóstico para COVID-19 publicadas en el periodo comprendido desde 1/7/2020 al 5/09/2020 (Objetivos 1 y 2).
- Estudio de cohorte para validación de RPC pronósticas, con datos de pacientes de la primera ola (16/03/2020-17/05/2020) (Objetivo 3).
- Estudio de cohorte para validación de la mejor RPC pronóstica, con datos de pacientes de la segunda ola (1/8/2020-15/11/2020) (Objetivo 4).
- Estudio de cohorte para comparar algunas características de los pacientes de la primera y la segunda ola (Objetivo 5).
- Estudio de cohorte para describir el uso y analizar el rendimiento del score ANDC con pacientes que acuden a urgencias (1/1/2021-28/2/2021) (Objetivo 6).
- Estudio de cohorte para análisis y caracterización de las pruebas diagnósticas de laboratorio. Pruebas para infección actual (RT-PCR y prueba de antígenos (Panbio™)) y para infección pasada (Serología) (Objetivo 7).
- RS de estudios de pruebas rápidas de antígenos (Panbio™) con combinación de resultados mediante MA, analizando por separado pacientes sintomáticos y asintomáticos (Objetivo 8).
- Estudio de cohorte para análisis del poder discriminatorio del SSC con pacientes de la 1ª ola (Objetivo 9 y 10).
- Estudio de cohorte para validación del Risk Score con pacientes de la 1ª ola (Objetivo 11).

2. ÁMBITO

El ámbito de actuación para esta tesis es la OSI-Donostialdea - Área de Gipuzkoa (País Vasco)- que cuenta con 32 Centros de Salud y el HUD. Dicha OSI abarca una población de referencia de cerca de 400.000 habitantes, que comprende San Sebastián y localidades como Oiartzoldea, Donostialdea y parte de Urola Kosta. Cabe destacar que la atención sanitaria de cuidados médicos intensivos es para toda la provincia de Gipuzkoa, que comprende aproximadamente 700.000 habitantes.

Hemos recabado información de las unidades que han tenido ingresos por COVID-19 en el HUD, que son: *unidades médicas* (medicina interna, enfermedades infecciosas y urgencias) y *unidades de críticos* (unidad de cuidados intensivos -UCI- y anestesia).

3. BÚSQUEDA BIBLIOGRÁFICA

3.1. PARA MODELOS PREDICTIVOS DE DIAGNÓSTICO Y DE PRONÓSTICO

Se realizaron varias estrategias de búsqueda en bases de datos bibliográficas: Medline y Embase en la plataforma Ovid, Web of Science, Base de Datos Cochrane de Revisiones Sistemáticas, Cochrane Central Register of Controlled Trials (CENTRAL), para seleccionar los estudios que derivaron o validaron un modelo multivariante o un sistema de puntuación (score) para resultados relacionados con la COVID-19 y basados en datos a nivel de paciente individual. No hubo restricción de idioma en ninguna de las búsquedas realizadas.

Por un lado, se realizó una estrategia de búsqueda para modelos predictivos diagnósticos (es decir, modelos que permitían predecir la presencia o ausencia de la enfermedad en pacientes con síntomas o signos compatibles con SARS-CoV-2) y, por el otro, se llevó a cabo una estrategia de búsqueda para modelos predictivos pronósticos (es decir, para modelos que permitían predecir qué pacientes estaban a riesgo de morir o ingresar en UCI).

En la primera búsqueda, los términos principales empleados fueron los siguientes: coronavirus, diagnostic or diagnosis, nomogram, tools, score, model predictive. Estos términos se buscaron a través de descriptores (MeSH) y términos en lenguaje libre (título, abstract).

Y en la segunda, los términos empleados fueron: coronavirus, prognostic, nomogram, tools, score, model predictive. Igualmente, se buscaron descriptores MesH y términos en lenguaje libre (título, abstract).

El período de tiempo establecido para las búsquedas comprendió desde el 1 de Julio de 2020 hasta el 5 de septiembre de 2020. Se seleccionó este periodo porque el grupo COVID-PRECISE había revisado todos los estudios publicados hasta el 1 de julio y consideramos apropiado revisar los 2 meses siguientes.

Para seleccionar los artículos pertinentes (en base al título y al resumen) de entre los identificados mediante las búsquedas bibliográficas, se llevó a cabo una prueba piloto de revisión supervisada en la que se seleccionaron, mediante muestreo sistemático, 50 estudios. Estos estudios fueron revisados por pares por dos evaluadores (MI y JIE) de forma independiente, en la fase de selección de texto completo. Se obtuvo un alto grado de acuerdo con una kappa=0.98. En base a estos resultados, se decidió que un único evaluador (MI) realizara el resto de las revisiones de manera individualizada, aunque se acudió al revisor que había realizado la revisión piloto (JIE) para solventar los casos de dudas.

3.2. PARA LA PRUEBA RÁPIDA DE ANTÍGENOS (PANBIO™)

Se establecieron alertas y se llevaron a cabo búsquedas en MedRxiv y LitCovid para seleccionar todos los estudios en los que se analizaba la eficacia de las pruebas rápidas de antígenos (Panbio™). El período de búsqueda comprendió desde 1/03/2020 hasta el 1/12/2020. Los términos seleccionados para realizar las búsquedas fueron "antigen rapid tests" y "Panbio™". No se aplicaron restricciones de idioma.

Se añadió a esta búsqueda el informe: *Evaluation du test antigénique ABBOTT*, realizado entre los días 16 y 17 de septiembre por el Laboratoire de Virologie des Hospices Civils de Lyon. Dicho informe fue obtenido mediante el seguimiento en twitter de científicos y/o personas relevantes de este tema.

4. SUJETOS DE ESTUDIO

Se incluyeron en nuestra base de datos todos los pacientes a los que se les cumplimentó formulario COVID-19 (tanto de AP como de HUD) durante el período comprendido entre el 16/03/2020 – 17/05/2020, todos los pacientes COVID-19 registrados en la segunda ola (1/8/2020 – 15/11/2020) y todos los pacientes que acudieron al servicio de urgencias desde el 1 de enero de 2021 hasta el 28 de febrero de 2021.

5. RECOGIDA DE DATOS

5.1. DATOS PARA LOS MODELOS PREDICTIVOS DE DIAGNÓSTICO Y DE PRONÓSTICO.

Se crearon dos bases de datos Excel, una de ellas para registrar los datos relativos a los modelos predictivos de diagnóstico y otra para registrar los datos de los modelos predictivos de pronóstico (las variables incluidas se pueden consultar en el apartado 8 “Variables del estudio”).

Estas bases de datos se han usado sólo de manera meramente informativa y sirven como resumen de todos los modelos revisados. La información recogida no se utilizó posteriormente para realizar análisis estadísticos.

5.2. DATOS DE PACIENTES DEL HUD.

Durante la primera ola de la pandemia (16 de marzo a 17 de mayo), se registraron consecutivamente datos referidos a los pacientes de la OSI-Donostialdea en las bases de datos clínicas habituales. Así mismo, se cumplimentaron formularios adicionales (que fueron creados expresamente para la enfermedad) con las características de los pacientes.

Se realizaron diferentes consultas en los sistemas de información clínica (SIC) -Oracle Analytic Server (OAS)-, disponibles en Osakidetza. Mediante esas consultas se extrajeron retrospectivamente datos de los pacientes de la OSI-Donostialdea durante los meses de marzo a junio de 2020.

Se han seleccionado aquellos pacientes que acudieron a:

- Atención primaria (AP) siendo sospechosos de estar infectados por la COVID. A estos pacientes se les había cumplimentado el *Formulario COVID19 de AP*.

- Las urgencias del HUD siendo sospechosos de estar infectados por la Covid-19. A estos pacientes se les había cumplimentado el *Formulario COVID19 de HUD*.

Así mismo, se extrajeron datos de analíticas, constantes, prescripciones, diagnósticos y procedimientos de los pacientes ingresados en esas fechas.

Con estas consultas obtuvimos las siguientes bases de datos, para los periodos que indicamos a continuación:

- **RT-PCR -Código 11810-** (1 de marzo – 15 de junio)
- **Serologías -Código 11937-** (10 de marzo – 29 de junio)
- **Pacientes con episodio abierto en AP** (7 de marzo – 10 de mayo)
- **Formularios COVID-19 en AP** (10 de marzo – 17 de mayo)
- **Formularios COVID-19 en HUD** (31 de marzo – 27 de julio)
- **Ingresos HUD** (16 de marzo – 17 de mayo). Y para esos pacientes:

<ul style="list-style-type: none"> - Prescripciones - Analíticas - Constantes - Procedimientos - Diagnósticos 	}	(16 de marzo – 2 de junio de 2020)
--	---	------------------------------------

Las fechas seleccionadas para cada una de las bases de datos extraídas se establecieron considerando diferentes criterios.

- Las bases de datos de RT-PCR y de Serologías (con un lapso de 15 días tras hacer la RT-PCR) comprenden todas las pruebas diagnósticas de laboratorio disponibles en el momento en el que se hicieron las correspondientes consultas (15 de junio y 29 de junio respectivamente).
- La base de datos de episodios abiertos en AP comprende del 7 de marzo al 10 de mayo de 2020 porque definimos la primera ola entre el 16 de marzo y el 17 de mayo de 2020, por lo que el episodio de los pacientes que acudieron a AP debía comenzar antes. Se fijó en 9 días la fecha previa porque hasta ese momento no se habían empezado a registrar, en AP, pacientes afectados por la enfermedad.

- La fecha para la base de datos de formularios AP comienza el 10 de marzo que es cuando se empezó a disponer de dicho formulario. Esta base termina el día que hemos definido que finaliza la 1ª ola.
- La base de datos del formulario de HUD comienza el día que estuvo disponible el formulario y finaliza el día 17 de mayo de 2020. Figura como última fecha de registro el 27 de julio porque los formularios fueron revisados posteriormente por un grupo de clínicos para cumplimentar los datos faltantes. Se extrajeron datos de las historias clínicas de esos pacientes para reducir al máximo posible los datos faltantes. Estas modificaciones hicieron que la fecha de registro del formulario fuese modificada.
- La fecha de los ingresos se fija en el periodo establecido para la primera ola. Y las bases de datos que dependen de esos ingresos (analíticas, constantes, ...) comienzan en la misma fecha, pero se finaliza el 2 de junio ya que consideramos que un paciente puede seguir ingresado más allá de la fecha fijada para el final de la 1ª ola.

Existen diferentes aplicaciones en los SIC para registrar datos y todas exportan datos a OAS que es donde se han realizado todas las consultas.

Detallamos a continuación cuáles son esas aplicaciones informáticas y qué información, de toda la que se exporta a OAS, ha sido extraída para su posterior análisis:

- GestLab: en esta aplicación se registra información de laboratorio y se han extraído datos de RT-PCR, serologías y analíticas.
- Osabide Global Primaria (OGP): se registra información de la historia clínica electrónica de los pacientes con episodio abierto en AP. Se han extraído datos de los episodios abiertos en AP y datos de formularios COVID-19 en AP (en el apartado *formularios*).
- Osabide Global en el ámbito de especializada (consultas externa y hospitalización): se registra información de la historia clínica electrónica de los pacientes. Se han extraído datos de los formularios de HUD (en el apartado *formularios*) y datos de las constantes de los pacientes (en el apartado *Constantes*).

- eOsabide Admisión: da soporte a Osabide Global. Se registra información de los movimientos de pacientes (ingresos, altas y traslados). Se han extraído datos de los ingresos que se han producido en el HUD.
- Presbide: se registra información de las prescripciones realizadas y esos son los datos que se han extraído.
- Kodifika: sirve como apoyo a la codificación clínica. Se han extraído datos de procedimientos y diagnósticos.

Además de extraer información de los datos de pacientes de la primera ola (16/03/2020-17/05/2020), se extrajo información retrospectivamente de los pacientes de la segunda ola (1/8/2020-15/11/2020), pero sólo de aquellas variables que nos permitían realizar una segunda validación externa del mejor modelo predictivo pronóstico (score ANDC) obtenido en las validaciones realizadas con los datos de la primera ola.

Por último, se extrajeron retrospectivamente datos de todos los pacientes que habían acudido al servicio de urgencias del HUD, desde el 1 de enero de 2021 hasta el 28 de febrero de 2021. Estos datos se emplearon para describir el uso y el rendimiento del score ANDC en el triaje de los pacientes que acudieron a las urgencias del HUD en dicho período.

El sistema de información clínica OAS asigna un identificador a cada paciente, por lo que, el traslado de la información de los pacientes a las bases de datos Excel se llevó a cabo de manera anonimizada.

5.3. DATOS DE LOS ESTUDIOS DE PRUEBAS RÁPIDAS DE ANTÍGENOS (Panbio™).

Para realizar la combinación de estudios de la prueba rápida de antígenos Panbio™ mediante MA se creó una base de datos en Excel. En dicha base de datos se registró información relevante de los estudios seleccionados (se puede consultar en el apartado 8 “Variables del estudio”).

6. CRITERIOS DE INCLUSIÓN Y EXCLUSIÓN

6.1. CRITERIOS DE INCLUSIÓN:

6.1.1. Criterios de inclusión para los modelos predictivos de diagnóstico y de pronóstico.

- Se seleccionaron los modelos cuyo objetivo fuera predecir qué pacientes, de los que presentaban síntomas compatibles con SARS-CoV-2, estaban realmente infectados.
- Se seleccionaron los modelos cuyo objetivo fuera predecir mortalidad o ingreso en UCI. Se incluyeron también scores pronósticos comúnmente usados en la práctica clínica que, aunque no se hubieran creado específicamente para la COVID19, se hubieran validado con este tipo de pacientes.
- Los modelos debían haber sido informados completamente (i.e. incluir la constante) y tenían que ser reproducibles.
- La capacidad predictiva del modelo debía ser ≥ 0.80 (tanto para el AUC como para el C-Statistic).

6.1.2. Criterios de inclusión para los pacientes de la OSI-Donostialdea.

Se han incluido en nuestra base de datos todos los pacientes que habían solicitado asistencia sanitaria tanto en AP como en el HUD en los períodos comprendidos entre los meses de marzo a mediados de mayo de 2020 (primera ola) y entre los meses de agosto hasta mediados de noviembre (segunda ola).

Se han incluido también todos los pacientes que acudieron al servicio de urgencias desde el 1 de enero de 2021 hasta el 28 de febrero de 2021.

6.1.3. Criterios de inclusión para los estudios de pruebas rápidas de antígenos (Panbio™).

Se han incluido todos los estudios de pruebas rápidas de antígenos Panbio™ (*Ag Rapid Test Device* de la empresa ABBOTT) que habían sido publicados hasta el 1 de diciembre de 2020.

6.2. CRITERIOS DE EXCLUSIÓN:

6.2.1. Criterios de exclusión para los modelos predictivos de diagnóstico y de pronóstico.

- Se excluyeron los modelos que no permitían su reconstrucción.
- Se excluyeron los modelos que no permitían el acceso al texto completo del artículo.
- Se excluyeron los modelos que estaban conformados por variables que, en nuestra base de datos, tenían un alto porcentaje de valores missing (entendemos que no es razonable llevar a cabo la imputación de datos cuando el porcentaje de dato faltantes sea superior al 10%).
- No se tuvieron en cuenta aquellos modelos que contenían variables que no estaban disponibles en nuestras bases de datos.
- No se contemplaron modelos cuyo rendimiento predictivo fuese bajo o moderado, es decir, los que tenían un AUC<0,8 o un C-Statistic<0,8.
- No consideramos los modelos para los que sólo se proporcionaba una calculadora web para obtener el score.

6.2.2. Criterios de exclusión para los pacientes de la OSI-Donostialdea.

- Se excluyen los pacientes que fueron atendidos tanto en AP como en el HUD en los periodos no seleccionados en los criterios de inclusión.
- Se excluyeron las unidades de ingresos que no estuvieron relacionadas directamente con las unidades en las que se ingresaban los pacientes COVID-19 positivo.

6.2.3. Criterios de exclusión para los estudios de pruebas rápidas de antígenos (Panbio™).

- Estudios que no se hubieran realizado empleando la prueba rápida de antígenos *Ag Rapid Test Device* de ABBOTT.

7. TAMAÑO MUESTRAL

7.1. TAMAÑO MUESTRAL PARA LA SELECCIÓN DE PACIENTES

No se ha estimado el tamaño muestral porque, en nuestro caso, estábamos interesados en estudiar a todos los pacientes que cumplieran los criterios de inclusión para los períodos de estudio.

8. VARIABLES DE ESTUDIO

8.1. VARIABLES DEPENDIENTES (OUTCOME PRINCIPALES)

8.1.1. Para los modelos predictivos diagnósticos.

- Covid positivo (Sí, No)

Nota: se considera Covid positivo si se obtiene, al menos, un resultado RT-PCR positivo.

8.1.2. Para los modelos predictivos pronósticos.

- Éxito (Sí/No)
- Ingreso en UCI (Sí/No)

8.1.3. Para las pruebas diagnósticas de laboratorio.

- Resultado test RT-PCR (+, -)
- Resultado test de Antígeno (+, -)
- Resultado test Serológico (+, -)

8.1.4. Para combinar estudios de la RS de la prueba rápida de antígenos (Panbio™).

- Sensibilidad (se combinan las sensibilidades de los estudios)

8.1.5. Para el Score de Sospecha Clínica (SSC).

- Sospecha de COVID positivo: baja, media, alta o muy alta

8.2. VARIABLES INDEPENDIENTES

8.2.1. Para los modelos predictivos de pronóstico.

Tabla 10. Variables para modelos predictivos de pronóstico.

Nombre de variable	Tipo de variable	Categorías	Medición
Id de estudio	Continua		
Fecha del estudio	Fecha		dd/mmm/aaaa
Autor	Texto		
Lugar (Ciudad /País)	Texto		
Nombre de la RPC	Texto		
Variables que conforman la RPC	Texto		
Tamaño muestral (D)	Continua		
AUC e IC95% (D)	Continua		
Tamaño muestral (V)	Continua		
AUC e IC95% (V)	Continua		
Tamaño muestral (V OSI)	Continua		
AUC e IC95% (V OSI)	Continua		
PROBAST (riesgo de sesgos)	Categoría	Bajo riesgo, alto riesgo, ?	
TRIPOD (reporte de resultados)	Categoría	Sí No	
Variable dependiente (outcome)	Categoría	Éxitos Críticos	

(D): muestra de Derivación de la RPC; (V): muestra de Validación de la RPC; (V OSI): muestra de Validación de la RPC (datos de la OSI); ?: incierto

8.2.2. Para las bases de datos de los pacientes de la OSI-Donostialdea.

Tabla 11. Variables de ingreso del paciente.

Nombre de variable	Tipo de variable	Categorías	Medición
Id de Paciente	Continua		
Episodio único	Continua		
Unidad de ingreso	Categoría	COVID, No COVID, gris, críticos	
Fecha de ingreso	Fecha		dd/mmm/aaaa
Traslado	Categoría (binaria)	Sí No	
Fechas de traslados	Fecha		dd/mmm/aaaa
Tipo de unidad de traslado	Categoría	COVID, No COVID, gris, críticos	
Fechas de alta	Fecha		dd/mmm/aaaa
Alta	Categoría	Alta médica domicilio habitual, alta médica hospitalización a domicilio, alta médica residencia social, alta voluntaria, fallecido, otras causas, traslado a hospital de agudo, traslado a media y larga estancia	
Tipo de unidad de alta	Categoría	COVID, No COVID, gris, críticos	
Fecha de defunción	Fecha		dd/mmm/aaaa
Estancia en no críticos	Continua		Días
Estancia en críticos	Continua		Días

Tabla 12. Variables sociodemográficas del paciente.

Nombre de variable	Tipo de variable	Categorías	Medición
Id de Paciente	Continua		
CIC	Continua		
Sexo	Categoría (binaria)	Hombre , Mujer	
Edad	Continua		Años
Situación	Categoría	Defunción, traslado, vigente	
Fecha de nacimiento	Fecha		dd/mmm/aaaa
Barthel	Continua		Puntuación de 0 a 100
Escalón estratificado	Categoría	Población sana, soporte autogestión, gestión enfermedad, gestión caso, Desconocida	
Marca PP	Categoría (binaria)	Sí No	
Formulario PPP	Categoría	Dependiente en domicilio, dependiente en gerontológico, estadio final, frágil con reingreso, pre-frágil, desconocido	
Marca estratificado	Categoría	IC, EP, EP e IC, DM, DM y EP y IC, No, PP, Desconocido	
Residenciado	Categoría (binaria)	Sí No	
Polimedicado	Categoría (binaria)	Sí No	
Dependencia	Categoría (binaria)	Sí No	
Demencia	Categoría (binaria)	Sí No	
Alerta coronavirus	Categoría (binaria)	Sí No	
Fecha de alerta	Fecha		dd/mmm/aaaa
Episodio en paliativo	Categoría (binaria)	Sí No	
Código cupo	Continua		
Código postal	Continua		
Charlson	Categoría	Alta comorbilidad, ausencia de comorbilidad, comorbilidad baja, desconocida	
¿Covid anterior?	Categoría (binaria)	Sí No	

Tabla 13. Variables de analíticas del paciente.

Nombre de variable	Tipo de variable	Categorías	Medición
Id de Paciente	Continua		
Fecha de la analítica	Fecha		dd/mm/aaaa
Hora de la analítica	Continua		hh:mm
Resultado texto	Categoría	No patológico, fuera de rango de normalidad, rango de pánico	
Resultado valor	Continua		
Id de Prueba	Continua		
Tipo de prueba realizada	Categoría	Pla-Dímero D, San Cloruro, ...	
Tipo de muestra	Categoría	Plasma citrato, plasma heparina, sangre total EDTA, sangre total Heparina, Litio, suero	
Área de realización	Categoría	Bioquímica especial, bioquímica general, equilibrio Ácido-Base, hematología general, hemostasia, proteínas	

Tabla 14. Variables de laboratorio.

Nombre de variable	Tipo de variable	Categorías	Medición
Id de Paciente	Continua		
Fecha de realización de la RT-PCR	Fecha		dd/mm/aaaa
Resultado RT-PCR	Categoría (binaria)	COVID positiva, COVID negativa	
Fecha de realización de Serología	Fecha		dd/mm/aaaa
Resultado Serología	Categoría (binaria)	COVID positiva, COVID negativa	

Tabla 15. Variables de constantes del paciente.

Nombre de variable	Tipo de variable	Categorías	Medición
Id de Paciente	Continua		
Fecha de toma de constantes	Fecha		dd/mm/aaaa
Hora de toma de constantes	Continua		hh:mm
Temperatura	Continua		Grados centígr
Estado neurológico	Categoría	Alerta, confuso/agitado, convulsión, no responde, responde a la llamada, responde al dolor	
Frecuencia respiratoria	Continua		rpm
Presión arterial sistólica (PAS)	Continua		mm Hg
Presión arterial diastólica (PAD)	Continua		mm Hg
Oxigenoterapia	Categoría	Aire ambiente, gafas O ₂ , IOT, Optiflow, VMK, VMNI	
Saturación O ₂	Continua		%
Frecuencia cardíaca	Continua		lpm

Tabla 16. Variables recogidas en el formulario COVID-19 de Atención Primaria.

Nombre de variable	Tipo de variable	Categorías	Medición
Id de Paciente	Continua		
Fecha ítem	Fecha		dd/mmm/aaaa
Inicio de síntomas	Fecha		dd/mmm/aaaa
Contacto epidemiológico	Categórica (binaria)	Sí No	
Fiebre	Categórica	Sí No No termometrada	
Disnea	Categórica	No, en reposo, ante tos o esfuerzo, ante actividad intensidad moderada - severa	
Vómitos	Categórica (binaria)	Sí No	
Diarrea	Categórica (binaria)	Sí No	
Tos	Categórica (binaria)	Sí No	
Expectoración	Categórica	Fétida, hemoptoica, mucosa, normal, purulenta, serosa	
Mialgia	Categórica (binaria)	Sí No	
Odinofagia	Categórica (binaria)	Leve, moderada, intensa, No	
Anosmia	Categórica (binaria)	Sí No	
Dolor torácico	Categórica (binaria)	Sí No	
EPOCAsma	Categórica (binaria)	Sí No	
Enfermedad cardiovascular (ECV)	Categórica (binaria)	Sí No	
Hipertensión arterial (HTA)	Categórica (binaria)	Sí No	
Diabetes Mellitus (DM)	Categórica (binaria)	Sí No	
Inmunosupresión	Categórica (binaria)	Sí No	
Enfermedad Renal Crónica (ERC) moderada – severa	Categórica (binaria)	Sí No	
Obesidad (IMC > 40)	Categórica (binaria)	Sí No	
Asplenia	Categórica (binaria)	Sí No	
Cefalea	Categórica (binaria)	Sí No	
Consumo de tabaco	Categórica	En deshabitación, ex fumador, fumador pasivo, no ocasional, sí	
Duración de la fiebre	Continua		Días
Malnutrición	Categórica (binaria)	Sí No	
Otalgia	Categórica (binaria)	Sí No	
Exantema COVID-19	Categórica (binaria)	Sí No	
Fecha fin seguimiento	Fecha		dd/mmm/aaaa
Lesiones acroisquémicas para estudio prospectivo de relación con COVID-19	Categórica (binaria)	Sí No	
Residente en área de transmisión comunitaria	Categórica (binaria)	Sí No	
Seguimiento de Salud Pública	Categórica (binaria)	Sí No	
Test contar disnea resultado	Continua		
Tiempo de evolución COVID-19	Continua		Días
Fatiga	Categórica (binaria)	Sí No	
Tipo de consulta	Categórica (binaria)	Presencial, telefónica	

Tabla 17. Variables recogidas en el formulario COVID-19 de Atención Primaria (continuación).

Nombre de variable	Tipo de variable	Categorías	Medición
Tipo seguimiento casos posibles COVID-19	Catagórica	Derivación hospitalaria, seguimiento equipo específico COVID-19, seguimiento intensivo, seguimiento ordinario	
Exantema	Catagórica (binaria)	Sí No	
¿Ha estado en región de riesgo?	Catagórica (binaria)	Sí No	
¿Ha realizado algún viaje en las últimas 2 semanas?	Catagórica (binaria)	Sí No	

Hay variables que se empezaron a recoger en el mes de Marzo (comienzo de la pandemia en nuestra OSI) pero que después se dejaron de registrar porque no se consideraron relevantes en el transcurso de la enfermedad o porque ya no eran de interés porque el virus estaba demasiado extendido (por ej: si había viajado a un país de riesgo) y, al contrario, otras que no se habían tenido en cuenta, pero comenzaron a registrarse a posteriori.

En la siguiente tabla se muestran dichas variables y se especifica cuáles fueron registradas cada mes y cuáles no.

Tabla 18. Características de los pacientes que sufrieron variaciones, en cuanto a su recogida o no, en la primera ola.

	Inicio de síntomas	Diarrea	Asplenia	Cefalea	Malnutrición	Otalgia	Fatiga	Vómito	Exantema	Ha estado en región de riesgo	Ha viajado en las últimas 2 semanas
Mar	No	No	Sí	Sí	Sí	Sí	Sí	No	No	Sí	Sí
Abr	Sí	Sí	No	No	No	No	No	Sí	Sí	Sí	Sí
May	Sí	Sí	No	No	No	No	No	Sí	Sí	No	No
Jun	Sí	Sí	No	No	No	No	No	Sí	Sí	No	No

Tabla 19. Variables recogidas en el formulario COVID-19 del HUD.

Nombre de variable	Tipo de variable	Categorías	Medición
Id de Paciente	Continua		
Fecha de registro	Fecha		dd/mmm/aaaa
Inicio de síntomas	Fecha		dd/mmm/aaaa
Contacto epidemiológico	Categoría (binaria)	Sí No	
Fiebre	Categoría	Sí N No termometrada	
Disnea	Categoría	No, en reposo, ante tos o esfuerzo, ante actividad intensidad moderada - severa	
Vómitos	Categoría (binaria)	Sí No	
Diarrea	Categoría (binaria)	Sí No	
NYHA	Categoría	I, II, III, IV	
Tos	Categoría (binaria)	Sí No	
Evolución subjetiva	Categoría	Peor, similar, mejor	
Expectoración	Categoría	Fétida, hemoptoica, mucosa, normal, purulenta, serosa	
Mialgia	Categoría (binaria)	Sí No	
Odinofagia	Categoría (binaria)	Leve, moderada, intensa, No	
Anosmia	Categoría (binaria)	Sí No	
Dolor torácico	Categoría (binaria)	Sí No	
Insuficiencia respiratoria	Categoría (binaria)	Sí No	
Afectación hemodinámica	Categoría (binaria)	Sí No	
Enfermedad pulmonar obstructiva Crónica (EPOC)	Categoría (binaria)	Sí No	
Enfermedad cardiovascular (ECV)	Categoría (binaria)	Sí No	
Hipertensión arterial (HTA)	Categoría (binaria)	Sí No	
Diabetes Mellitus (DM)	Categoría (binaria)	Sí No	
Inmunosupresión	Categoría (binaria)	Sí No	
Enfermedad Renal Crónica (ERC) moderada – severa	Categoría (binaria)	Sí No	
IMC	Continua		Kg / m ²
Obesidad (IMC > 40)	Categoría (binaria)	Sí No	
Tumor	Categoría (binaria)	Sí No	
Leucemia	Categoría (binaria)	Sí No	
Factor de riesgo CV	Categoría (binaria)	Sí No	
Gravedad	Categoría	Baja, media, alta	
Linfoma	Categoría	En deshabitación, ex fumador, fumador pasivo, no ocasional, sí	
RT-PCR orofaríngeo	Categoría (binaria)	Positivo, negativo	
RT-PCR esputo	Categoría (binaria)	Positivo, negativo	

NYHA: Clasificación funcional New York Heart Association

Tabla 20. Variables para la segunda validación externa del score ANDC.

Nombre de variable	Tipo de variable	Categorías	Medición
IdPaciente	Continua		
Proteína C reactiva	Continua		
Dímero D	Continua		
Ratio Neutrófilos/Linfocitos	Continua		
Edad	Continua		Años
COVID-19 positivo	Categoría (binaria)	Sí No	
Éxito	Categoría (binaria)	Sí No	

Tabla 21. Variables para el análisis del uso y rendimiento del score ANDC en Urgencias.

Nombre de variable	Tipo de variable	Categorías	Medición
IdPaciente	Continua		
Edad	Continua		Años
Días de evolución	Continua		Días
PAS	Continua		mmHg
PAD	Continua		mmHg
Frecuencia cardíaca	Continua		lpm
Frecuencia respiratoria	Continua		Rpm
Saturación O2	Continua		%
PAFI (PaO ₂ / FiO ₂)	Continua		mmHg / %
Temperatura	Categoría (binaria)	>=37.5 ° <37.5 °	Grados centígr.
Factores de riesgo	Categoría	0 1 >=2	
Radiografía de tórax (RX)	Categoría	Unilobar, Multilobar, Normal	
CURB65	Categoría	0 1 >=2	
Tienen Oxígeno	Categoría (binaria)	Sí No	
Criterio analítico	Categoría (binaria)	Sí No	
Destino al alta (en Urgencias)	Categoría	Hospitalización, UCI, UCRI, Domicilio	
Retornan al HUD	Categoría (binaria)	Sí No	
Insuficiencia respiratoria	Categoría (binaria)	Sí No	
Ingreso	Categoría (binaria)	Sí No	
Motivo de ingreso	Categoría	Alteración respiratoria, imagen radiológica, Otros, no se especifica	
Ingreso en UCI	Categoría (binaria)	Sí No	
Días transcurridos hasta el retorno			Días
Score ANDC	Categoría	Bajo, moderado, alto	

8.2.3. Para la base de datos de los estudios de pruebas rápidas de antígenos (Panbio™).

De cada estudio revisado, se ha recogido la información que mostramos en la siguiente tabla, siempre y cuando estuviese disponible en el artículo publicado.

Tabla 22. Variables para el MA del test de antígenos Panbio™.

Nombre de variable	Tipo de variable	Categorías	Medición
Id de estudio	Continua		
Fecha del estudio	Fecha		dd/mm/aaaa
Autor	Texto		
Lugar (Ciudad /País)	Texto		
Point of care	Categoría (binaria)	Sí No	
Síntomas	Categoría (binaria)	Sí No	
Notas del estudio (observaciones)	Texto		
Verdaderos positivos (tp)	Continua		
Falsos positivos (fp)	Continua		
Verdaderos negativos (tn)	Continua		
Falsos negativos (fn)	Continua		
Enfermos (ndis) (tp +fn)	Continua		
No enfermos (nndis) (fp + tn)	Continua		
Total (enfermos + no enfermos)	Continua		
Prevalencia	Continua		
Lím. inferior del IC95%	Continua		
Lím. superior del IC95%	Continua		
Sensibilidad	Continua		
Lím. inferior del IC95%	Continua		
Lím. superior del IC95%	Continua		
Especificidad	Continua		
Lím. inferior del IC95%	Continua		
Lím. superior del IC95%	Continua		

ndis: pacientes enfermos; nndis: pacientes no enfermos.

9. ANÁLISIS ESTADÍSTICO

Los datos se han descrito mediante los estadísticos más adecuados a la naturaleza y escala de medición de cada variable: frecuencias absolutas y relativas en porcentajes, y media y desviación estándar para variables continuas, o mediana y rango intercuartílico si la distribución de los datos así lo aconsejaba.

Hemos empleado las representaciones gráficas adecuadas a cada tipo de estudio y a las variables representadas: gráfico de líneas, forest plot, pirámide de población y gráficos de calibración.

Para todos los análisis llevados a cabo se han excluido los registros con datos inconsistentes e ilógicos y se ha establecido un nivel de significación estadística del 0,05.

Las bases de datos que se han empleado para realizar los análisis, se han importado al software estadístico STATA SE v.16, el cual ha sido empleado para realizar todos los análisis estadísticos.

MODELOS PREDICTIVOS DE PRONÓSTICO

Se han validado externamente los modelos predictivos incluidos en la RS de modelos predictivos de pronóstico. Se ha calculado la calibración y la discriminación de cada modelo. Y para aquellos modelos en los que los autores habían establecido unos puntos de corte de uso clínico, se ha evaluado la bondad clasificatoria de dichos puntos de corte en nuestra muestra.

Se ha seleccionado el score ANDC como el mejor de entre todos los evaluados. Y posteriormente, se ha realizado una segunda validación externa para dicho score, con los datos de los pacientes de la segunda ola del HUD, para analizar si el rendimiento predictivo era similar al obtenido con la primera validación. Para ello, se han obtenido las medidas de rendimiento de calibración y de discriminación para dicho modelo.

Se ha realizado una comparativa de los pacientes ingresados en el HUD en la primera y en la segunda ola de la pandemia. Para ello, se han analizado las diferencias respecto a sexo, edad y letalidad entre las dos olas. Para el análisis de letalidad se ha llevado a cabo un análisis de supervivencia y se

ha aplicado el test log-rank para analizar las diferencias entre grupos de edad (" >85 años" vs " ≤ 85 años").

Hemos estudiado el uso y el rendimiento del score ANDC en el servicio de urgencias del HUD mediante un estudio descriptivo de todos los pacientes que han acudido al servicio de urgencias durante los meses de enero a febrero de 2021.

PRUEBAS DIAGNÓSTICAS DE LABORATORIO

Para evaluar la validez de las pruebas diagnósticas de laboratorio empleadas en nuestro hospital: RT-PCR, la prueba rápida de antígenos (Panbio™) y la prueba de determinación de anticuerpos (Serología), se han empleado tablas de contingencia 2x2 y se han obtenido la Se, la Sp, el valor predictivo positivo (VPP) y el valor predictivo negativo (VPN).

Para el cálculo de los intervalos de confianza al 95% se han utilizado las tablas científicas, que proporcionan límites exactos de confianza para proporciones para la distribución binomial (J.R. Geigy, Basilea. 1965).

Para estudiar la sensibilidad combinada de los diferentes estudios seleccionados con la RS realizada para la prueba rápida de antígenos Panbio™ hemos llevado a cabo la combinación de estudios, mediante meta-análisis, cuando los datos eran suficientes (≥ 3 estudios) y cuando los estudios eran lo suficientemente homogéneos con respecto a la variable clave: población de estudio. Para que las características de las diferentes poblaciones fueran similares se han analizado, por un lado, pacientes asintomáticos y, por otro, pacientes sintomáticos.

Se han aplicado modelos de efectos aleatorios de DerSimonian-Laird⁽¹⁷⁰⁾ para combinar los estudios y obtener el estimador ponderado. El modelo de efectos aleatorios se ha seleccionado para tener en cuenta la componente de variabilidad entre estudios además de la varianza de cada estudio individual (que es la única variabilidad considerada cuando aplicamos un modelo de efectos fijos).

Se ha empleado el estadístico I^2 para estudiar la heterogeneidad entre los estudios⁽¹⁶⁷⁾. Y se han definido valores de $I^2 < 25\%$ como baja heterogeneidad, valores de 25% a 50% como heterogeneidad moderada, valores de 50% a 75% como heterogeneidad sustancial y valores $> 75\%$ como alta heterogeneidad^(157,171).

SCORES CREADOS EN EL HUD

Para evaluar la capacidad discriminadora del Score de Sospecha Clínica (SSC) hemos obtenido la Se y la Sp, así como el VPP y el VPN para diferentes puntos de corte. Se ha empleado el índice de Youden para seleccionar el punto de corte óptimo.

Para el cálculo de los intervalos de confianza al 95%, se han utilizado las tablas científicas que proporcionan límites exactos de confianza para proporciones para la distribución binomial (J.R. Geigy, Basilea. 1965).

Se ha empleado el test exacto de Fisher para comparar las sensibilidades obtenidas en el SSC para los diferentes grupos de edad establecidos.

Para evaluar la capacidad predictiva del Risk Score se ha llevado a cabo la validación externa del score con los datos de los pacientes de la 1ª ola y se han obtenido las medidas de rendimiento de calibración y de discriminación para dicho modelo.

10. CONSIDERACIONES ÉTICAS

Para el desarrollo del estudio se solicitó al Comité ético de investigación clínica (CEIC) del área sanitaria de Gipuzkoa la aprobación del proyecto de investigación y la exención del consentimiento informado debido a la naturaleza del estudio.

El uso de los datos de los participantes y el procesamiento de los datos se realizó de acuerdo con la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales que deroga la Ley Orgánica 15/1999, de 5 de diciembre, de protección de datos personales.

Los únicos datos personales relevantes de los pacientes en este estudio son la edad y el sexo, sin interesar su nombre ni sus apellidos u otros datos identificativos. En todo momento se ha asegurado el anonimato de los pacientes, preservando los datos identificativos de los pacientes separados de los datos clínicos, como se indica en el artículo 16.3 de la ley 41/2002, de 14 de noviembre, básica reguladora de la autonomía del paciente y de derechos y obligaciones en materia de información y documentación clínica. En ningún momento se publicarán los datos personales.

Para la realización del estudio, a cada paciente se le asignó un *identificador de paciente* (Id o código ficticio que no permitiera su identificación), que fue recogido en las bases de datos. La persona responsable de la base de datos ha sido la autora de la tesis (MI) y la ventana de tiempo en la que se ha accedido a la historia clínica de los pacientes ha sido el tiempo suficiente para extraer los datos de los períodos: (16/3/2020–17/5/2020), (1/8/2020–15/11/2020) y (1/1/2021–28/2/2021).

RESULTADOS

RESULTADOS

- DESCRIPCIÓN DE LOS PACIENTES DE LA OSI-DONOSTIALDEA
 - Características de los pacientes ingresados en el HUD en la 1ª ola.
 - Características de los pacientes COVID-19 del HUD en la 1ª ola.
- MODELOS PREDICTIVOS DE DIAGNÓSTICO Y DE PRONÓSTICO
 - Modelos predictivos de diagnóstico.
 - Modelos predictivos de pronóstico.
- CARACTERIZACIÓN RUEBAS DIAGNÓSTICAS DE LABORATORIO
 - Prueba RT-PCR.
 - Prueba rápida de antígenos PANBIO™ (Ag Rapid test Device).
 - Prueba de determinación de anticuerpos (Serología).
- SCORES CREADOS EN EL HUD
 - Score de sospecha clínica (SSC).
 - Risk score.

1. DESCRIPCIÓN DE LOS PACIENTES DE LA OSI-DONOSTIALDEA.

Se registraron prospectivamente y de manera consecutiva a todos los pacientes, mayores de 18 años, que fueron ingresados en el HUD durante el período comprendido entre el 16 de marzo hasta el 17 de mayo de 2020. El número total de pacientes ingresados fue de 1.712.

Se cumplimentó el formulario COVID-19 a 987 pacientes que eran sospechosos de padecer la enfermedad y que acudieron al HUD en dicho período (99 no fueron ingresados).

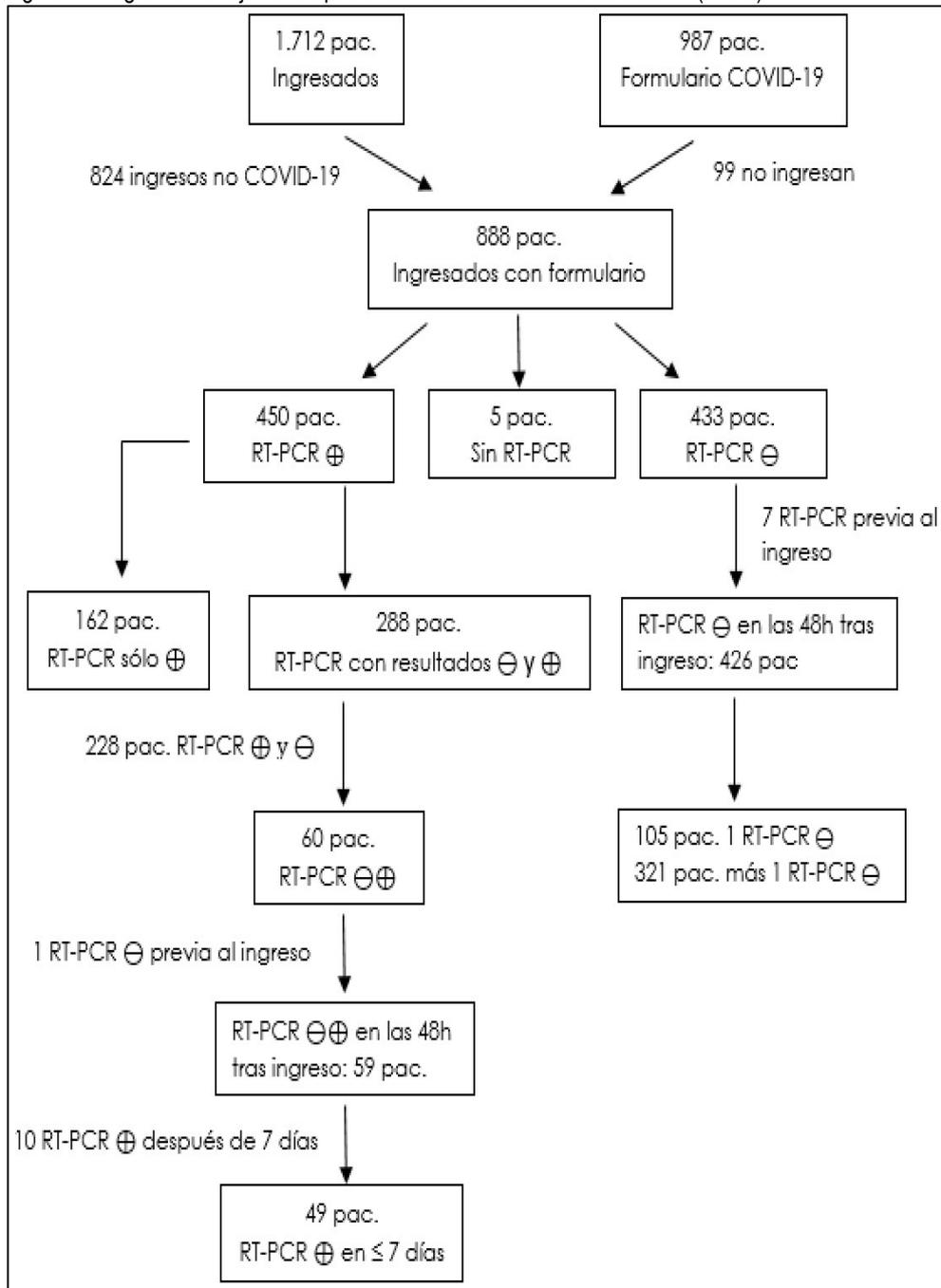
En total 888 pacientes tenían cumplimentado el formulario COVID-19 y fueron ingresados. Para 5 de ellos no se pudo encontrar un resultado de RT-PCR, 450 tuvieron un resultado positivo en al menos una de las pruebas RT-PCR que se les realizaron durante su ingreso y 433 pacientes tuvieron resultado negativo (que indica que todas las RT-PCR que se les realizaron durante el ingreso fueron negativas).

De los 450 pacientes con alguna RT-PCR positiva, 162 obtuvieron sólo resultados positivos en las RT-PCR que se les realizaron y 288 resultados tanto positivos como negativos.

De los 433 pacientes que sólo obtienen resultados negativos en las pruebas RT-PCR realizadas, 426 de ellos tienen esa primera RT-PCR negativa en las 48h tras el ingreso. A 105 de ellos sólo se les realizó una prueba RT-PCR y al resto de pacientes, 321, se les realizaron más de 1 prueba RT-PCR durante su ingreso.

De los 288 pacientes que obtuvieron RT-PCR con resultados tanto positivos como negativos en las pruebas RT-PCR que se les realizaron, en 60 de ellos ocurre que la primera RT-PCR es negativa y, posteriormente, se obtiene alguna RT-PCR positiva (es decir, que la persona ingresa dando un resultado negativo, pero se positiviza posteriormente). Consideramos para este análisis que la primera RT-PCR negativa lo sea en las primeras 48h tras el ingreso, para evitar así tener RT-PCR extrahospitalarias. De este modo, nos quedamos con 59 pacientes. De éstos, 49 tienen una RT-PCR positiva en la primera semana tras haber obtenido el resultado negativo en la primera RT-PCR intrahospitalaria. Y los otros 10 pacientes tuvieron una RT-PCR positiva pero después de 7 días tras el ingreso. Un resumen de estos datos se puede ver en el siguiente diagrama de flujo (Figura 9).

Figura 9. Diagrama de flujo de los pacientes considerados en este estudio (1ª ola).



CARACTERÍSTICAS DE PACIENTES INGRESADOS EN EL HUD EN LA 1ª OLA

Hemos recogido datos de algunos de los signos y síntomas que consideramos de interés. En la tabla 23 se hace referencia a si el paciente ha tenido o no, en algún momento del ingreso, los signos y síntomas indicados. Desglosamos la información en función de si son pacientes COVID-19 (los que tienen al menos un resultado RT-PCR positivo) o no COVID-19 (los que no tienen ninguna RT-PCR positiva).

Tabla 23. Características seleccionadas de los pacientes ingresados en el HUD en la 1ª ola.

Características % (n)	COVID-19 (n=450)	No COVID-19 (n=433)	Población Total (n=883)
Edad (años) \bar{x} ,SD (n)	66.25, 13.84 (435)	69.63, 16.80 (419)	67.88, 15.42 (859)
Hombres	58.39 (435)	52.27 (419)	55.41 (859)
Diabetes mellitus	18.00 (450)	18.94 (433)	18.46 (883)
Enferm. cardiovascular	29.40 (449)	43.95 (430)	36.52 (879)
EPOC	16.00 (450)	28.67 (429)	22.18 (879)
Enferm. renal crónica	9.78 (450)	16.86 (433)	13.25 (883)
HTA	38.26 (447)	46.60 (427)	42.33 (874)
Obesidad	11.78 (450)	8.31 (433)	10.08 (883)
Factores de riesgo	46.05 (228)	47.8 (318)	47.07 (546)
Fiebre	76.00 (450)	57.67 (430)	67.05 (880)
Disnea	54.14 (447)	60.79 (431)	57.40 (878)
Tos	67.26 (449)	58.56 (432)	63.00 (881)
NYHA			
I	33.85 (130)	29.92 (111)	32.92 (241)
II	37.50 (144)	29.65 (110)	33.64 (254)
III	18.49 (71)	21.83 (81)	20.13 (152)
IV	10.16 (39)	18.60 (69)	14.30 (108)
Expectoración	26.29 (426)	25.56 (399)	25.94 (825)
Mialgias	40.04 (447)	22.70 (423)	31.61 (870)
Dolor torácico	13.39 (448)	17.29 (428)	15.30 (876)
Diarrea	21.23 (438)	12.89 (419)	17.15 (857)
Vómitos	6.71 (447)	8.33 (432)	7.51 (879)
Insufic. respiratoria	37.39 (444)	24.21 (409)	31.07 (853)
Tumor	10.11 (445)	17.14 (426)	13.55 (871)
Leucemia	1.36 (442)	1.64 (426)	1.50 (868)
Linfoma	1.82 (439)	2.33 (426)	2.08 (865)
Éxitus	13.59 (449)	15.01 (433)	14.27 (882)
Críticos	12.44 (450)	2.08 (433)	7.59 (883)

EPOC: enfermedad pulmonar obstructiva crónica; HTA: hipertensión arterial; NYHA: clasificación funcional New York Heart Association.

En esta tabla no aparecen dos síntomas que han sido importantes en esta pandemia, ya que dos tercios de las personas infectadas por el SARS-CoV-2 han referido padecerlos. Dichos síntomas son la anosmia (falta de olfato) y la ageusia (falta del gusto). El motivo por el que no se han incorporado es porque, en la primera ola, la anosmia se empezó a registrar a mediados del mes de abril, no desde el 16 de marzo que es el período al que están referidas el resto de las variables registradas en esta tabla. Y la ageusia no se ha registrado en ningún momento de la 1ª ola.

CARACTERÍSTICAS DE LOS PACIENTES COVID-19 DEL HUD DE LA 1ª OLA

Describimos en la tabla 24 las características de los pacientes COVID-19 desglosados en función de si los pacientes fallecieron o no.

El porcentaje de hombres fallecidos fue menor que el de mujeres y la edad de los fallecidos fue, en promedio, 14 años mayor que la edad de los que no fallecieron. Así mismo, la enfermedad cardiovascular está más presente en los fallecidos (aproximadamente en el 50%), al igual que la diabetes mellitus (26.23% vs 16.75%), la enfermedad renal crónica (16.39% vs 8.76%), la HTA (55.74% vs 35.05%), los factores de riesgo (29.5% vs 22.42%) y la insuficiencia respiratoria (65.75% vs 32.22%), además presentan más tumores (18.03% vs 8.76%). En el grupo de pacientes fallecidos se toman valores mayores en la escala NYHA y menores en el grupo de pacientes que no fallecieron.

Hay un porcentaje menor de pacientes obesos (8.2% vs 12.37%) y EPOC (13.11% vs 16.49%) en los fallecidos respecto a los que no fallecen. Lo mismo sucede con la expectoración (18.03% vs 26.03%), las mialgias (27.87% vs 41.75%) y el dolor torácico (8.2% vs 14.18%).

En cuanto a la fiebre, tos y disnea se producen en un alto porcentaje de pacientes, tanto en el grupo de los fallecidos como en el de los que no fallecen.

Nota: hay un paciente para el que no pudimos averiguar si estaba vivo o había fallecido, por este motivo la población total es de 449 pacientes y no de 450.

Tabla 24. Características seleccionadas de los pacientes COVID-19 ingresados (HUD) en la 1ª ola.

Características % (n)	Éxito (n=61)	Vivo (n=388)	Población Total (n=449)
Edad (años) \bar{X} ,SD	78.62, 10.4	64.52, 13.4	66.25, 13.84
Hombres	25 (40.78)	228 (58.76)	253 (56.35)
Diabetes mellitus	16 (26.23)	65 (16.75)	81 (18.00)
Enferm. cardiovascular	28 (45.90)	103 (26.55)	131 (29.18)
EPOC	8 (13.11)	64 (16.49)	72 (16.00)
Enferm. renal crónica	10 (16.39)	34 (8.76)	44 (9.78)
HTA	34 (55.74)	136 (35.05)	170 (37.86)
Obesidad	5 (8.20)	48 (12.37)	53 (11.80)
Factores de riesgo	18 (29.51)	87 (22.42)	105 (23.39)
Fiebre	46 (75.41)	295 (76.03)	341 (75.95)
Disnea	38 (62.30)	204 (52.58)	242 (53.90)
Tos	36 (59.02)	266 (68.56)	302 (67.26)
NYHA			
I	12 (19.67)	117 (30.15)	129 (28.73)
II	12 (19.67)	132 (34.02)	144 (32.07)
III	13 (21.31)	58 (14.95)	71 (15.81)
IV	18 (29.51)	21 (5.41)	39 (8.69)
Expectoración	11 (18.03)	101 (26.03)	112 (24.94)
Mialgias	17 (27.87)	162 (41.75)	179 (39.87)
Dolor torácico	5 (8.20)	55 (14.18)	60 (13.36)
Diarrea	6 (9.84)	87 (22.42)	93 (20.71)
Vómitos	8 (13.11)	22 (5.67)	30 (6.68)
Insuficiencia respiratoria	40 (65.57)	125 (32.22)	165 (36.75)
Tumor	11 (18.03)	34 (8.76)	45 (10.02)
Leucemia	2 (3.28)	3 (0.77)	5 (1.11)
Linfoma	2 (3.28)	8 (2.06)	10 (2.22)
Críticos	13 (21.31)	44 (11.34)	57 (12.69)

EPOC: enfermedad pulmonar obstructiva crónica; HTA: hipertensión arterial; NYHA: clasificación funcional New York Heart Association.

2. MODELOS PREDICTIVOS DE DIAGNÓSTICO Y DE PRONÓSTICO.

A continuación, mostramos los resultados de las dos RS realizadas. En primer lugar, la RS de los modelos predictivos de diagnóstico y, posteriormente, los resultados de la RS de modelos predictivos de pronóstico.

2.1. MODELOS PREDICTIVOS DE DIAGNÓSTICO.

En el diagrama de flujo (Figura 10) se muestran los resultados de la búsqueda bibliográfica realizada para identificar modelos predictivos diagnósticos que permitan determinar qué pacientes, de los que presentan sintomatología compatible con el SARS-Cov2, están realmente infectados.

Tras examinar el título y el abstract de los estudios identificados mediante las búsquedas realizadas, se excluyeron: 135 (78%) estudios de los 173 identificados por Medline y 118 (84.3%) de los 140 estudios recuperados por Embase por no ser pertinentes.

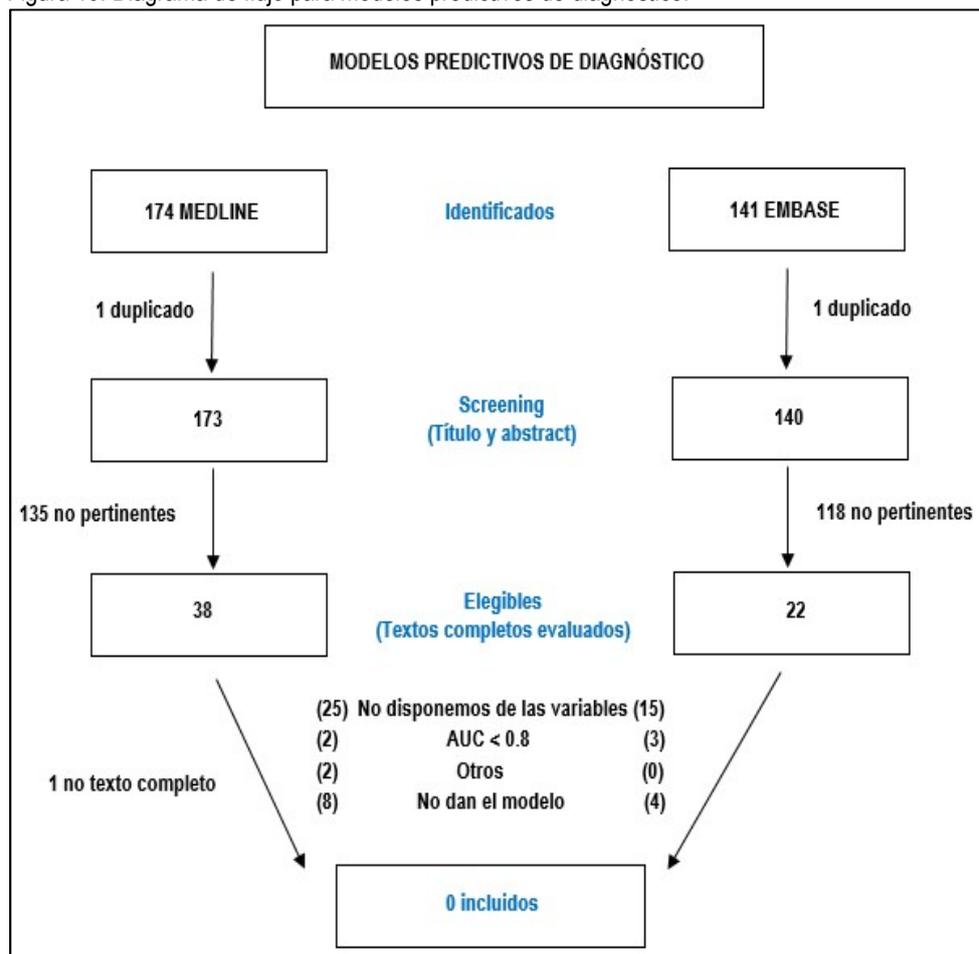
El motivo por el que se excluyeron esos estudios fue porque no consideraban nuestro outcome de interés: establecían modelos para predecir cómo se iba a propagar la infección (modelos SIR o SEIR) o bien predecían qué pacientes iban a acudir a las urgencias del hospital o iban a ser ingresados. Se identificaron estudios cuyo objetivo era detectar la enfermedad, pero se basaban en analizar la IgG y la IgM de las pruebas diagnósticas de determinación de anticuerpos (Serología) y, por lo tanto, se referían a infección pasada y no a infección actual, que era nuestro objetivo.

De los estudios elegibles: 38 con la búsqueda de Medline y 22 con la Embase, un alto porcentaje de ellos fueron descartados, porque no disponíamos de las variables que se necesitaban para validar el modelo (25 de Medline y 15 de Embase), bien porque no habían sido recogidas (código postal, VIH, etc.) o porque no teníamos acceso a ellas (pruebas de imagen). 12 de 38 para Medline y 7 de 22 para Embase, o no daban el modelo o tenían un AUC < 0.8 o se clasificaron en la categoría "otros". Uno de los estudios de la categoría "otros" era una carta al editor y el otro estudio de esa categoría se excluyó porque no estaba bien desarrollado (los autores no daban la ecuación del modelo sólo la puntuación otorgada a cada una de las variables incluidas en el modelo -1 pto cada una -, y a su vez, reportaron las OR de dichas variables, cuyos valores oscilaban entre 1 y 33, por lo que no era creíble la puntuación asignada a cada variable del score).

En los 12 estudios en los que no se proporcionaba el modelo, los autores daban qué variables eran importantes para el diagnóstico de la enfermedad, pero no daban la ecuación de regresión para poder reproducir dicho modelo. En ocasiones, sólo mostraban las OR de las variables implicadas, pero no era posible obtener el modelo completo porque no daban la constante.

Concluimos que, en la RS realizada para modelos predictivos de diagnóstico, finalmente, ningún estudio cumplió los criterios para ser incluido y, por lo tanto, no se pudo llevar a cabo la validación externa de ninguno de ellos con los datos de los pacientes de nuestra OSI.

Figura 10. Diagrama de flujo para modelos predictivos de diagnóstico.



2.2. MODELOS PREDICTIVOS DE PRONÓSTICO.

En la figura 11 se muestra el diagrama de flujo con los estudios identificados mediante la búsqueda bibliográfica realizada para encontrar modelos predictivos de pronóstico que permitan conocer qué pacientes, de los que se han infectado por el SARS-Cov2, presentan mayor riesgo de sufrir un peor pronóstico o un desenlace fatal, es decir, quiénes van a necesitar ingresar en la UCI o van a fallecer.

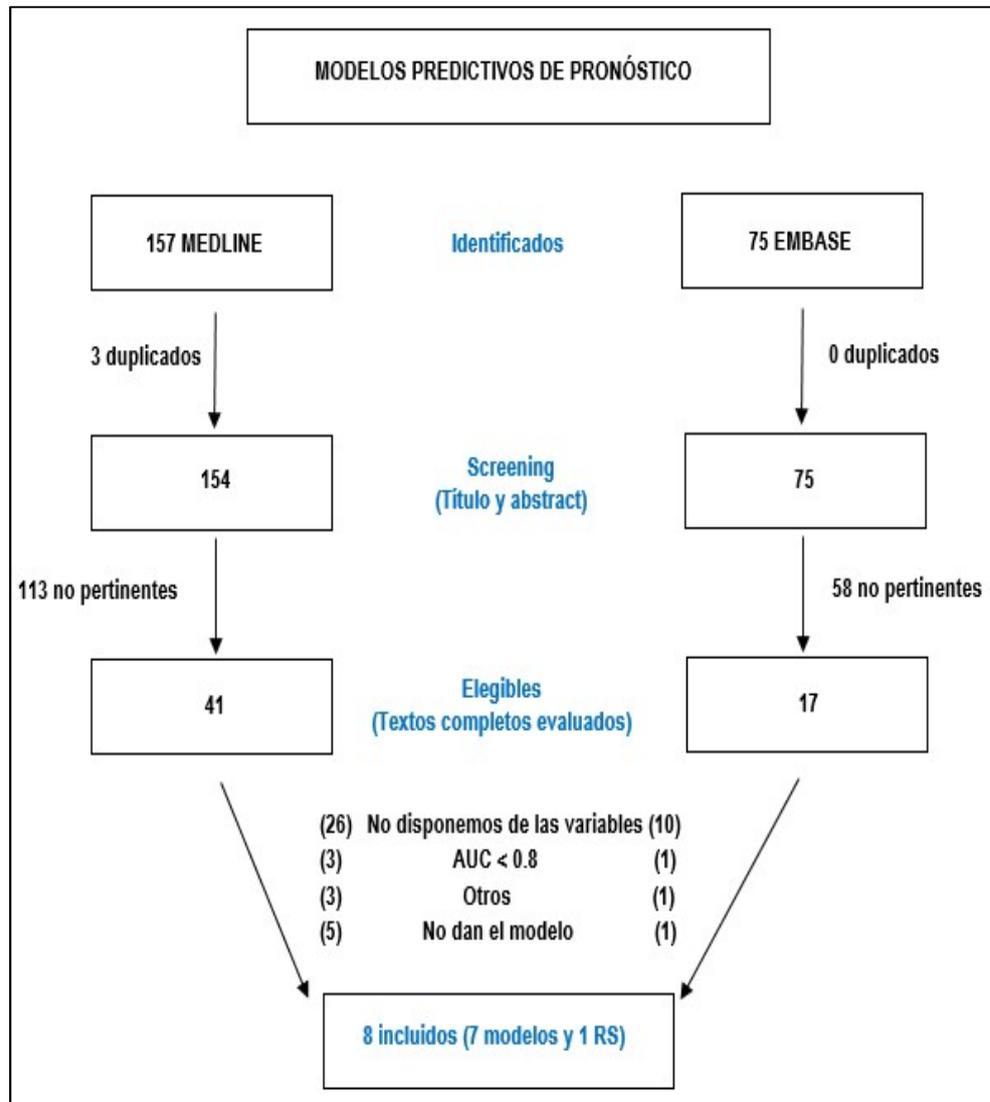
De los 232 estudios identificados mediante la búsqueda bibliográfica, sólo 8 fueron finalmente seleccionados para la posterior validación externa con los datos de pacientes de nuestra OSI obtenidos en la 1ª ola (Tabla 25). De los 8 estudios, 7 son para modelos predictivos de pronóstico y 1 estudio es una RS⁽¹³⁰⁾ que comprende estudios tanto de pronóstico como de diagnóstico.

En esta RS, que se corresponde con la última actualización de la Living Systematic Review que está realizando el equipo COVID-PRECISE, los revisores propusieron dos modelos para que fuesen validados externamente en otras poblaciones: uno para diagnóstico y otro para pronóstico. El modelo propuesto para diagnóstico fue el desarrollado por Lara Jehi et al^(131,132) que no pudo ser reproducido con los datos de nuestra población porque no disponíamos de algunas de las variables que conformaban dicho modelo (ej: raza, etnia, geolocalización,..), ya que no habían sido recogidas en nuestra OSI. Y el modelo propuesto para pronóstico fue el desarrollado por Stephen R. Knight et al⁽¹³³⁾, denominado score de mortalidad 4C, del que hablaremos posteriormente.

Del mismo modo que en la RS realizada para modelos predictivos de diagnóstico, en la RS realizada para seleccionar modelos predictivos de pronóstico, se excluyeron un alto porcentaje de los estudios elegibles porque no disponíamos de los predictores que se empleaban en dichos modelos. En este caso, el porcentaje asciende a un 62%, de los cuales, el 41.4% corresponden a modelos basados en pruebas de imagen y el resto a otros predictores. Un 10% de los estudios elegibles no proporcionaban el modelo y, por lo tanto, no permitían la reproducibilidad del mismo.

De la búsqueda realizada en Embase, 4 estudios fueron seleccionados finalmente: PREDI-CO⁽¹⁷²⁾, Living Systematic Review⁽¹³⁰⁾, HNC-LL⁽¹⁷³⁾ y Score ANDC⁽¹⁷⁴⁾. Y de la búsqueda realizada en Medline se seleccionaron otros 4 estudios: CSS⁽¹⁷⁵⁾, el Score de Zhang⁽¹⁷⁶⁾, el NEWS2⁽¹⁵⁶⁾ y el PRE Score⁽¹⁷⁷⁾, tal y como se muestra en la Figura 11.

Figura 11. Diagrama de flujo para modelos predictivos de pronóstico.



2.2.1. Resumen de los modelos predictivos incluidos en la RS.

En la tabla 25 se muestra un resumen de los estudios seleccionados para realizar la validación externa con los datos de pacientes de la primera ola de nuestra OSI.

Tabla 25. RPC para pronóstico seleccionadas para ser validadas en nuestra población de estudio.

Autor	Lugar	Score	Outcome
Knight	Reino Unido	4C score mortality	Mortalidad
Bartoletti	Italia	PREDI-CO	IRS (Críticos)
Weng	Wuhan (China)	ANDC	Mortalidad
Zhang	Wuhan (China)	*	Mortalidad
Shang	Wuhan (China)	CSS	Mortalidad en críticos
Liu	Wuhan (China)	PRE	Mortalidad en críticos
Xiao	Jiangxi (China)	HNC-LL	Severidad (Críticos)
Gidari	Italia	NEWS2	Ingreso en críticos

(*) No asignan ningún nombre al score; IRS= Insuficiencia respiratoria severa.

Cabe mencionar que, para llevar a cabo las validaciones externas, se ha considerado como población de estudio a los 450 pacientes COVID-19 que obtuvieron un resultado positivo en alguna de las pruebas diagnósticas de laboratorio RT-PCR que se les realizaron durante su ingreso en el HUD en la primera ola. Destacamos que, debido al elevado número de datos faltantes en algunas de las variables empleadas en las RPC, no se pudieron validar las reglas con el total de los pacientes.

A continuación, explicamos los scores validados externamente.

- *4C Mortality score*⁽¹³³⁾: es el score de Mortalidad 4C (Coronavirus Clinical Characterisation Consortium) que fue desarrollado por el consorcio ISARIC (International Severe Acute Respiratory and emerging Infections Consortium) y en el que participaron 260 hospitales de Inglaterra, Escocia y Gales. El consorcio ISARIC está formado por médicos y científicos del Reino Unido comprometidos con responder preguntas urgentes sobre COVID-19 de manera rápida, abierta y para el beneficio de todos.

El score 4C está formado por 8 variables: edad, sexo, frecuencia respiratoria, número de comorbilidades, saturación de oxígeno, nivel de conciencia, urea y proteína C reactiva (PCR).

Los autores compararon el rendimiento del 4C score con otros scores: SOFA, qSOFA, SMARTCOP, Surgisphere, SCAP, NEWS, DL score, CRB-65, COVID-GRAM, CURB-65, DS-CRB65, Xie SCORE, A-DROP, PSI Y E-CURB65 y todos obtuvieron un rendimiento menor que el obtenido con el score 4C. El AUC obtenida para la cohorte de derivación (n=35.463) fue 0.79 (IC95%: 0.78-0.79) y para la cohorte de validación (n=22.361) de 0.77 (IC95%: 0.76-0.78) con una excelente calibración. El AUC para la cohorte de validación externa, con los datos de la primera ola de nuestra OSI, fue 0.70 (IC95%: 0.59-0.80).

Nota: a pesar de tener un AUC<0.8 (criterio de exclusión) tanto en la derivación como en la validación, se ha considerado este score para llevar a cabo la validación externa con nuestros datos, dado que la diferencia de 0.79 a 0.80 era mínima y a que era el único score para pronóstico propuesto por el grupo COVID-PRECISE en la *Living Systematic Review*.

- *PREDI-CO score*⁽¹⁷²⁾: fue desarrollado por Bartoletti et al., para predecir insuficiencia respiratoria severa (IRS) en pacientes COVID-19. En este estudio multicéntrico participaron 11 hospitales de Italia, con un total de 1.113 pacientes diagnosticados con COVID-19.

Las variables que componen el score son 8: edad, obesidad, temperatura, frecuencia respiratoria, linfocitos, creatinina, PCR y lactato deshidrogenasa (LDH).

Los autores compararon el rendimiento del PREDI-CO score con otros scores: SOFA, qSOFA, CURB-65 y el MEWS. Todos obtuvieron un rendimiento bastante más bajo que el score propuesto por los autores. El PREDI-CO score proporcionó un AUC=0.89 con un (IC95%: 0.86-0.92) para la cohorte de derivación (n=644) y fue similar en la cohorte de validación (n=469) con un AUC=0.85 (IC95%: 0.81-0.88). Para la validación externa con nuestros pacientes hemos considerado el ingreso en UCI, dado que no disponíamos de la variable IRS. Y hemos obtenido un AUC=0.64 (IC95%: 0.56-0.71).

- *Score de riesgo ANDC*⁽¹⁷³⁾: en el estudio desarrollado por Weng et al⁽³¹⁾ participan 2 centros, Main District y Tumor Center, del Union Hospital de la Universidad de Ciencia y tecnología de Huazhong en Wuhan (China). Se

propone el score ANDC, que es un score de alerta temprana para predecir el riesgo de mortalidad por COVID-19.

El ANDC está formado por 4 predictores: edad, dímero D, PCR y la ratio de Neutrófilos/linfocitos.

El modelo ha demostrado una buena calibración y discriminación con un AUC en la cohorte de derivación (n=176) de 0.92 (IC95%: 0.84-0.97) y AUC=0.98 con (IC95%: 0.95-1.00) en la cohorte de validación (n=125). El AUC para la validación externa, con los datos de la primera ola de nuestra OSI, fue 0.82 (IC95%: 0.77-0.87).

- *Zhang score*⁽¹⁷⁴⁾: es un score creado por Zhang et al para predecir mortalidad a los 14 días y a los 28 días tras el ingreso (nosotros hemos considerado mortalidad intrahospitalaria, sin tener en cuenta el momento en el que se produce el fallecimiento). Este score fue desarrollado en Wuhan seleccionando datos de dos hospitales (*Union hospital* y *Central Hospital*).

El score está formado por las siguientes 4 variables: edad, LDH, ratio de neutrófilos/linfocitos y bilirrubina directa.

El score se derivó con 516 pacientes del Union Hospital de Wuhan obteniendo un C-Statistic=0.89 con un (IC95%: 0.87-0.90). Los autores seleccionaron 186 pacientes del mismo hospital para la validación y obtuvieron un C-Statistic=0.88 con (IC95%: 0.86-0.90). Llevaron a cabo una segunda validación con un total de 126 pacientes del Central Hospital de Wuhan con C-Statistic=0.84 con (IC95%: 0.80-0.88). El resultado obtenido para la validación externa en nuestra cohorte fue de 0.74 (IC95%: 0.66-0.81).

Los autores compararon el rendimiento del score con otros scores: qSOFA y CURB-65 pero ambos obtuvieron un rendimiento bastante más bajo que el score propuesto por los autores, siendo C-Statistic=0.67 y C-Statistic=0.78 los resultados obtenidos respectivamente.

- *CSS score*⁽¹⁷⁵⁾: este score (*Scoring System of COVID-19*) fue derivado en Wuhan por Shang et al. Permite predecir riesgo de fallecer en pacientes COVID-19 ingresados en la unidad de críticos. Se recogieron datos

completos de 1.830 pacientes, con 452 casos de pacientes severos, que fueron hospitalizados en el hospital Zhongnan, en el hospital de Leishenshan o en otros 7 hospitales de Wuhan.

El score está formado por 5 variables: Edad, enfermedad coronaria, linfocitos (%), Procalcitonina y Dímero D.

Para la cohorte de derivación se consideraron 113 pacientes críticos y se obtuvo una buena discriminación $AUC=0.92$ (IC95%: 0.87-0.97) y calibración ($p=0.264$). 339 pacientes fueron incluidos en la cohorte de validación y se obtuvo un $AUC=0.94$ (IC95%: 0.90-0.97). Nosotros hemos validado la RPC para predecir mortalidad en todos los pacientes infectados por la COVID-19 que estaban ingresados en el HUD durante la 1ª ola, no sólo en aquellos pacientes críticos, ya que no disponíamos de datos para algunas de las variables del score en los pacientes críticos que habían fallecido. El AUC para la cohorte de validación externa, con los datos de la primera ola de nuestra OSI, fue 0.83 (IC95%: 0.77-0.87).

- *PRE score*⁽¹⁷⁶⁾: este score fue desarrollado por Liu et al para predecir mortalidad en pacientes críticos. Reclutaron 336 pacientes de dos pabellones del Union Hospital de Wuhan (Tongji Medical College y Huazhong University of Science and Technology).

El PRE score está formado únicamente por 3 variables: BUN, linfocitos (%) y Dímero D.

Obtuvieron un $AUC=0.99$ (IC95%: 0.98-1) en la muestra de derivación. Para la validación externa con nuestros datos seleccionamos pacientes ingresados en la unidad de críticos, pero la muestra se redujo a 21 pacientes. Se obtuvo un $AUC=0.85$ (IC95%: 0.64-0.97). Dado el número tan reducido de pacientes, decidimos validar la RPC con todos los pacientes (críticos y no críticos) y obtuvimos un $AUC=0.79$ (IC95%: 0.74-0.84).

- *HNC-LL score*⁽¹⁷⁷⁾: es un score que permite predecir severidad en pacientes COVID-19. Fue derivado y validado por Xiao et al, en la provincia de Jiangxi (sureste de China).

El score está formado por las siguientes 5 variables: neutrófilos, linfocitos, PCR, LDH e hipertensión.

Se derivó (n=231) y validó internamente (n=101) con pacientes del Honghu Hospital. Y se obtuvieron valores de AUC=0.86 (IC95%: 0.8–0.92) y AUC=0.87 (IC95%: 0.77–0.97) respectivamente. Se realizó una validación externa con pacientes del Nanchang Hospital (n=110) con un AUC=0.83 (IC95%: 0.75–0.91).

Los autores han comparado este score con el MuLBSTA, CURB-65 y la ratio neutrófilos/linfocitos obteniendo, en la cohorte de validación Nanchang los siguientes resultados: AUC=0.81 con un IC95%: (0.73–0.89), AUC=0.64 con un IC95%: (0.63–0.75) y AUC=0.78 con un IC95%: (0.69–0.88) respectivamente. En la validación externa, con los pacientes de nuestra OSI, obtuvimos un AUC=0.55 con un IC95%: (0.43–0.68).

- NEWS2⁽¹⁴⁷⁾ (National Early Warning Score 2) es un score de alerta temprana que permite predecir deterioro a las 24h tras el ingreso. No fue creado expresamente para la COVID-19 aunque se ha considerado en esta tesis por ser el que ha proporcionado mejores resultados de entre todos los que se han validado en pacientes COVID-19 pero que no fueron creados para esta enfermedad.

El score NEWS2 está formado por variables: frecuencia respiratoria, saturación de oxígeno, presión arterial sistólica, frecuencia cardíaca, estado de conciencia y temperatura.

Gidari et al⁽¹⁵⁶⁾, validaron el NEWS2 en pacientes COVID-19 para predecir ingresos en la unidad de críticos y obtuvieron un AUC=0.90 (IC95%: 0.82–0.97). La validación externa, con los datos de la primera ola de nuestro hospital, se llevó a cabo considerando los datos recogidos al ingreso. Con estos datos, se obtuvo un AUC de 0.71 (IC95%: 0.61–0.80).

En la tabla 26, se resume la información de las RPC validadas con los datos de los pacientes de nuestra OSI. De izquierda a derecha vemos, en primer lugar, la RPC o Score que se valida. A continuación, el resultado (outcome) que se quiere predecir (mortalidad o ingreso en UCI). Tras el resultado evaluado, vemos el número de pacientes que se han tenido en cuenta para validar las reglas (de los 450 que tuvieron RT-PCR positiva). A continuación, las variables que componen cada score, con el AUC y el IC95% resultante de la validación en nuestra población. En color gris, vemos los valores obtenidos en la población donde se derivó y se validó la RPC.

Tabla 26. Resumen de las RPC validadas externamente en el HUD con datos de la 1ª ola.

Score	Resultado	n	Variables	AUC (IC95%)	n (D)	AUC (IC95%)	n (V)	AUC (IC95%)
4C Mortality	Éxito	78	Edad, Sexo, N° comorb. FR, Sat O ₂ , Nivel conciencia, Urea, PCR	0.70 (0.59-0.80)	35463	0.79 (0.78-0.79)	22361	0.77 (0.76-0.77)
PREDICO	IR severa (ingreso en críticos)	161	Edad, Obesidad, T ^a , FR, Linfocitos, Creatinina, PCR, LDH	0.64 (0.56-0.71)	644	0.89 (0.86-0.92)	469	0.85 (0.81-0.88)
ANDC	Éxito	247	Edad, Ratio N/L, Dímero D, PCR	0.82 (0.77-0.87)	176	0.92 (0.83-0.97)	125	0.97 (0.94-1)
Zhang et al	Éxito	134	Edad, Ratio N/L, LDH, Bilirrubina directa	0.74 (0.66-0.81)	516	0.89 (0.87-0.90)	126	0.84 (0.80-0.88)
CSS	Éxito	238	Edad, Enf. Coronaria, linfocitos (%), Procalcitonina, Dímero D.	0.83 (0.77-0.87)	113	0.92 (0.87-0.97)	339	0.94 (0.90-0.97)
PRE	Éxito	258	BUN, Dímero D, Linfocitos (%)	0.79 (0.74-0.84)	336	0.99 (0.98-100)	-	-
HNC-LL	Ingreso en críticos	257	HTA, linfocitos, neutrófilos, PCR, LDH	0.55 (0.43-0.68)	231	0.86 (0.80-0.92)	110	0.83 (0.75-0.91)
NEWS2	Ingreso en críticos	100	FR, FC, Sat O ₂ , PAS, Nivel Conc.; T ^a	0.74 (0.64-0.82)	-	-	68	0.90 (0.82-0.97)

IRS: insuficiencia respiratoria severa; FR: frecuencia respiratoria; Sat O₂: saturación de oxígeno; PCR=Proteína C reactiva; T^a=Temperatura; LDH=lactato deshidrogenasa; BUN=nitrógeno ureico en la sangre; HTA: hipertensión arterial; FC: frecuencia cardíaca; PAS: presión arterial sistólica; n(D)=n° de individuos empleados para la derivación de la regla, n(V)=n° de individuos empleados para la validación de la regla. n=n° de individuos empleados en la validación de la RPC con los datos de la OSI.

Llama la atención que en 2 (PRE, HCN-LL) de los 7 modelos derivados para COVID-19 no se considere la edad en el modelo final. La edad es el predictor más relevante para predecir mortalidad y determina, en gran medida, el desenlace del paciente COVID-19, por lo que resulta llamativo que no haya sido incluido en los modelos finales de esos dos scores. Aunque, en el estudio del PRE score, indican que ellos se centran en estudiar mortalidad en pacientes severos y creen que el efecto de la edad en ese grupo es relativamente pequeño y, por eso, no es un predictor del modelo final. Sin embargo, en el HNC-LL no se menciona nada al respecto, lo cuál es curioso porque en el análisis univariante la edad tiene un OR=4.207 aunque no sale significativo en el análisis multivariante.

Además de la edad, los parámetros más comunes en los modelos predictivos de pronóstico para COVID-19 han sido: dímero D, linfocitos, saturación de oxígeno, proteína C reactiva (PCR) y plaquetas (aunque no aparece en ninguno de los modelos que hemos incluido en esta revisión). Otros parámetros menos comunes han sido: lactato deshidrogenasa (LDH), ratio neutrófilos/linfocitos (NLR), la procalcitonina (PCT) y la urea. Todos ellos y, algunos más, están contemplados en los modelos que hemos analizado. Siendo el ANDC y el PRE los que incluyen los parámetros más habituales.

El rendimiento predictivo obtenido en la validación externa está condicionado por diferentes factores, entre ellos, la prevalencia del outcome en la población validada. En este caso, las prevalencias de los outcomes analizados, en las poblaciones donde se derivan las reglas, son diferentes a las prevalencias de dichos outcomes en nuestra población. Lo detallamos en la tabla 27.

Tabla 27. Prevalencias de las RPC validadas externamente y de los datos de la 1ª ola del HUD.

Score	Prevalencia HUD	Prevalencia (D)
4C Mortality	7.7%	32.2%
PREDI-CO	5%	29%
ANDC	11%	11.9%
Zhang	17.2%	18.6%
CSS	10.9%	43.4%
PRE	17.7%	10.1%
HNC-LL	8,7%	17.3%
NEWS2	11%	38%

(D)= población de derivación de la regla

Sólo en el caso del score ANDC y el Zhang score, las prevalencias son casi iguales. En el caso del estudio de Shang (score CSS), las prevalencias son muy distintas, aunque obtenemos un buen rendimiento del score en la muestra de validación. Los autores del score 4C de mortalidad, indican que la RPC puede no funcionar bien en aquellas poblaciones con bajas prevalencias de mortalidad, como nos sucede a nosotros (7.7%), ya que la prevalencia en la población de derivación es del 32.2%. Además, nuestra población de validación únicamente está formada por 78 pacientes.

2.2.2. Evaluación de riesgo de sesgos (PROBAST).

Hemos llevado a cabo la evaluación del riesgo de sesgos (RoB), mediante la herramienta PROBAST, para cada uno de los estudios incluidos en la RS. Como puede verse en la tabla 28, todos los estudios tienen alto riesgo de sesgo en el 4º dominio, es decir, en el dominio de *Análisis*, salvo el 4C score. Esto es debido, principalmente, a que hay pasos dentro del procedimiento de derivación o validación de los modelos que se llevan a cabo de manera incorrecta.

Tabla 28. Evaluación del riesgo de sesgos con PROBAST en los estudios seleccionados.

ESTUDIO	RoB				APLICABILIDAD			GLOBAL	
	PARTICIPANTES	PREDICTORES	OUTCOME	ANÁLISIS	PARTICIPANTES	PREDICTORES	OUTCOME	RoB	CRA
4C	+	+	+	+	+	+	+	-	+
PREDI-CO	+	+	+	-	+	+	+	-	+
ANDC	+	+	+	-	+	+	+	-	+
ZHANG	+	+	+	-	+	?	+	-	?
CSS	+	+	+	-	+	+	+	-	+
PRE	+	+	+	-	+	+	+	-	+
HNC-LL	?	+	+	-	+	+	+	-	+
NEWS2	+	+	-	-	+	+	+	-	+

RoB: Risk of Bias; +: bajo RoB / bajo CRA; -: Alto RoB / Alto CRA; ?: incierto.

CRA: Concern Regarding Applicability.

En la tabla 29 mostramos algunas características de los estudios analizados.

Tabla 29. Características destacadas de los estudios seleccionados.

ESTUDIO	ASPECTOS RELEVANTES DEL ESTUDIO								
	CÁLCULO DEL TAMAÑO MUESTRAL	PREDICTORES CANDIDATOS	FÓRMULA o SCORE Puntuación	TÉCNICA ESTADÍSTICA	VALIDACIÓN INTERNA	VALIDACIÓN EXTERNA	GRÁFICO DE CALIBRACIÓN	CALCULADORA WEB / NOMOGRAMA	MISSING DATA
4C	No	41	Score	LASSO, ML	No	Sí	GC	C web	IM
PREDI-CO	Sí	36	Fórmula y score	RL	Sí	Sí	GC	No	ACC
ANDC	No	24	Fórmula	RL (-)	Sí	Sí	GC	Nomog	RF
ZHANG	Sí	30*	Score	Cox	No	Sí	GC	Nomog	IM
CSS	No	30*	Score	RL (-)	No	Sí	GC	No	IM
PRE	No	25	Fórmula y score	RL	No	No	No	No	No
HNC-LL	No	29*	Fórmula	RL, Cox	Sí	Sí	H-L	No	No
NEWS2	-	-	Score	-	No	Sí	No	C web	ACC

ML: Machine Learning; RL: regresión logística; Cox: regression de Cox; GC: gráfico de calibración; H-L: Hosmer-Lemeshow; C web: calculadora web; Nomog: nomograma; IM: imputación múltiple; ACC: análisis de casos completos; RF: random forest.

(-): emplean LASSO para seleccionar predictores y reducir la dimensionalidad.

(*) Se contabilizan como predictores las variables dummies

De izquierda a derecha, dichos aspectos son:

- *Tamaño muestral*: si se ha hecho el cálculo del tamaño muestral mínimo requerido, previo a la derivación del modelo, tal y como recomienda Riley et al^(42,43,45).
- *Predictores candidatos*: número de predictores candidatos considerados para la derivación del modelo.
- *Fórmula completa de regresión o score obtenido mediante un sistema de puntuación*: si se da la fórmula completa para el cálculo del riesgo o si se ha calculado un score otorgando una puntuación a cada predictor.

- *Técnica estadística*: indica si se ha creado el modelo final mediante regresión logística, regresión de LASSO u otra técnica estadística.
- *Validación interna*: si mencionan haber realizado validación interna.
- *Validación externa*: si han llevado a cabo validación externa.
- *Gráfico de calibración*: si se proporciona el gráfico de calibración, no sólo el test de Hosmer-Lemeshow.
- *Calculadora web o nomograma*: si se proporciona una calculadora web o nomograma.
- *Missing data*: cómo se han manejados los datos faltantes.

Sólo en 2 de los 8 estudios (score PREDI-CO y score de Zhang) los autores mencionan haber aplicado los 4 pasos (B1, B2, B3 y B4) propuestos por Riley et al. para calcular el tamaño muestral mínimo requerido en la derivación de un modelo de predicción. En dichos estudios, se han considerado los predictores finales incluidos en el modelo, no el número de predictores candidatos, para realizar los cálculos. En el resto de los estudios, no se ha hecho un cálculo previo de cuál sería el tamaño muestral mínimo requerido, aunque el tamaño considerado para derivar el score 4C es muy grande.

Hemos seguido las indicaciones marcadas por Riley para obtener cuáles deberían haber sido los tamaños muestrales mínimos requeridos para cada uno de los estudios, y de ese modo, poder valorar si los tamaños muestrales empleados en las derivaciones de los modelos fueron o no adecuados.

En la tabla 30 se muestra, de izquierda a derecha, el tamaño muestral que fue empleado para derivar el modelo, la prevalencia en la cohorte de derivación, el número de eventos en la cohorte de derivación según el outcome de cada estudio, el valor C-Statistic (AUC), número de predictores candidatos (PC), número de predictores incluidos en el modelo final (PF), el número máximo del tamaño muestral mínimo requerido obtenido entre los 3 criterios calculados (i, ii y iii), tamaño muestral mínimo requerido para el paso B2 y EPP.

Tabla 30. Cálculo del tamaño muestral mínimo requerido para los modelos predictivos seleccionados.

Score	n (D)	Prev (D)	E (D)	AUC	PC	PF	n máx(i,ii,iii)	B2	EPP
4C Mort.	35463	32.2%	11419	0.79	41	8	1.460	1.762	13.84
PREDI-CO	644	29%	189	0.89	36	8	800	1.486	11.97
ANDC	176	11.9%	21	0.92	24	4	779	668	3.86
Zhang	516	18.6%	96	0.89	30	4	799	1.016	6.30
CSS	113	43.4%	49	0.92	30	5	571	1.521	22.00
PRE	336	10.1%	34	0.99	25	3	798	642	3.22
HNC-LL	231	17.3%	40	0.86	29	5	916	951	5.67

PC: predictores candidatos; PF: predictores finales; EPP: eventos por variable predictora; E: eventos.
 (*) Se ha marcado en negrita cuál debería haber sido el tamaño muestral mínimo requerido para cada estudio según la regla de Riley.

Todos los valores obtenidos en EPP son inferiores a 20, salvo el del CSS score. Como ya se comentó al comienzo de la tesis, es complicado decidir si el tamaño muestral es adecuado cuando el valor EPP está entre 10-20. Pero dicho valor sí parece indicar que los tamaños muestrales empleados en la derivación de los modelos ANDC, Zhang, PRE y HNC-LL eran insuficientes. En el caso del NEWS2, no hemos hecho los cálculos de cuál sería el tamaño mínimo requerido porque este modelo no fue derivado para COVID-19. Nosotros nos hemos basado en una validación externa para considerar este estudio y no en un estudio de derivación, por lo que no tiene sentido realizar dicho cálculo.

En los modelos analizados, la mayoría de los autores (en 6 de los 8 estudios) han elegido proporcionar un score de puntuación, aunque en 4 de los 8 estudios se proporciona la fórmula de regresión. Sólo dos estudios dan ambas opciones. Cabe destacar que el PREDI-CO score emplea ambos métodos, aunque no proporciona el valor de la constante para el modelo final y cuando calcula las puntuaciones de cada uno de los predictores para obtener el valor total del score lo hace de manera incorrecta. Por lo que, consideramos que no se deberían tener en cuenta los resultados obtenidos con este score.

La técnica estadística empleada con más frecuencia ha sido la regresión logística (PREDI-CO, ANDC, CSS, PRE y HNC-LL), seguida de la regresión de Cox (Zhang y HNC-LL), que son las dos técnicas que habitualmente se emplean en medicina. Sólo un modelo, el 4C score, se ha creado empleando regresión de LASSO y Machine Learning.

Un requisito indispensable dentro de la etapa de derivación es realizar una validación interna del modelo, pero sólo 3 estudios (PREDI-CO, ANDC y HNC-LL) mencionan haberlo hecho. Aunque todos menos PRE score llevan a cabo una validación externa.

Las medidas empleadas son fundamentales en el reporte de los resultados, las más frecuentes para evaluar las predicciones son la de discriminación y la de calibración. Todos los estudios informan el AUC como medida de discriminación, aunque no todos reportan la calibración (como sucede en: PRE score y del NEWS2). En el estudio del HNC-LL score únicamente se informa del resultado del test de Hosmer-Lemeshow como medida de calibración. Y el 4C score, PREDI-CO, ANDC, Zhang y CSS score, reportan el gráfico de calibración además del test de Hosmer-Lemeshow.

De los estudios que hemos analizado, en 2 de ellos (ANDC y score de Zhang) se reporta el nomograma y, en otros dos (4C score y NEWS2), una calculadora web, aunque la del NEWS2 era previa a la pandemia.

2.2.3. Evaluación del reporte de resultados (TRIPOD).

Hemos revisado si el reporte de los resultados de cada uno de los estudios se había realizado de acuerdo a la declaración TRIPOD que establece las recomendaciones basadas en la evidencia para informar de forma estándar los estudios de los modelos de predicción en ciencias biomédicas. Tan sólo el estudio de Knight (4C score) indicó haber seguido la recomendación TRIPOD para reportar los resultados. En el resto de los estudios, los autores no lo mencionaron expresamente. En ninguno de los estudios se reportaron todos los ítems indicados en la guía, siendo el ítem de análisis estadístico uno de los que más dudas generaron, tal y como hemos mencionamos anteriormente en la evaluación de riesgos de sesgos realizada con la herramienta PROBAST.

2.2.4. Selección del mejor modelo predictivo.

De todos los scores para pronóstico que se validaron con los datos de la primera ola, el CSS fue el que obtuvo mejor rendimiento, con un AUC=0.83. El ANDC score obtuvo un AUC=0.82. La diferencia entre las dos AUC no es estadísticamente significativa. De hecho, la estimación por intervalos es exactamente igual, aunque la estimación puntual difiere mínimamente.

El principio de parsimonia (Ockham's Razor) establece que debemos seleccionar el modelo que contenga el menor número de variables si el rendimiento es similar ya que es más viable para su uso en la práctica clínica. El ANDC score está formado por 4 variables mientras que el CSS está compuesto por 5 variables.

De todos los modelos validados, el ANDC score es el que tiene una prevalencia similar, en nuestra población, a la que indican los autores de este modelo. En el caso del score CSS, la prevalencia que proporcionan los autores es casi 4 veces mayor que la prevalencia obtenida en nuestra población.

Por todo ello, seleccionamos el score ANDC como el mejor score de todos los validados en nuestra población de estudio.

Los autores del estudio establecen dos puntos de corte (59 y 101 puntos respectivamente) para el score ANDC. De manera que se obtienen tres grupos de riesgo: bajo, moderado y alto. Con los datos de nuestra población, la probabilidad de morir en cada uno de los grupos de riesgo, viene detallada en la tabla 31. Se observa que dichas probabilidades se corresponden con los valores que cabría esperar en esos grupos de riesgo.

Tabla 31. Probabilidad de morir en función del grupo de riesgo asignado según el ANDC score.

Score ANDC	Éxitus (n)	Probabilidad de morir	IC 95%
Bajo riesgo	0 (52)	0%	(0 - 6.85)
Riesgo moderado	8 (140)	5.71%	(2.48 - 11)
Alto riesgo	21 (55)	38.2%	(25.4 - 52.3)

2.2.5. Segunda validación externa del Score ANDC.

Tras la obtención de los resultados de la primera validación en pacientes de nuestra OSI, se llevó a cabo una segunda validación externa del score ANDC con los pacientes ingresados durante la segunda ola de la pandemia (1/8/2020-15/11/2020). Se obtuvo un **AUC=0.84 (IC95%: 0.81-0.87)** para dicha población (n=721), con una prevalencia de mortalidad del 8.5%. Las características nosológicas evaluadas del ANDC proporcionaron resultados similares a los obtenidos con los pacientes de la primera ola y el rendimiento fue un poco mejor que el que se había obtenido con los datos de la 1ª ola.

2.2.6. Comparativa de los pacientes de la 1ª y de la 2ª ola del HUD.

Se analizó si los pacientes de la 1ª ola eran similares a los pacientes de la 2ª ola en cuanto a la edad, al porcentaje de hombres que estaban enfermos o al número de fallecimientos (Tabla 32).

Tabla 32. Edades medianas y rangos intercuartílicos para los pacientes de la 1ª y 2ª ola del HUD.

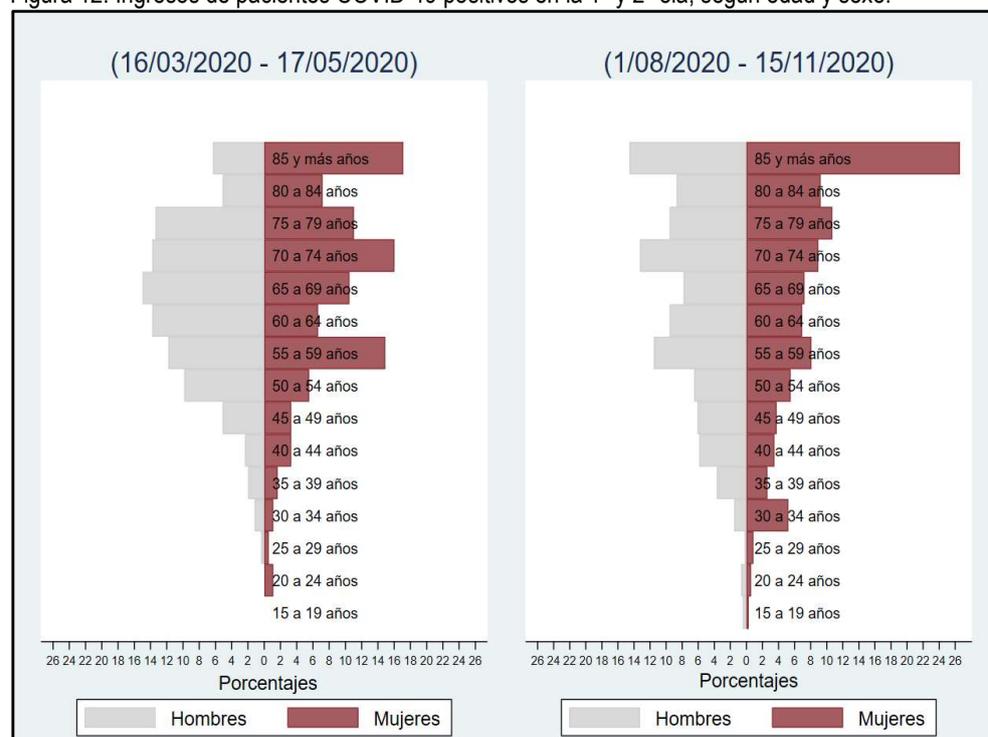
Período de estudio	Hombres	Mujeres	Total
1ª ola	Me = 66 (RI: 56, 74)	Me = 70 (RI: 57, 79)	Me = 67 (RI: 57, 76)
2ª Ola	Me = 67 (RI: 54.5, 78)	Me = 72.5 (RI: 55, 85)	Me = 70 (RI: 55, 81)

Me= mediana; RI= rango intercuartílico.

La población de la 2ª ola estaba un poco más envejecida que la de la 1ª ola, aunque las diferencias no fueron estadísticamente significativas.

Representamos gráficamente los ingresos que se produjeron en cada una de las olas, desglosando por sexos (Figura 12).

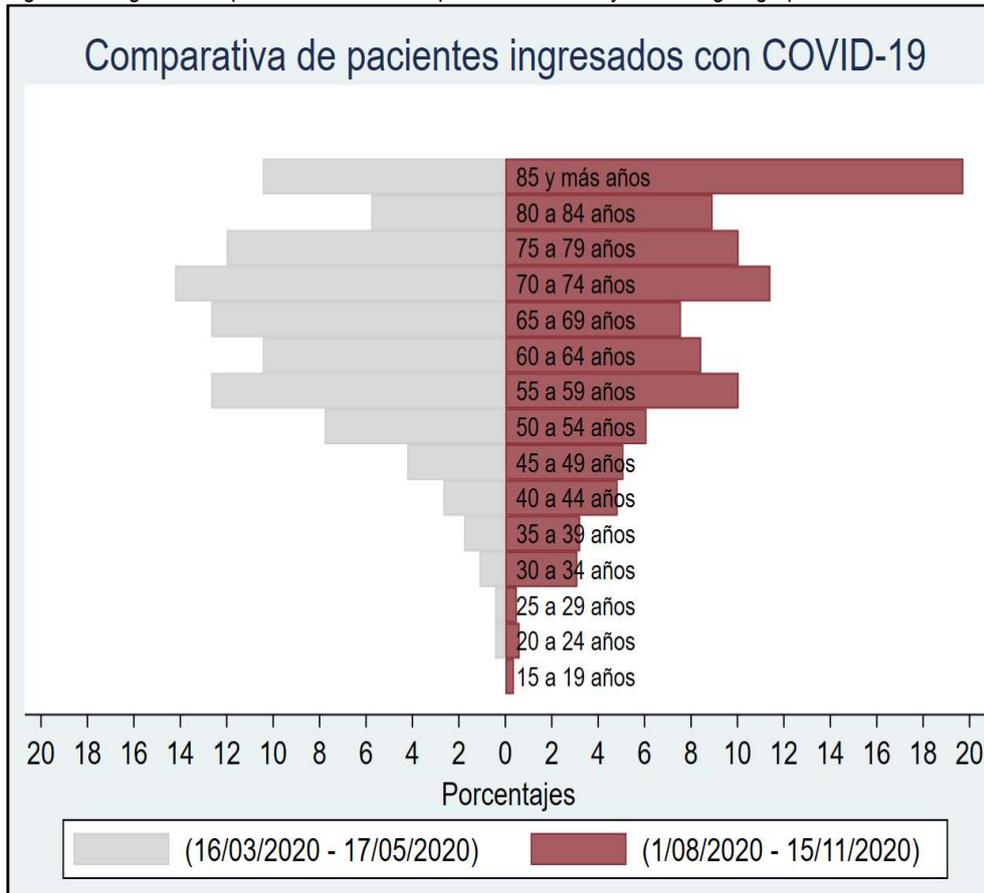
Figura 12. Ingresos de pacientes COVID-19 positivos en la 1ª y 2ª ola, según edad y sexo.



El 42.4% de los pacientes que ingresan con alguna RT-PCR positiva en el primer período, tienen más de 70 años frente al 50.1% del mismo tipo de pacientes en el segundo período. El 16.2% y el 28.7%, 1ª y 2ª ola respectivamente, corresponden a los pacientes que ingresaron en esos períodos y que tenían más de 80 años.

Representamos gráficamente (Figura 13) los ingresos de cada uno de los períodos sin desglosar por sexos, ya que se comprobó que no había diferencias estadísticamente significativas entre hombres y mujeres.

Figura 13. Ingresos de pacientes COVID-19 positivos en la 1ª y 2ª ola según grupos de edad.



Elaboramos la tabla 33 con los datos de fallecidos de cada período analizado.

Tabla 33. Fallecimientos de pacientes ingresados en la 1ª y la 2ª ola, desglosados por sexo.

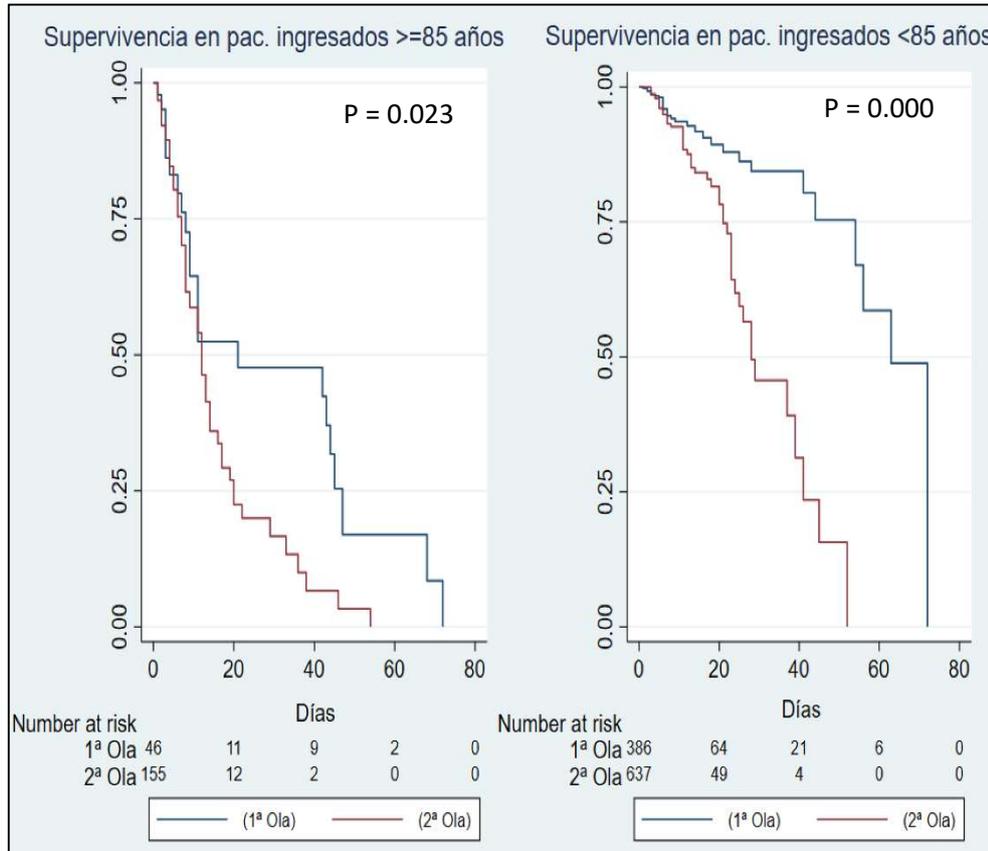
	Éxito n (%)	Vivo n (%)	Total n
(16/03/2020 - 17/05/2020)			
Hombres	25 (9.8)	229 (90.2)	254
Mujeres	28 (15.5)	153 (84.5)	181
Desconocido	8 (53.3)	7 (46.7)	15
Total	61 (13.6)	389 (86.4)	450
(1/08/2020 - 15/11/2020)			
Hombres	72 (15.7)	388 (84.3)	460
Mujeres	52 (15.0)	294 (85.0)	346
Total	124 (15.4)	682 (84.6)	806

En la 1ª ola se producen un 13.6% de fallecimientos intrahospitalarios frente al 15.4% en la 2ª ola. Del total de hombres ingresados en la 1ª ola, el 9.84% fallecen frente al 15.47% de mujeres, pero esas diferencias no se observan cuando incrementamos el tamaño muestral con la 2ª ola, en la que el 15.65% de los hombres ingresados fallecen frente al 15.03% de las mujeres.

Se analizaron también las curvas de supervivencia de ambas olas, obteniendo un tiempo mediano de: 56 días para la primera ola, con un rango intercuartílico de (41, 72) y de 24 días para la segunda, con un rango intercuartílico de (13, 39). En la 2ª Ola, el tiempo desde el ingreso hasta que el paciente fallece es más breve que en la primera ola. El valor obtenido con el test log Rank ($p=0.000$) indica que sí existen diferencias de mortalidad intrahospitalaria entre ambas olas.

Según la pirámide de población anterior parece que hay más ingresos en los mayores de 85 años en la segunda ola. Analizamos la supervivencia considerando los grupos creados con ese corte de edad (Figura 14).

Figura 14. Curvas de supervivencia en pacientes ingresados en la 1ª y la 2ª ola, desglosado por edad.



En el grupo de < 85 años, el tiempo mediano para la primera ola es 63 con un rango intercuartílico (54, 72) y, en la 2ª Ola, el tiempo mediano es muy inferior con un valor de 28 y un rango intercuartílico (21, 41). Para el grupo de ≥ 85 años, en la primera ola el tiempo mediano fue 21 con un rango intercuartílico (8, 47) y, en la 2ª ola, 12 con un rango intercuartílico (7, 20). Los tests log-rank realizados para cada uno de los grupos de edad ($p=0.000$ y $p=0.023$ respectivamente) indican que hay diferencias estadísticamente significativas entre los fallecidos en las dos olas y entre los grupos de edad de los pacientes.

Analizamos la letalidad en ambos períodos y por grupos de edad (Tabla 34).

Tabla 34. Letalidad de los pacientes ingresados en la 1ª y la 2ª ola, por grupos de edad.

Grupos de edad	INGRESOS CON COVID-19 CON FORMULARIO HUD					
	(16/03/2020 - 17/05/2020)			(1/08/2020 - 15/11/2020)		
	Éxito	n	Letalidad	Éxito	n	Letalidad
15 a 19 años	-	-	-	-	3	-
20 a 24 años	-	2	-	-	5	-
25 a 29 años	-	2	-	-	4	-
30 a 34 años	-	5	-	-	25	-
35 a 39 años	-	8	-	-	26	-
40 a 44 años	-	12	-	-	39	-
45 a 49 años	1	19	0.05	-	41	-
50 a 54 años	1	35	0.03	-	49	-
55 a 59 años	1	57	0.02	5	81	0.06
60 a 64 años	3	47	0.06	4	68	0.06
65 a 69 años	3	57	0.05	5	61	0.08
70 a 74 años	4	64	0.06	13	92	0.14
75 a 79 años	11	54	0.20	9	81	0.11
80 a 84 años	7	26	0.27	21	72	0.29
85 y más años	22	47	0.47	67	159	0.42
.	8	15	-	-	-	-

Se observa que los valores de letalidad son muy similares en ambas olas salvo en 2 grupos de edad: en el grupo de [70-74 años] en la 1ª ola fallecen un 6% de los ingresados contagiados frente a un 14% en la 2ª ola, y en el grupo de edad de [75-79 años], se pasa de un 20% de fallecidos del total de ingresados contagiados en la 1ª ola a un 11% en la 2ª ola.

Así mismo, se observa que en la 1ª ola la edad de corte a partir de la cual fallecen la mayoría de los ingresados es de 75 o más años, pero en la 2ª ola, dicho punto de corte disminuye, bajando hasta los 70 o más años.

Nota: hablamos de letalidad y no de mortalidad porque en el denominador se consideran los infectados y no la población total de cada grupo de edad.

2.2.7. Descripción y análisis del rendimiento del algoritmo de triaje modificado de urgencias tras la implementación del score ANDC en dicho algoritmo.

Una vez que se demostró que el score ANDC proporcionaba un buen rendimiento predictivo en nuestra población, se expusieron los resultados obtenidos a las diferentes unidades (urgencias, infecciosas, respiratorio, medicina interna, anestesia) implicadas en el manejo y cuidado de los pacientes COVID-19.

En ese tiempo, mediados de octubre de 2020, se observaba un incremento en el número de ingresos del HUD. Las urgencias comenzaban a sobrecargarse y para realizar el triaje de los pacientes, tal y como establecía el Ministerio de Sanidad, Servicios Sociales e Igualdad de España, se empleaba un algoritmo que daba mucho peso al patrón radiológico (ver Anexos), por lo que la mayoría de los pacientes eran ingresados si tras una radiografía de tórax se observaba neumonía multilobar, al margen de otras variables que, a juicio de los clínicos, eran también relevantes en dicha valoración.

La forma de trabajar en las diferentes unidades que están implicadas en el cuidado y manejo de pacientes COVID-19, ha ido cambiando a lo largo de los meses de pandemia, y tras conocer los resultados obtenidos con el ANDC score, se decidió modificar el algoritmo de triaje de urgencias e incluir este score como una característica más a tener en cuenta para decidir qué pacientes podían o no ser dados de alta tras acudir a urgencias con síntomas compatibles con la enfermedad. En diciembre de 2020, se realizó esta última modificación del algoritmo y el resultado final viene descrito en la Figura 15.

Se decidió también incluir el ANDC score dentro de los sistemas de información de Osakidetza, en la aplicación GestLab, que se emplea para emitir los informes de laboratorio del HUD (Figura 16), De esta manera se proporcionaba la probabilidad de riesgo de fallecer, junto con la clasificación correspondiente del riesgo (alto, moderado o bajo), de manera automática y como un parámetro más de dicho informe.

Figura 15. Algoritmo para el triaje de los pacientes en el servicio de urgencias del HUD (Dic 2020).

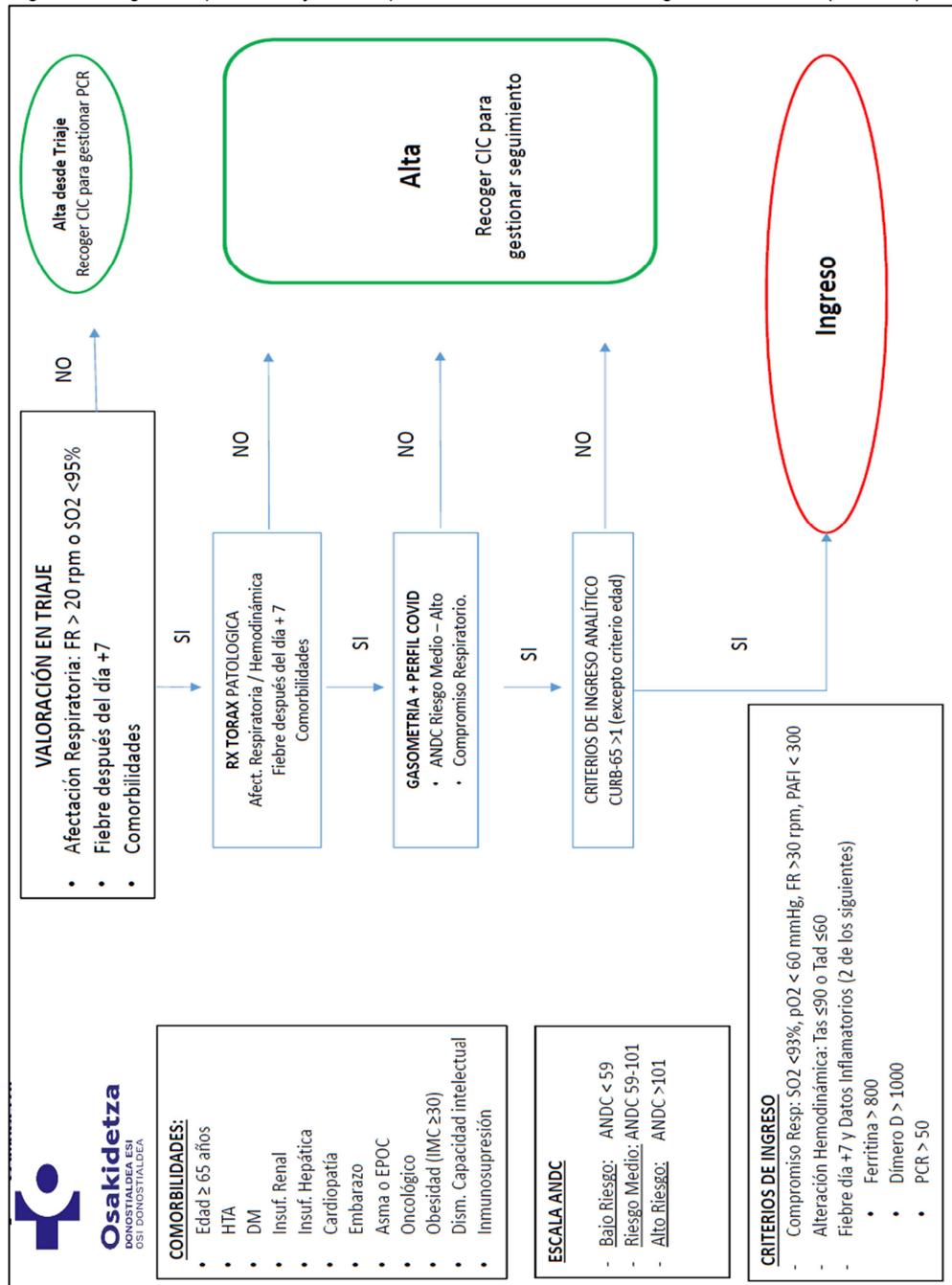


Figura 16. Informe de resultados de pruebas de laboratorio, con el cálculo del score ANDC, de un paciente atendido en el servicio de urgencias del HUD.

 Osakidetza LABORATORIOS HOSPITAL BIDASOA HOSPITAL MENDIARÓ HOSPITAL ZUMARRAGA HOSPITAL UNIVERSITARIO DONOSTIA Begiristain dk. pasealekua, z.g. 20014 Donostia-Gipuzkoa-España	CIC [REDACTED] NSS [REDACTED]	Fecha Muestra [REDACTED] Nº Petición [REDACTED]	
	1º Apellido [REDACTED] 2º Apellido [REDACTED] NOMBRE [REDACTED]		
INFORME DE RESULTADOS DE PRUEBAS DE LABORATORIO	F.Nacimiento [REDACTED] Edad 74 años Sexo H CIP-AUT [REDACTED] DNI [REDACTED]		
Procedencia HOSPITAL UNIVERSITARIO DONOSTIA Cama BOX04	Servicio URGENCIAS GENERALES (URG) Solicitante [REDACTED]		
Diagnostico: Fiebre Observaciones:			
INCIDENCIAS			
Muestra no recibida JERINGA GASES (SANGRE TOTAL-HEPARINA LITIO)			
Determinación	Resultado	Unidades	V.Referencia
Covid Escala ANDC Grupo de bajo riesgo: ANDC < 59 Grupo de riesgo moderado: ANDC entre 59 y 101 Grupo de alto riesgo: ANDC >101	96,60		
BIOQUIMICA GENERAL			
BIOQUIMICA SUERO			
Srm-Glucosa	98	mg/dL	[70 - 110]
Srm-Urea	30	mg/dL	[10 - 71]
Srm-Creatinina	0,80	mg/dL	[0,70 - 1,20]
Srm-Aspartato aminotransferasa (AST/GOT)	* 156	U/L	[0 - 37]
Srm-Alanina aminotransferasa (ALT/GPT)	* 121	U/L	[0 - 41]
Srm-Bilirrubina	0,5	mg/dL	[0,0 - 1,1]
Srm-Calcio	9,3	mg/dL	[8,8 - 10,2]
Srm-Fosfato	* 1,6	mg/dL	[2,5 - 4,5]
Srm-Magnesio	1,73	mg/dL	[1,59 - 2,56]
Srm-Proteinas	7,5	g/dL	[6,6 - 8,7]
Srm-Albumina	4,3	g/dL	[3,7 - 5,1]
Srm-Sodio	138	mEq/L	[135 - 145]
Srm-Potasio	4,1	mEq/L	[3,7 - 5,4]
Srm-Cloruro	99	mEq/L	[93 - 110]
Srm-Proteina C reactiva (PCR)	* 109,60	mg/L	[0,00 - 5,00]
Srm-Procalcitonina (PCT)	0,35	ng/mL	[<0,5]
Probabilidad de sepsis: baja con valores <0.5 ng/mL, y alta con valores >2 ng/mL (excepto en neonatos)			
Srm-Creatinquinasa (CK)	* 195	U/L	[0 - 189]

Tras la implementación en urgencias del nuevo algoritmo de triaje, se recogieron datos de manera prospectiva de todos los pacientes que acudieron al servicio de urgencias del HUD durante los meses de enero a febrero de 2021.

El objetivo de este estudio era analizar si el score ANDC podía ayudar a los clínicos a la hora de decidir qué pacientes tenían que ser ingresados (hospitalización en planta, UCI o UCRI) y cuáles podían ser dados de alta al domicilio habitual.

La lógica nos indica que:

- Si el ANDC era alto: la mayoría de los pacientes deberían haber sido ingresados en la UCI y algunos pacientes deberían haber sido ingresados en planta (aunque en menor proporción que los ingresados en UCI).
- Si el ANDC era moderado: la mayoría de los pacientes deberían haber sido ingresados en planta y algunos pacientes deberían haber sido dados de alta al domicilio habitual (aunque en menor proporción que los ingresados en planta del HUD).
- Si el ANDC era bajo: la mayoría de los pacientes deberían haber sido dados de alta a su domicilio habitual y algunos pacientes deberían haber sido ingresados en planta (aunque en menor proporción que los que se fueron de alta al domicilio habitual).

Nota: hay pacientes que están ingresados en UCI, otros en UCRI y otros que pasan por ambas unidades en algún momento del ingreso en el HUD. Los pacientes que pasan por ambas unidades se han contabilizado únicamente en la unidad de UCI en las tablas 35, 36, 37, 38, 39 y 40.

En la tabla 35 vemos que, de los 572 pacientes que acuden a urgencias en el periodo de estudio, 244 (42.7%) son dados de alta al domicilio habitual, 327 (57.2%) son ingresados en el HUD y 1 fallece en urgencias.

La clasificación del riesgo según el score ANDC, para los 572 pacientes que acuden a urgencias, fue: 27.6% riesgo alto, 48.6% riesgo moderado y 23.8% riesgo bajo.

De los clasificados como *alto riesgo*, el 65.2% fueron ingresados y el resto fue dado de alta al domicilio habitual. De los 102 pacientes ingresados, 16 (10.1%) ingresaron en UCI y 5 (3.2%) en UCRI en algún momento de su ingreso.

De los clasificados como *riesgo moderado*, el 68.7% fueron ingresados y el resto fueron dados de alta a su domicilio habitual. De los 191 pacientes ingresados, 33 (11.9%) fueron a UCI y 4 (1.4%) a UCRI en algún momento del ingreso.

Y, por último, de los clasificados como *bajo riesgo*, el 75% fueron dados de alta a su domicilio habitual, el resto fue ingresado. De los 34 pacientes ingresados, 6 (4.4%) fueron a UCI, en algún momento del ingreso, y ninguno fue a UCRI.

Se encuentran diferencias estadísticamente significativas entre la edad promedio (46.5) de los pacientes con ANDC bajo y la edad promedio de los pacientes con ANDC moderado y alto (69.2 y 68.7 respectivamente).

El tiempo de evolución promedio de los pacientes que acuden a urgencias es un poco mayor, aproximadamente un día, en el grupo ANDC de bajo riesgo.

No se observan diferencias notables en las variables analizadas entre los diferentes grupos de riesgo ni entre los destinos al alta analizados en la tabla 35.

En la tabla 36 vemos que, de los 572 pacientes que acudieron a urgencias durante el período de estudio, fallecieron 63 (11%). 3 pacientes pertenecen al grupo de pacientes que fueron dados de alta al domicilio habitual (244 pacientes) tras su primera visita a urgencias, 1 falleció en urgencias y el resto (59 pacientes) son del grupo de pacientes que fueron ingresados tras acudir a urgencias.

De los 327 pacientes ingresados, 263 (80.4%) fueron ingresados en planta, 30 (9.2%) en UCI, 9 (2.8%) en UCRI y 25 (7.6%) tanto en UCRI como en UCI. De los 327 pacientes, 7 fueron dados de alta y posteriormente volvieron a ingresar. De esos, 3 fallecieron (ninguno estuvo ingresado ni en UCI ni en UCRI). Del resto pacientes (320), fallecieron 56: 8 de ellos habían estado ingresados en UCRI, 9 en UCI y los 39 restantes estuvieron ingresados en planta de hospitalización.

Tabla 35. Características de pacientes (variables continuas) según score ANDC y destino al alta tras el triaje en urgencias.

ANDC (\bar{x} , σ)	DESTINO AL ALTA EN URGENCIAS				Total
	PLANTA	UCI	UCRI	DOMICILIO	
RIESGO ALTO	(n=81)	(n=16)	(n=5)	(n=55)	(n=158) (*)
Edad (años)	83.3 (9.1)	68.7 (6.3)	77.9 (4.1)	45.5 (15.9)	68.7 (21.0)
Días Evol.	4.7 (4.4)	2.9 (3.3)	6.8 (3.8)	4.2 (4.6)	4.4 (4.4)
TAS	70.1 (10.5)	69.2 (11.0)	67.3 (5.9)	76.6 (11.7)	72.1 (11.2)
TAD	139.0 (23.3)	137.4 (21.7)	146.0 (18.6)	128.8 (19.4)	135.7 (22.2)
FC	82.6 (15.3)	91.9 (17.6)	86.5 (18.6)	76.5 (13.8)	81.7 (15.7)
FR	24.8 (5.5)	28.0 (7.8)	28 (3.3)	17.9 (3.8)	23.1 (6.4)
Sat. O ₂	93.2 (4.1)	90.5 (5.9)	92.5 (5.2)	97.4 (2.0)	94.3 (4.5)
PAFI	340.0 (106.1)	229.4 (90.1)	242.0 (107.7)	412.2 (74.0)	325.8 (112.5)
RIESGO MODERADO	(n=154)	(n=33)	(n=4)	(n=87)	(n=278)
Edad (años)	70.0 (10.9)	63.1 (10.1)	77.7 (4.6)	69.6 (9.8)	69.2 (10.6)
Días Evol.	4.8 (4.0)	3.5 (2.8)	2.3 (2.6)	5.1 (3.8)	4.7 (3.8)
TAS	72.4 (11.1)	75.0 (14.2)	73.0 (12.8)	74.9 (12.1)	73.5 (11.8)
TAD	134.8 (20.2)	131.9 (23.5)	149.0 (30.9)	141.5 (21.7)	136.7 (21.4)
FC	78.5 (15.1)	89.1 (19.0)	77.3 (13.5)	78.3 (13.5)	79.7 (15.4)
FR	20.3 (4.5)	23.1 (6.3)	21.8 (3.9)	19.1 (3.5)	20.2 (4.5)
Sat. O ₂	93.8 (4.1)	91.6 (6.3)	93.5 (3.0)	95.4 (1.9)	94.0 (4.0)
PAFI	352.9 (85.0)	311.8 (85.8)	391.5 (123.3)	382.0 (61.5)	356.5 (82.0)
RIESGO BAJO	(n=28)	(n=6)	(n=0)	(n=102)	(n=136)
Edad (años)	46.7 (9.3)	44.6 (9.5)	-	46.5 (8.4)	46.5 (8.6)
Días Evol.	5.4 (3.9)	3.5 (3.0)	-	5.5 (4.2)	5.4 (4.1)
TAS	80.0 (8.1)	74.5 (2.5)	-	77.0 (10.7)	77.5 (10.1)
TAD	128.5 (13.1)	124.5 (8.4)	-	126.1 (14.5)	126.5 (14.0)
FC	86.3 (14.9)	86.7 (9.4)	-	80.0 (14.3)	81.6 (14.5)
FR	20.3 (5.7)	22.4 (2.2)	-	19.2 (4.0)	19.6 (4.4)
Sat. O ₂	92.0 (16.8)	92.8 (4.4)	-	96.4 (2.1)	95.4 (7.9)
PAFI	355.5 (50.8)	348.6 (37.4)	-	386.9 (98.9)	376.7 (87.3)

(*) Un paciente fallece en urgencias.

UCI: unidad de cuidados intensivos; UCRI: unidad de cuidados respiratorios intermedios; Días Evol: días de evolución de la enfermedad; TAS: Tensión arterial sistólica; TAD: Tensión arterial diastólica; FC: frecuencia cardíaca; FR: frecuencia respiratoria; Sat. O₂: Saturación de oxígeno; PAFI: (PaO₂/FiO₂).

Tabla 36. Fallecidos según grupos de riesgo del score ANDC y destino al alta tras el triaje en urgencias.

ANDC n (%)	DESTINO AL ALTA EN URGENCIAS				Total
	PLANTA	UCI	UCRI	DOMICILIO	
RIESGO ALTO	(n=81)	(n=16)	(n=5)	(n=55)	(n=158) (*)
Éxito	27 (33.3)	5 (31.3)	5 (100)	1 (1.8)	38 (24.1)
Vivo	54 (66.7)	11 (68.7)	-	54 (98.2)	119 (75.3)
RIESGO MODERADO	(n=154)	(n=33)	(n=4)	(n=87)	(n=278)
Éxito	15 (9.7)	4 (12.2)	3 (75.0)	2 (2.3)	24 (8.6)
Vivo	139 (90.3)	29 (81.8)	1 (25.0)	85 (97.7)	254 (91.4)
RIESGO BAJO	(n=28)	(n=6)	(n=0)	(n=102)	(n=136)
Éxito	-	-	-	-	-
Vivo	28 (100)	6 (100)	-	102 (100)	136 (100)

(*) Un paciente fallece en urgencias.

Si nos fijamos en la Tabla 37, observamos que dentro del grupo *alto riesgo*, asignado mediante el score ANDC, hubo un mayor porcentaje de pacientes con fiebre entre los que ingresaron comparado con aquellos que se fueron de alta al domicilio. Lo mismo ocurre con el oxígeno, necesitaron más los que estaban ingresados, y como es lógico, el porcentaje mayor fue en el grupo de pacientes ingresados en UCI y/o UCRI.

El criterio analítico, los factores de riesgo y la afectación multilobar fueron una característica de los que ingresan, así mismo, estos pacientes tuvieron un CURB-65 mayor de 2.

5 de los pacientes que fueron dados de alta al domicilio habitual con *alto riesgo* según el score ANDC, retornaron a urgencias.

Tabla 37. Características de pacientes (variables categóricas) según score ANDC (riesgo alto) y destino al alta tras el triaje en urgencias.

ANDC n (%)	DESTINO AL ALTA EN URGENCIAS				Total
	PLANTA	UCI	UCRI	DOMICILIO	
RIESGO ALTO	(n=81)	(n=16)	(n=5)	(n=55)	(n=158) (*)
T^a					
≥37.5°	20 (24.7)	4 (25.0)	1 (20.0)	8 (14.6)	34 (21.5)
<37.5°	61 (75.3)	12 (75.0)	4 (80.0)	47 (85.5)	124 (78.5)
Factores de riesgo					
0	2 (2.5)	5 (31.3)	-	36 (65.5)	43 (27.2)
1	14 (17.3)	5 (31.3)	-	8 (14.5)	27 (17.1)
2	61 (75.3)	6 (37.4)	5 (100)	7 (12.7)	79 (50.0)
RX					
Unilobar	10 (12.3)	-	-	6 (10.9)	16 (10.1)
Multilobar	46 (56.8)	15 (93.8)	5 (100)	2 (3.6)	68 (43.0)
Normal	20 (24.7)	1 (6.2)	-	39 (70.9)	60 (38.0)
CURB-65					
0	-	1 (6.2)	1 (20.0)	30 (54.5)	32 (20.3)
1	11 (13.6)	5 (31.3)	-	6 (10.9)	22 (13.9)
≥2	55 (67.9)	9 (56.3)	4 (80.0)	2 (3.6)	70 (44.3)
Necesitan oxígeno					
Sí	25 (30.9)	8 (50.0)	2 (40.0)	1 (1.8)	37 (23.4)
No	54 (66.7)	8 (50.0)	3 (60.0)	43 (78.2)	105 (66.5)
Criterio Analítico					
Sí	47 (58.0)	12 (75.0)	4 (80.0)	1 (1.8)	64 (40.5)
No	30 (37.0)	4 (25.0)	1 (20.0)	42 (76.4)	77 (48.7)
Retornan al hospital					
Sí	1 (1.2)	0 (0.0)	0 (0.0)	5 (9.1)	6 (3.8)
No	80 (98.8)	16 (100)	5 (100)	50 (90.9)	152 (96.2)

(*) Un paciente fallece en urgencias

T^a: temperatura; RX: radiografía de tórax; UCI: unidad de cuidados intensivos; UCRI: unidad de cuidados intermedios.

En la tabla 38 se muestra que en el grupo de *riesgo moderado* hubo un mayor porcentaje de pacientes con fiebre entre los que ingresaron comparado con aquellos que se fueron de alta al domicilio habitual.

Un 33.3% de los pacientes de UCI necesitaron oxígeno. 1 paciente de la UCRI, 18 de los hospitalizados en planta y 2 de los pacientes que fueron dados de alta al domicilio.

La mayoría de los pacientes ingresados tuvieron afectación multilobar y factores de riesgo. Y de los 87 pacientes dados de alta al domicilio, 20 retornaron.

Tabla 38. Características de pacientes (variables categóricas) según score ANDC (riesgo moderado) y destino al alta tras el triaje en urgencias.

ANDC n (%)	DESTINO AL ALTA EN URGENCIAS				Total
	PLANTA	UCI	UCRI	DOMICILIO	
RIESGO MODERADO	(n=154)	(n=33)	(n=4)	(n=87)	(n=278)
T ^a					
≥37.5°	32 (20.8)	9 (27.3)	1 (25.0)	11 (12.6)	53 (19.1)
<37.5°	122 (79.2)	24 (72.7)	3 (75.0)	76 (87.4)	225 (80.9)
Factores de riesgo					
0	27 (17.5)	7 (21.2)	-	19 (21.8)	53 (19.1)
1	39 (25.3)	10 (30.3)	-	32 (36.8)	81 (29.1)
2	88 (57.4)	16 (48.5)	4 (100)	36 (41.4)	144 (51.8)
RX					
Unilobar	33 (21.4)	4 (12.1)	1 (25.0)	14 (16.1)	52 (18.7)
Multilobar	92 (59.7)	27 (81.8)	1 (25.0)	27 (31.0)	147 (52.9)
Normal	27 (17.5)	2 (6.1)	2 (50.0)	46 (52.9)	77 (27.7)
CURB-65					
0	19 (12.3)	5 (15.2)	-	18 (20.7)	42 (15.1)
1	50 (32.5)	10 (30.3)	3 (75.0)	28 (31.2)	91 (32.7)
≥2	59 (38.3)	13 (39.4)	1 (25.0)	26 (29.9)	99 (35.6)
Necesitan oxígeno					
Sí	18 (11.7)	11 (33.3)	1 (25.0)	2 (2.3)	32 (11.5)
No	126 (81.8)	19 (57.6)	2 (50.0)	82 (94.25)	229 (82.4)
Criterio Analítico					
Sí	74 (48.1)	21 (63.6)	-	4 (4.6)	99 (35.6)
No	80 (51.9)	12 (36.4)	4 (100)	83 (95.4)	179 (64.4)
Retornan al hospital					
Sí	5 (3.3)	0 (0.0)	0 (0.0)	20 (23.0)	25 (9.0)
No	149 (96.7)	33 (100.0)	4 (100.0)	67 (77.0)	253 (91.0)

T^a: temperatura; RX: radiografía de tórax; UCI: unidad de cuidados intensivos; UCRI: unidad de cuidados intermedios.

Por último, en la Tabla 39 vemos que, de los 136 pacientes en el grupo de *bajo riesgo*, 102 pacientes fueron dados de alta al domicilio y 34 de ellos (25%) fueron ingresados. Un tercio, aproximadamente, tuvieron fiebre. Y 3 recibieron oxígeno (1 en la UCI).

La mayoría de los pacientes ingresados presentaban afectación multilobar, pero no tenían factores de riesgo ni criterio analítico.

De los 102 pacientes que se fueron de alta al domicilio, retornaron 16 (15.7%).

Tabla 39. Características de pacientes según (variables categóricas) score ANDC (riesgo bajo) y destino al alta tras el triaje en urgencias.

ANDC n (%)	DESTINO AL ALTA EN URGENCIAS				Total
	PLANTA (n=28)	UCI (n=6)	UCRI (n=0)	DOMICILIO (n=102)	
RIESGO BAJO					(n=136)
T^a					
≥37.5°	6 (21.4)	3 (50.0)	-	13 (12.7)	22 (16.2)
<37.5°	22 (78.6)	3 (50.0)	-	89 (87.3)	114 (83.8)
Factores de riesgo					
0	16 (57.1)	4 (66.6)	-	65 (63.7)	85 (62.5)
1	8 (28.6)	1 (16.7)	-	31 (30.4)	40 (29.4)
2	4 (14.3)	1 (16.7)	-	6 (5.9)	11 (8.1)
RX					
Unilobar	5 (17.9)	2 (33.3)	-	29 (28.4)	36 (26.5)
Multilobar	20 (71.4)	4 (66.7)	-	31 (30.4)	55 (40.4)
Normal	2 (7.1)	-	-	41 (40.2)	43 (31.6)
CURB-65					
0	13 (46.4)	6 (100)	-	74 (72.5)	93 (68.4)
1	10 (35.7)	-	-	21 (20.6)	31 (22.8)
≥2	-	-	-	-	-
Necesitan oxígeno					
Sí	2 (7.1)	1 (16.7)	-	2 (2.0)	5 (3.7)
No	24 (85.7)	4 (66.7)	-	93 (91.2)	121 (89.0)
Criterio Analítico					
Sí	9 (32.1)	1 (16.7)	-	2 (2.0)	12 (8.8)
No	19 (67.9)	5 (83.3)	-	100 (98.0)	124 (91.2)
Retornan al hospital					
Sí	1 (3.6)	0 (0.0)	-	16 (15.7)	17 (12.5)
No	27 (96.4)	6 (100.0)	-	86 (84.3)	119 (87.5)

T^a: temperatura; RX: radiografía de tórax; UCI: unidad de cuidados intensivos; UCRI: unidad de cuidados intermedios.

Analizamos los pacientes que retornaron al hospital. En total hay 48 pacientes que retornaron, de los cuales, 7 estuvieron ingresados la primera vez que acudieron a las urgencias del HUD y tras unos días de ingreso fueron dados de alta, aunque posteriormente volvieron al hospital (éstos no se han considerado en la siguiente tabla). El resto de los pacientes, 41, fueron dados de alta al domicilio la primera vez que acudieron a urgencias y posteriormente regresaron (son los que mostramos en la Tabla 40).

Sabíamos que antes de incluir el ANDC en el algoritmo de triaje, el número de pacientes que ingresaba era muy elevado debido a que el patrón radiológico tenía mucho peso en el algoritmo. Después de implementar el score, se pudieron dar 244 altas al domicilio, de las cuáles 41 (16,8%) retornaron, y de esos, 5 pacientes tuvieron que ser ingresados en la UCI y 1 en la UCRI. De los pacientes ingresados en UCI, 3 fallecieron (1 con riesgo alto en el score ANDC y 2 con riesgo moderado). Por lo tanto, el porcentaje de pacientes que fallecen de los que se fueron a casa dados de alta tras acudir a urgencias es del 1.2%.

En la primera visita a urgencias, el tiempo desde el inicio de síntomas hasta que acuden a urgencias, de esos 3 pacientes fallecidos, fue: 3, 2 y 5 días respectivamente y, en la segunda visita a urgencias (cuando retornan), el tiempo desde el inicio de síntomas hasta que acuden a urgencias fue: 6, 4 y 9 días respectivamente.

De los 41 pacientes que retornaron al HUD, 10 fueron dados de alta al domicilio habitual. De los 31 restantes, 5 ingresaron (en algún momento) en UCI, 1 en la UCRI y el resto (25 pacientes) en planta. El retorno se produjo, en promedio, a los 3.1 días (SD=2.4) tras haber acudido la primera vez a las urgencias del HUD.

Se comprobó que uno de los motivos por los que estos pacientes regresaron al hospital fue porque en el momento en el que fueron atendidos por primera vez en las urgencias del HUD, llevaban pocos días de evolución de la enfermedad y, por lo tanto, estaban en una fase temprana de la misma. El tiempo promedio de evolución, la primera vez que acuden a urgencias es de 3.9 (SD=2.9). El tiempo de evolución cuando retornan es, en promedio, de 6.7 (SD=3.7).

Hemos aplicado también el algoritmo inicial propuesto por el Ministerio (ver Anexos) para saber cuál habría sido el destino al alta para cada uno de los pacientes si se hubiese empleado dicho algoritmo en lugar de usar el algoritmo modificado (Figura 15).

Los resultados que obtuvimos indican que: 171 (29.9%) de los pacientes habrían sido dados de alta al domicilio y 400 (69.9%) habrían sido hospitalizados. Por lo tanto, con la inclusión del score ANDC en el algoritmo de triaje, se evitaron 73 ingresos. Por lo que se redujeron los ingresos en un 12.8%.

Tabla 40. Pacientes que retornan al HUD, según ANDC y destino al alta tras el triaje en urgencias cuando retornan (2ª visita).

ANDC RETORNO	DESTINO AL ALTA TRAS RETORNO				Total
	PLANTA	UCI	UCRI	DOMICILIO	
RIESGO ALTO	(n=5)	(n=0)	(n=1)	(n=0)	(n=6)
Días Evolución (\bar{X} , σ)	8.0 (1.0)	-	6 (-)	-	7.7 (1.2)
Días hasta retorno (\bar{X} , σ)	3.4 (1.8)	-	3 (-)	-	3.3 (1.6)
Insuficiencia n (%)					
Sí	1 (20.0)	-	-	-	1 (16.7)
No	4 (80.0)	-	1 (100)	-	5 (83.3)
Motivo de ingreso n (%)					
Alter. Respiratoria	1 (20.0)	-	-	-	1 (16.7)
Imagen radiológica	3 (60.0)	-	-	-	3 (50.0)
Otros	1 (20.0)	-	1 (100)	-	2 (33.3)
No se especifica	-	-	-	-	-
RIESGO MODERADO	(n=13)	(n=5)	(n=0)	(n=4)	(n=22)
Días Evolución (\bar{X} , σ)	5.9 (3.5)	5.6 (3.1)	-	10.5 (4.7)	6.7 (3.9)
Días hasta retorno (\bar{X} , σ)	3.8 (1.9)	3.8 (2.1)	-	6.5 (5.9)	4. (3.0)
Insuficiencia n (%)					
Sí	-	3 (60.0)	-	2 (50.0)	5 (22.7)
No	13 (100)	2 (40.0)	-	2 (50.0)	17 (77.3)
Motivo de ingreso n (%)					
Alter. Respiratoria	3 (23.1)	3 (60.0)	-	-	6 (27.3)
Imagen radiológica	3 (23.1)	1 (20.0)	-	3 (75.0)	7 (31.8)
Otros	7 (53.8)	1 (20.0)	-	1 (25.0)	9 (40.9)
No se especifica	-	-	-	-	-
RIESGO BAJO	(n=7)	(n=0)	(n=0)	(n=6)	(n=13)
Días Evolución (\bar{X} , σ)	5.3 (4.2)	-	-	7.2 (4.4)	6.2 (4.2)
Días hasta retorno (\bar{X} , σ)	2.4 (1.1)	-	-	4.8 (4.6)	3.5 (3.3)
Insuficiencia n (%)					
Sí	1 (14.3)	-	-	-	1 (7.7)
No	6 (85.7)	-	-	6 (100)	12 (92.3)
Motivo de ingreso n (%)					
Alter. Respiratoria	5 (71.4)	-	-	-	5 (38.5)
Imagen radiológica	-	-	-	1 (16.7)	1 (7.7)
Otros	2 (28.6)	-	-	5 (83.3)	7 (53.8)
No se especifica	-	-	-	-	-

Alter. Respiratoria: alteración respiratoria; (\bar{X} , σ): (media, desviación típica); UCI: unidad de cuidados intensivos; UCRI: unidad de cuidados intermedios.

3. CARACTERIZACIÓN DE LAS PRUEBAS DIAGNÓSTICAS DE LABORATORIO.

No disponemos de ninguna prueba diagnóstica que permita predecir en qué pacientes está presente la enfermedad y en cuáles no con una fiabilidad del 100%. Los estudios que se han realizado al respecto se han desarrollado en el laboratorio y han reportado rendimientos de las diferentes pruebas diagnósticas de laboratorio. pero en el momento en el que se desarrolló este estudio, no existían estudios que proporcionasen información del rendimiento de las pruebas de laboratorio en entornos clínicos.

Es por ello, que nos hemos centrado en analizar las tres pruebas diagnósticas de laboratorio principales. En primer lugar, las dos pruebas de detección de infección activa, que son: la prueba de la RT-PCR y la prueba rápida de anticuerpos Panbio™ y, en segundo lugar, la prueba de infección pasada, que es la prueba de determinación de anticuerpos (Serología). Nuestro objetivo era determinar la sensibilidad de cada una de estas pruebas cuando se han aplicado en nuestro medio (con los datos de la OSI) y poder determinar la fiabilidad del resultado obtenido.

3.1. PRUEBA DE LA RT-PCR (REVERSE TRANSCRIPTION POLYMERASE CHAIN REACTION)

En la 1ª ola de la pandemia, comprendida entre (16/03/2020-17/05/2020) se realizaron 66.970 pruebas RT-PCR en el HUD, de las cuales: 2.888 (4.3%) obtuvieron resultado positivo y 63.320 (94.5%) resultado negativo. Se obtuvieron 762 resultados dudosos o indeterminados (Tabla 41 y 42).

Tabla 41. Resultados de la prueba RT-PCR realizada en el HUD en la primera ola.

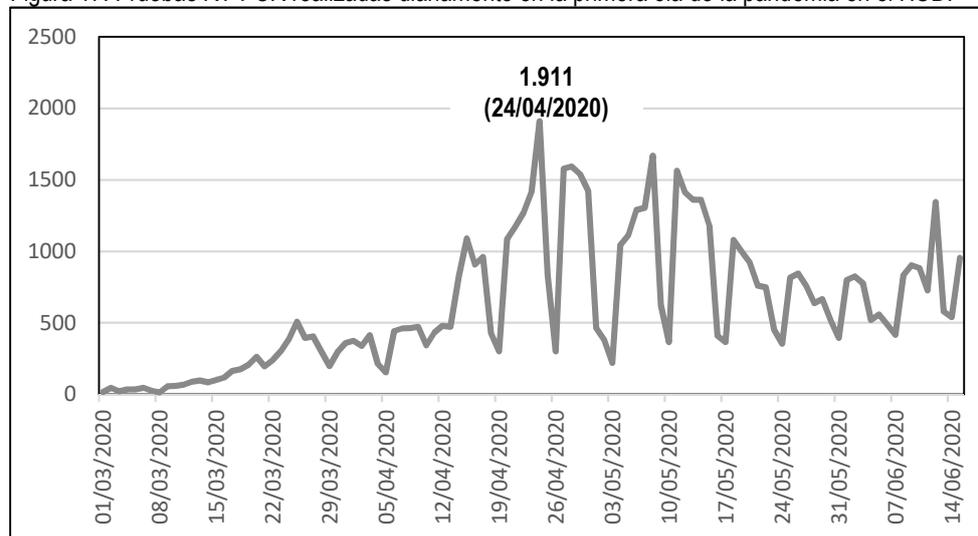
	Resultados de la RT-PCR				Total
	Sólo ⊕	Sólo ⊖	⊕ y ⊖	Indeterminadas	
Nº pruebas	851	60.928	4.429	762	66.970
Pacientes	790	50.410	1.539	735	53.474

Tabla 42. Resultados de la prueba RT-PCR realizada en el HUD en la primera ola, sin considerar los resultados dudosos o indeterminados.

	Resultados de la RT-PCR			
	Sólo ⊕	Sólo ⊖	⊕ y ⊖	Total
Nº pruebas	851	60.928	4.429	66.208
Pacientes	790	50.410	1.539	52.739

Representamos gráficamente (Figura 17) las RT-PCR realizadas diariamente en el periodo de estudio. En promedio, se realizaron aproximadamente 619 pruebas RT-PCR/día (SD= 416.4). Y la Me y el RI de la muestra fueron 471 y (299-911) respectivamente. El 24 de abril fue el día que más pruebas se realizaron, con un total de 1.911 pruebas.

Figura 17. Pruebas RT-PCR realizadas diariamente en la primera ola de la pandemia en el HUD.



3.1.1. Sensibilidad de la primera RT-PCR con los datos de pacientes de la 1ª ola del HUD.

Consideramos que la especificidad de la prueba RT-PCR es, por definición, del 100% (es decir, que una RT-PCR \ominus indica que el paciente es no COVID-19) y que un individuo presenta la enfermedad del SARS-CoV-2 si tiene, al menos, una RT-PCR positiva (según las guías de la OMS).

Basándonos en los datos proporcionados en el diagrama de flujo presentado en la Figura 9 (pág. 132) de los pacientes de la OSI, seleccionamos a los pacientes que tienen la 1ª RT-PCR \oplus , que son: los 162 pacientes que tuvieron todas las RT-PCR \oplus y los 228 (288 – 60) pacientes que tuvieron la 1ª prueba RT-PCR \oplus y, posteriormente, alguna RT-PCR \ominus . En total, 390 pacientes con la 1ª RT-PCR \oplus . Del mismo modo, seleccionamos a los pacientes con la primera RT-PCR negativa, 475 en total. Con esos datos, elaboramos la siguiente tabla de contingencia (Tabla 43).

Tabla 43. Pacientes ingresados con formulario COVID-19 y con una RT-PCR en las 48h tras el ingreso.

	Covid	No Covid	Total
1ª PCR ⊕	390	0	390
1ª PCR ⊖	49	426	475
Total	439	426	865

De las 865 pruebas RT-PCR realizadas en las 48h tras el ingreso, 475 (55%) fueron negativas. De ellas, el 10.3% (49/475) se positivizan en los siguientes 7 días.

La sensibilidad para la 1ª RT-PCR realizada en el HUD en las 48h tras el ingreso fue de 88.8% (390/439) y el IC95%: (85.9-91.8).

3.1.2. Sensibilidad de la segunda RT-PCR con los datos de pacientes de la 1ª ola del HUD.

Para calcular la sensibilidad de la 2ª RT-PCR, establecimos la consideración de que el paciente hubiese obtenido una RT-PCR negativa y se le realizara otra RT-PCR (con un máximo de 48h entre ambas pruebas). Con los datos: $Se=88.8\%$ y $Sp=100\%$ de la 1ª RT-PCR en el HUD y la prevalencia del 10.3%, obtenemos la siguiente tabla 2x2 (Tabla 44):

Tabla 44. Sensibilidad de la 2ª RT-PCR con los datos de la OSI-Donsotialdea (1ª Ola).

	Covid	No Covid	Total
2ª PCR ⊕	44	0	44
2ª PCR ⊖	5	426	431
Total	49	426	475

La sensibilidad de la 2ª RT-PCR es del 89.8% (IC95%: 77.8-96.6).

3.1.3. Probabilidad de que un paciente sospechoso de COVID-19 se positivice (tras obtener dos RT-PCR negativas consecutivas con un máximo de 48h entre ambas) en los siguientes 7 días.

Podemos abordar este análisis de diferente manera:

- Aplicando el Teorema de Bayes: podemos realizar el cálculo de 2 formas diferentes.

1. Empleando la tabla 2x2 (Tabla 40):

Con esos datos, el 1.2% (5/431) de los pacientes que han obtenido dos resultados negativos consecutivos de la RT-PCR, se positivarán.

2. Aplicamos la fórmula del Teorema de Bayes:

$$1 - VPN = 1 - \frac{(0.897 * 1)}{(0.897 * 1) + (0.103 * 0.112)} = 0.01269 \quad (1.3\%)$$

Por lo tanto, la probabilidad calculada de positivizarse tras dos RT-PCR \ominus consecutivas, con un máximo de 48h entre ellas, a la siguiente semana es de 1.3%.

Nota: los datos obtenidos por los dos métodos, no son exactamente iguales porque cuando aplicamos los datos de Se y Sp para hacer la tabla 2x2, cogemos valores enteros y no con decimales y eso afecta al cálculo.

- Utilizando los datos de RT-PCR observados en la 1ª ola y calculando la probabilidad de positivizarse.
 - Pacientes que tienen sólo RT-PCR \ominus :
 - Más de 1 RT-PCR \ominus (\leq 48h tras ingreso): 321 pacientes.
 - \leq 48h entre 1ª RT-PCR y 2ª RT-PCR \ominus : 236 pacientes.
 - Sólo 2 RT-PCR \ominus : 127 pacientes.
 - Más de 2 RT-PCR \ominus : 109 pacientes.
 - Pacientes que tienen RT-PCR \ominus y \oplus :
 - $\ominus\ominus\oplus$: 10 pacientes.
 - \leq 48h entre 1ª RT-PCR \ominus y 2ª RT-PCR \ominus : 8 pacientes.
 - RT-PCR \oplus \leq 7 días: 5 pacientes.

Con estos datos, calculamos qué porcentaje de los pacientes se van a positivar: $5/(236+5) = 0.021$ (2.1% ; IC95%: 0.64-4.64).

Podemos concluir que la probabilidad observada de positivizarse tras 2 RT-PCR \ominus consecutivas, con un máximo de 48h entre ellas, a la siguiente semana es de 2.1%.

3.2. PRUEBA RÁPIDA DE ANTÍGENOS PANBIO™ (Ag RAPID TEST DEVICE)

En la búsqueda bibliográfica realizada hasta el 1 de diciembre de 2020 se recopilaron 14 estudios en los que se analizaba la sensibilidad y la especificidad de la prueba rápida de antígenos Panbio™.

La prueba rápida de antígenos Panbio™ parece tener una especificidad muy alta, próxima al 100%. De hecho, varios de los estudios encuentran especificidades del 100%. Por ello, no hemos combinado los resultados de especificidad, sólo nos hemos ceñido a los datos de sensibilidad.

Los estudios de Linares⁽¹⁷⁸⁾, Albert⁽¹⁷⁹⁾, Gremmels⁽¹⁸⁰⁾ y Fenollar⁽¹⁸¹⁾ se llevaron a cabo de forma prospectiva realizando la prueba rápida de antígenos a pacientes que acuden a los centros. El estudio de Mitja et al⁽¹⁸²⁾, se realizó con muestras congeladas, y no queda bien explicado cómo fue la recogida de las muestras. Los artículos de Berger⁽¹⁸³⁾, Masiá⁽¹⁸⁴⁾, Drevinek⁽¹⁸⁵⁾ y Bulilete⁽¹⁸⁶⁾ reclutan también prospectivamente, haciendo la prueba point-of-care. El artículo de Merino-Amador⁽¹⁸⁷⁾ recluta pacientes ingresados. El estudio de Abdulrahman⁽¹⁸⁸⁾ recluta a pacientes levemente sintomáticos, Billaud⁽¹⁸⁹⁾ analiza de manera separada sintomáticos y asintomáticos (aunque no queda muy claro a quién considera asintomáticos), Schwob⁽¹⁹⁰⁾ sólo seleccionó a pacientes sintomáticos y Torres⁽¹⁹¹⁾ sólo a pacientes asintomáticos.

3.2.1. Resumen de los estudios incluidos en la RS.

En la tabla 45, se resume la información de los 14 estudios identificados en la RS realizada. De izquierda a derecha vemos, en primer lugar, el nombre del autor del artículo, fecha de publicación, número de pacientes considerados, lugar de realización y características de los pacientes en el estudio.

Tabla 45. Estudios seleccionados para MA de la Se del test rápido de antígenos Panbio™.

Artículo	Fecha	n	Lugar	Características de los pacientes
Linares	16-oct (15-20/09)	141 67	Madrid (UH y AP)	Sx: <7d. Dif en Sx UH y AP (fiebre, disnea) aSx: con contacto próximo
Albert	20-oct (2/9-7/10)	412	Valencia (AP)	Sx: 85 niños <16
Gremmels	20-oct (22/9-9/10)	947	Holanda (AP)	Sx: 84% <7d aSx: 37, sin datos de Ag
Fenollar	2-nov (21/9-1/10)	182 159	Francia	Sx: todos PCR ⊕ aSx: con contacto próximo
Alemaný-Mitja	3-nov (fechas?)	446 473 487	Cataluña	Sx: "Case" Frozen swabs ?Sx: "Contact" aSx: "Screening"
Berger	23-nov (9 – 23/10)	535	Suiza (1 centro)	Sx
Masiá	17-nov (15/9-29/10)	617 296	Alicante 3 CAP, 1 DE	Sx: 491 < 7d aSx: con contacto próximo
Bulilete	16-nov (2-25/10)	680 689	Mallorca (4 CAP)	Sx: < 7d aSx: 87%, con contacto próximo
Merino- Amador	20-nov (/9 y /10)	830 128	ISCIII (10 hosp.)	Sx: pacientes ingresados aSx: con contacto próximo
Billaud	30-sep	147 315	Francia	Sx aSx ¿?
Drevinek	24-nov	290 301	Praga 3 clinics	Sx aSx con contacto próximo
Schwob	24-nov	271	Suiza	Sx
Abdulrahman	13-nov	4.183	Bahrain	Sx (levemente sintomáticos)
Torres	-	634	Valencia (INCLIVA)	aSx con contacto estrecho

Sx: Sintomáticos; aSx: asintomáticos; HU: Hospital universitario; AP: Atención primaria; CAP: Centro de Atención Primaria; DE: Departamento de emergencias

Detallamos a continuación (Tabla 46), los valores de prevalencia, sensibilidad y especificidad extraídos de cada uno de los estudios seleccionados para el análisis de la prueba diagnóstica de laboratorio del test rápido de antígenos Panbio™.

Tabla 46. Se, Sp y prevalencia de los estudios seleccionados para MA del test rápido de antígenos Panbio™.

Artículo	Pacs.	n	Prev. (%)	Se (%)	Sp (%)
Linares	Sx	141	26.2	86.5	-
	aSx	67	14.9	50	100
Albert	Sx	412	13.1	79.6	-
Gremmels	Sx	947	10.7	74.3	-
Fenollar	Sx	182	"100"	79.1	-
	aSx	159	13.8	45.5	94.5
Mitja	Sx	446	93.9	92.6	-
	?Sx	473	87.7	94.2	-
	aSx	487	24.0	79.5	98.6
Berger	Sx	535	23.2	85.5	-
Masiá	Sx	617	25.3	67.3	100
	aSx	296	13.2	33.3	100
Bulilete	Sx	680	12.8	80.5	99.7
	aSx	689	7.74	56.6	100
Merino-Amador	Sx	830	37.5	90.5	98.8
	aSx	128	24.2	100	100
Billaud	Sx	147	50.3	54	98.6
	aSx	315	7.94	52	100
Drevinek	Sx	290	57.9	67	-
	aSx	301	18.3	44	-
Schwob	Sx	271	45.0	86	-
Abdulrahman	Sx	4183	17.5	82.1	99.1
Torres	aSx	634	12.5	48.1	100

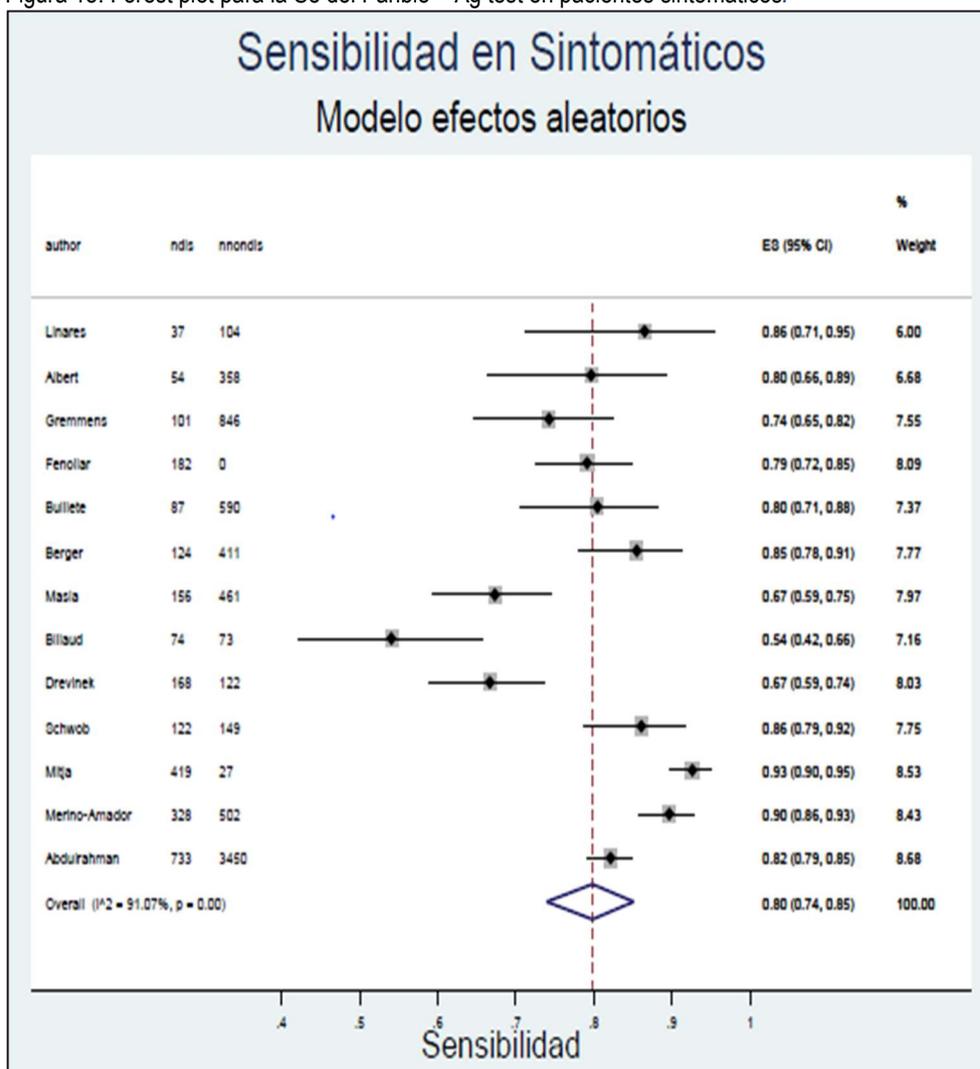
Pacs: pacientes; Sx: sintomáticos; aSx: asintomáticos; Se: sensibilidad; Sp: especificidad.

Para llevar a cabo la combinación de estudios mediante meta-análisis, por una parte, se analizó la sensibilidad de la prueba Panbio™ Ag test en pacientes sintomáticos de menos de 7 días de evolución. Para ello, se consideraron 13 estudios con un total de 9.681 pacientes. Por otra parte, analizamos la sensibilidad en asintomáticos, considerando 9 estudios con un total de 3.076 pacientes.

3.2.2. Combinación de estudios en pacientes sintomáticos.

Tras aplicar el MA para los 13 estudios con un modelo de efectos aleatorios en pacientes sintomáticos (Figura 18), se obtuvo una sensibilidad relativa del 0.8 (IC95%: 0.74-0.85) y una I^2 de 91.1%, lo que indica una gran heterogeneidad, fruto sin duda de las diferentes formas de seleccionar las muestras, del lugar de la toma de las muestras (AP u Hospitalaria) y/o de las diferentes prevalencias, etc.

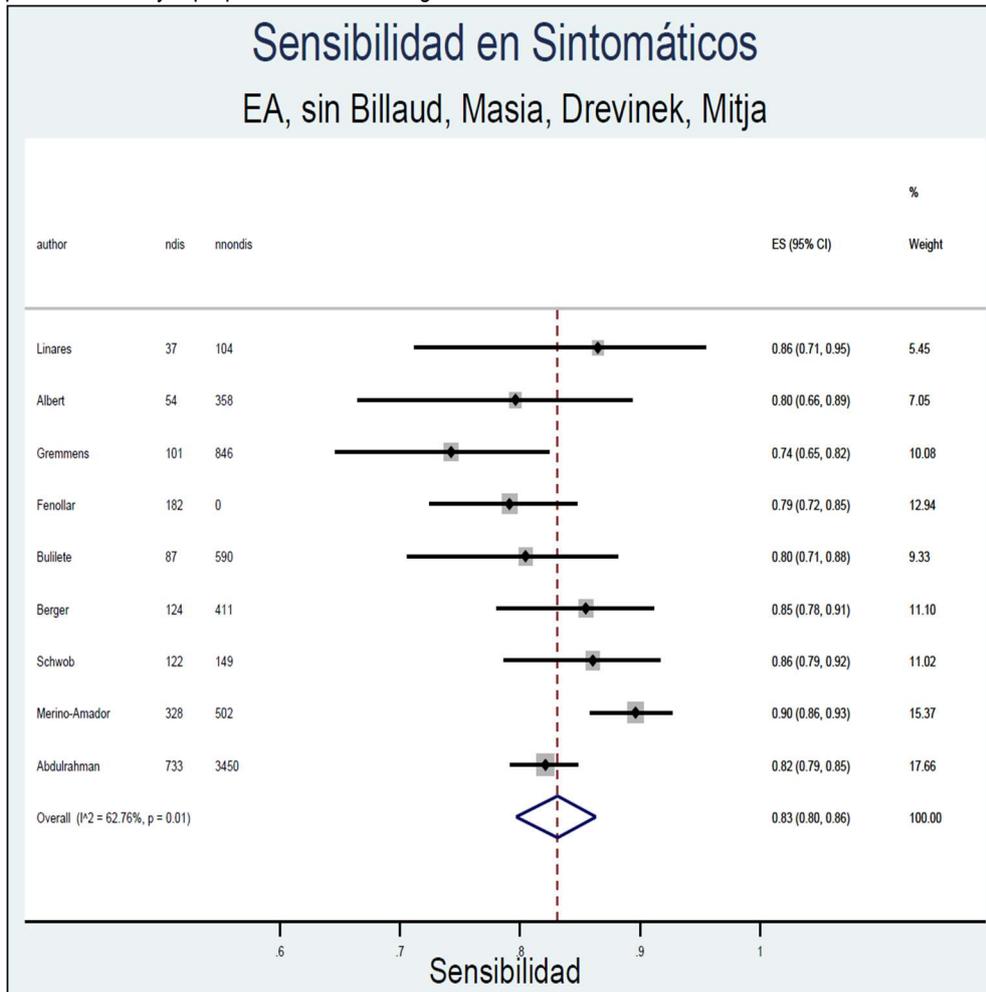
Figura 18. Forest plot para la Se del Panbio™ Ag test en pacientes sintomáticos.



ndis: pacientes enfermos; nnondis: pacientes no enfermos y ES: sensibilidad.

Se eliminaron aquellos estudios (Billaud, Masiá, Drevinek y Mitja) que parecían ser responsables de la mayor proporción de la heterogeneidad. De este modo, se disminuyó la I^2 hasta 62.8% para este MA, aunque el valor seguía siendo elevado. En la Figura 19 se proporciona la estimación de la sensibilidad del Panbio™ Ag test, que fue 0.83 (IC95%: 0.80-0.86), por lo tanto, similar a la sensibilidad obtenida en el forest plot anterior.

Figura 19. Forest plot para la Se del Panbio™ Ag test en pacientes sintomáticos sin 4 estudios que producen la mayor proporción de la heterogeneidad.



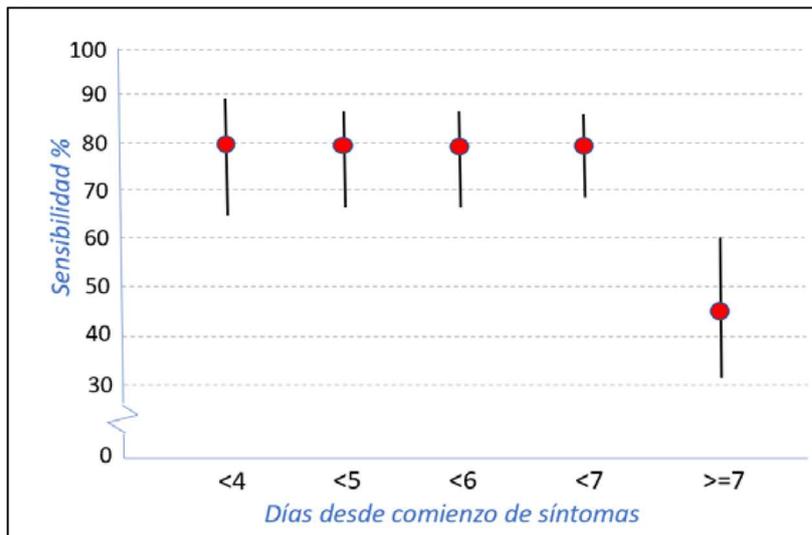
ndis: pacientes enfermos; nnondis: pacientes no enfermos y ES: sensibilidad.

La sensibilidad relativa del test de antígenos Panbio™ se ha obtenido respecto a la RT-PCR (es decir, se considera que la RT-PCR es el Gold-Standard), por lo tanto, la sensibilidad absoluta del Panbio™ es de 0.73 con un (IC95%: 0.70-0.76) para pacientes sintomáticos con menos de 7 días de evolución (considerando que la sensibilidad de la RT-PCR en nuestro hospital es de 0.88).

Nota: la sensibilidad absoluta se obtiene calculando $0.88 \times 0.83 = 0.73$

Sólo uno de los estudios, el de Masiá et al⁽¹⁸⁴⁾, describe el efecto de la duración de los síntomas en la sensibilidad de la prueba de antígeno. En la Figura 20, se observa que la sensibilidad permanece más o menos estable hasta el 7º día desde el comienzo de los síntomas, momento a partir del cual se aprecia un descenso notable en la sensibilidad de la prueba.

Figura 20. Efecto de la duración de los síntomas en la sensibilidad.



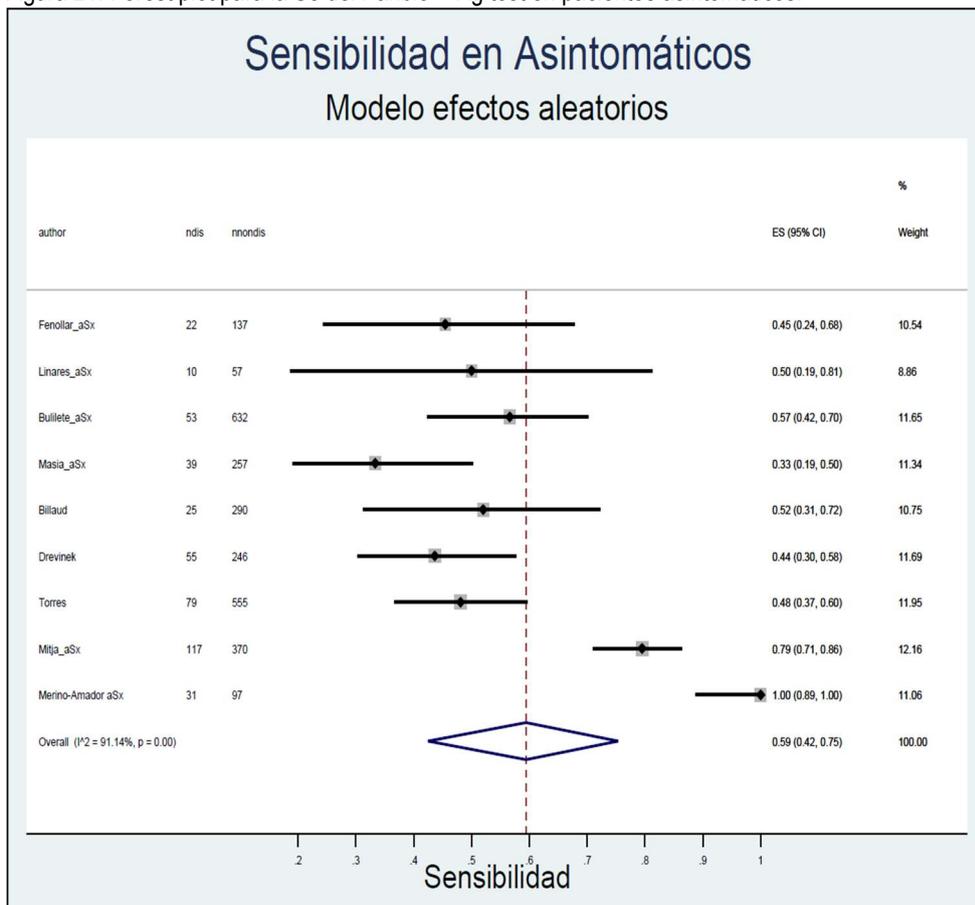
3.2.3. Combinación de estudios en pacientes asintomáticos.

Analizamos ahora la sensibilidad de la prueba de antígenos, pero en paciente asintomáticos. Consideramos para ello 9 de los 14 estudios. En la Figura 21, nuevamente, observamos una importante heterogeneidad ($I^2=91.14\%$) entre estudios, lo que resta credibilidad a la estimación conjunta. La sensibilidad obtenida es 0.59 (IC95%: 0.43-0.75).

Hay que señalar que, en la mayoría de los estudios (salvo el de Billaud, que no lo especifica), se llevan a cabo en asintomáticos que han tenido un contacto estrecho con un paciente COVID-19 positivo. En el estudio de Mitja (Alemany) hemos seleccionado los pacientes que ellos definen como *asintomáticos screening*, dado que su prevalencia está más próxima -aún y todo mayor- al resto de estudios que el grupo que los autores definen como "contactos".

Nota: se considera contacto estrecho o próximo a una persona que haya estado con un positivo en COVID-19 durante más de 15 minutos, sin mascarilla y sin mantener la distancia de seguridad (establecida en 1.5 m).

Figura 21. Forest plot para la Se del Panbio™ Ag test en pacientes asintomáticos.



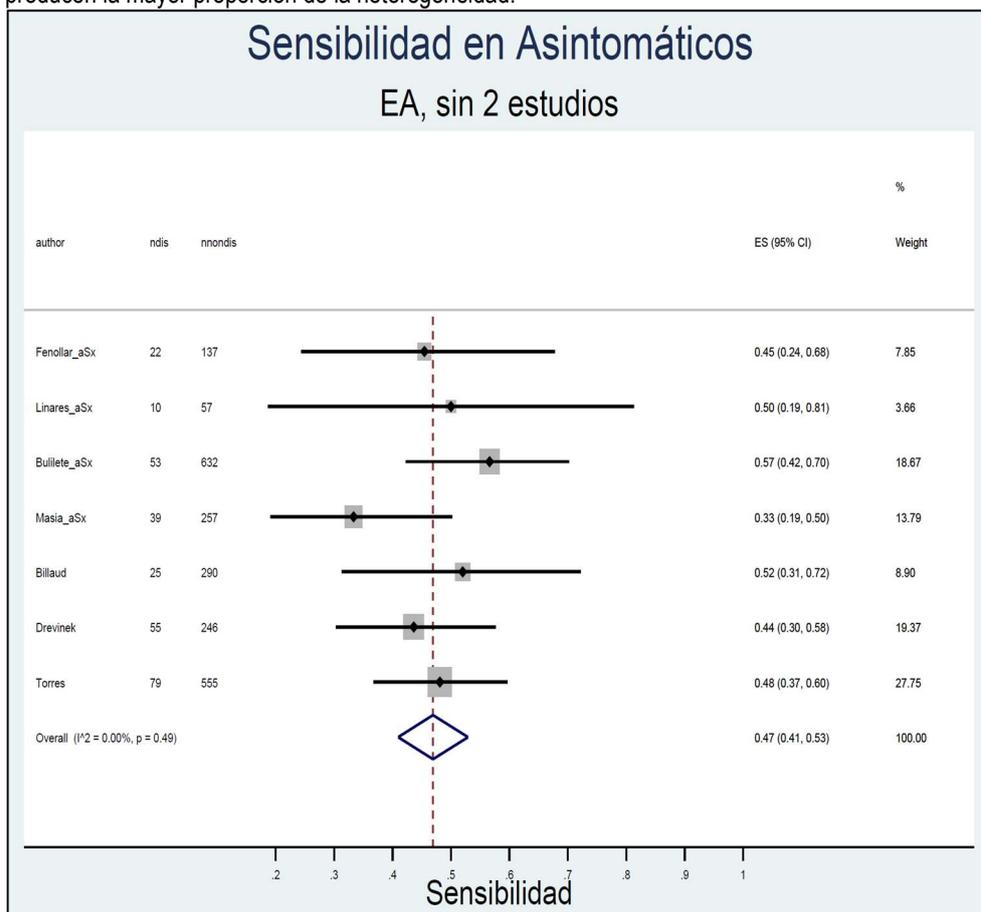
ndis: pacientes enfermos; nnondis: pacientes no enfermos y ES: sensibilidad.

En la Figura 21 se observa claramente que 2 de los estudios (Mitja y Merino-Amador) parecían ser discordantes respecto al resto, por lo que fueron eliminados, ya que eran los que aportaban mayor heterogeneidad debido a que eran los que tenían mayores prevalencias. En la Figura 22 observamos que desaparece toda la heterogeneidad que habíamos detectado anteriormente, obteniendo una $I^2=0.0\%$. La estimación de la sensibilidad relativa fue de 0.47 con un (IC95%: 0.41-0.53).

Se calculó la sensibilidad absoluta del Panbio™ Ag test, suponiendo con optimismo la misma sensibilidad de la RT-PCR en pacientes asintomáticos, y obtuvimos una $Se=0.41$ con un (IC95%: 0.36-0.47), claramente insuficiente puesto que un 59% de los asintomáticos con RT-PCR positiva serán descartados por el Panbio™.

Nota: la sensibilidad absoluta se ha calculado haciendo $0.47 \times 0.88 = 0.41$

Figura 22. Forest plot para la Se del Panbio™ Ag test en pacientes asintomáticos sin 2 estudios que producen la mayor proporción de la heterogeneidad.



ndis: pacientes enfermos; nndis: pacientes no enfermos y ES: sensibilidad.

3.2.4. Sensibilidad de la prueba rápida de antígenos Panbio™ (Ag Rapid Test Device) relativa a la RT-PCR, con datos de pacientes de la 1ª ola del HUD.

Tras analizar los resultados de la RS anterior, nos centramos en el estudio que llevó a cabo el Instituto de Salud Carlos III (ISCIII) y que se ha tenido en cuenta en la RS. Merino-Amador realizó un estudio multicéntrico en 10 hospitales universitarios, en dos comunidades autónomas (Madrid y País Vasco) de España. Los hospitales de Madrid que participaron en este estudio fueron: Hospital Universitario Ramón y Cajal, Hospital Clínico Universitario San Carlos, Hospital Universitario La Paz, Hospital Universitario Doce de Octubre y Hospital General Universitario Gregorio Marañón. Los hospitales del País Vasco, que engloban la Red de diagnóstico biológico de Osakidetza, y que participaron en el estudio son: Hospital Universitario de Álava, Hospital Universitario de Cruces, Hospital Universitario de Basurto, HUD y Hospital Universitario de Galdakao-Usansolo.

En total, se obtuvo una muestra de 958 pacientes, con sospecha clínica o epidemiológica. Se seleccionaron muestras de pacientes para los que no hubieran pasado más de 7 días desde el comienzo de los síntomas. Las muestras se realizaron mediante exudado nasofaríngeo y se realizaron pruebas diagnósticas de RT-PCR simultáneamente.

Dentro del estudio del ISCIII, en el subconjunto de pacientes de Osakidetza, se obtuvieron los siguientes datos (límites exactos de los IC95% -binomial-) (Tabla 47).

Tabla 47. Subgrupo Osakidetza del HUD del estudio realizado por el ISCIII para conocer la Se de la prueba rápida de antígenos Panbio™.

	RT-PCR ⊕	RT-PCR ⊖	Total
Ag ⊕	33	0	33
Ag ⊖	6	69	75
Total	39	69	108

Se = 84.6% (IC95%: 69.5-94.1)

Sp = 100% (IC95%: 94.8-100)

En la Tabla 48, resumimos los datos obtenidos en el estudio del ISCIII, en el subgrupo de pacientes del HUD y mostramos los datos que proporcionó la empresa ABBOTT sobre el rendimiento diagnóstico de la prueba rápida de antígenos Panbio™.

Tabla 48. Datos de Se y Sp del Estudio del ISCIII, del Subgrupo Osakidetza y Abbott.

	n	Sensibilidad	Especificidad
Estudio ISCIII	958	90.5% (IC95%: 86.0-93.6)	98.8% (IC95%: 98.0-99.7)
Subgrupo Osakidetza	108	84.6% (IC95%: 69.5-94.1)	100% (IC95%: 94.8-100)
Datos de ABBOTT Panbio® Ag Rapid Test	241	93.3% (IC95%: 83.8-98.2)	99.4% (IC95%: 97-100)

Cabe destacar que tanto en el estudio del ISCIII como en el estudio realizado por la empresa ABBOTT no se recogieron las muestras "point of care". Ambos estudios se realizaron con muestras proporcionadas por el laboratorio.

3.3. PRUEBA DE LA DETERMINACIÓN DE ANTICUERPOS (SEROLOGÍA)

De las 32.566 pruebas de determinación de anticuerpos (serologías) realizadas en el HUD durante la primera ola, 1.425 fueron positivas (4.38%) y 31.010 fueron negativas (95.22%). 131 (0.4%) de las pruebas obtuvieron un resultado: dudoso, muestras no recibida, no procede, no realizado o no valorable. A este grupo lo denominamos "El resto".

Según estos datos, obtenemos el siguiente resumen (Tabla 49 y 50).

Tabla 49. Resultados de la prueba de determinación de anticuerpos (Serología ICT) realizada en el HUD en la 1ª ola.

	Resultados de la Serología			
	⊕	⊖	El resto	Total
Nº pruebas	1.425	31.010	131	32.566
Pacientes	1.425	31.010	131	32.566

Tabla 50. Resultados de la prueba de determinación de anticuerpos (Serología ICT) realizada en el HUD en la 1ª ola sin considerar las pruebas con resultado dudoso, no procede, muestra no recibida, no realizada o no valorable.

	Resultados de la Serología		
	⊕	⊖	Total
Nº pruebas	1.425	31.010	32.435
Pacientes	1.425	31.010	32.435

3.3.1. Sensibilidad de la prueba de determinación de anticuerpos (Serología), relativa a la RT-PCR, con datos de pacientes de la 1ª ola del HUD.

Se consideraron sólo a aquellos pacientes a los que se les habían realizado tanto la prueba RT-PCR como la serología, y obtuvimos un total de 23.121 pacientes (Tabla 51).

Tabla 51. Pacientes con ambas pruebas (RT-PCR y Serología) en la 1ª ola.

Serología	Resultados de la RT-PCR			Se=66.0% IC 95% (62.4 – 69.5)	Sp=96.8% IC 95% (96.5 – 97.0)	VPP=38.3% IC 95% (35.6 – 41.1)	VPN=98.9% IC 95% (98.8 – 99.1)
	⊕	⊖	Total				
⊕	449	723	1.172				
⊖	231	21.718	21.949				
Total	680	22.441	23.121				

La proporción de fallos es del 4.13% (Fallos (%) = $((723 + 231) / 23.121) * 100$).

Comprobamos que, en algunas ocasiones, la serología era previa a la RT-PCR y pensamos que no era adecuado tener en cuenta esas pruebas en el análisis. Por dicho motivo, seleccionamos los datos que cumplían que la RT-PCR era previa a la serología y que los días transcurridos entre ambas pruebas fueran ≥ 15 días. Con esta premisa, obtuvimos los datos que se muestran en la Tabla 52. Calculamos a partir de ellos, la Se y la Sp con sus respectivos intervalos de confianza al 95%.

Tabla 52. Pacientes con ambas pruebas (RT-PCR y Serología) en la 1ª ola y se cumple que la RT-PCR es previa a la Serología y el tiempo transcurrido entre ambas pruebas es ≥ 15 días.

Serología	Resultados de la RT-PCR			Se=63.0% IC 95% (58.9 – 66.9)	Sp=99.8% IC 95% (99.8 – 99.9)	VPP=91.7% IC 95% (88.5 – 94.0)	VPN=98.9% IC 95% (98.7 – 99.0)
	⊕	⊖	Total				
⊕	352	32	384				
⊖	207	18.715	18.922				
Total	559	18.747	19.306				

La proporción de fallos es del 1.24% (Fallos (%) = $((207 + 32) / 19.306) * 100$).

La prueba de determinación de anticuerpos relativa a la RT-PCR, cuando se consideró que la serología se había realizado posterior a la RT-PCR con un mínimo de 15 días entre ambas pruebas tuvo una sensibilidad del 63% con un IC95% (58.8%-66.9%). La especificidad fue del 99.8% con un IC95% (99.8%-99.9%).

4. SCORES CREADOS EN EL HUD.

4.1. SCORE DE SOSPECHA CLÍNICA (SSC)

4.1.1. Uso del SSC para discriminar entre enfermos COVID-19 y no enfermos COVID-19.

Se evaluó si el SSC permitía discriminar entre pacientes enfermos y no enfermos por la COVID-19. Para realizar este análisis, se consideraron 987 pacientes a los que se les cumplimentó formulario COVID-19 durante la primera ola de la pandemia, es decir, consideramos a aquellos pacientes que presentaban síntomas compatibles con la enfermedad o eran sospechosos de padecerla. De esos pacientes, 479 tuvieron una RT-PCR positiva y 483 una negativa (para 25 pacientes no pudimos obtener el resultado de la RT-PCR o no fue realizada). A su vez, 109 pacientes tuvieron un diagnóstico confirmado al ingreso y a 150 no se les registró inicialmente el SSC. La sospecha que se le asignó a cada uno de ellos está reflejada en la siguiente tabla (Tabla 53).

Tabla 53. SSC inicial en relación al resultado de la RT-PCR.

Score de sospecha inicial	RT-PCR						Total
	⊕	%	⊖	%	.	%	
Diagnóstico confirmado	99	10.03	7	0.71	3	0.30	109
Muy alta	101	10.23	56	5.67	1	0.10	158
Alta	116	11.75	146	14.79	5	0.51	267
Media	36	3.65	135	13.68	4	0.41	175
Baja	20	2.03	101	10.24	7	0.70	128
.	107	10.84	38	3.85	5	0.51	150
Total	479	48.53	483	48.94	25	2.53	987

Por lo tanto, la tabla resumida (Tabla 54) sin considerar a los pacientes que tenían diagnóstico confirmado y sin considerar a aquellos a los que no se les asignó SSC o los que no tenían resultado de la RT-PCR, queda del siguiente modo:

Tabla 54. SSC inicial en relación al resultado de la RT-PCR (simplificada).

Score de sospecha inicial	RT-PCR				Total
	⊕	%	⊖	%	
Muy alta	101	14.21	56	7.88	157
Alta	116	16.32	146	20.53	262
Media	36	5.06	135	18.99	171
Baja	20	2.81	101	14.20	121
Total	273	38.40	438	61.60	711

Como cabría esperar, cuanto mayor es la sospecha mayor es el porcentaje de resultados positivos en la prueba diagnóstica RT-PCR y el porcentaje de resultados negativos es mayor cuando la sospecha es más baja.

En base a estos datos, establecimos diferentes puntos de corte en el SSC para poder conocer cuál era el punto de corte óptimo.

El primer punto de corte que establecimos fue, "Sospecha muy alta" vs "Sospecha alta, media o baja" (Tabla 55).

Tabla 55. "Sospecha muy alta" vs "Sospecha alta, media o baja" en relación al resultado de la RT-PCR.

Score de sospecha inicial	RT-PCR		Total
	⊕	⊖	
Muy alta	101	56	157
Alta, Media o baja	172	382	554
Total	273	438	711

Se = 37% IC 95%: (30.1 – 42.3)

Sp = 87.2% IC 95%: (83.9 – 90.6)

Índice J de Youden = 0.24 (Se + Sp – 1)

Posteriormente, establecimos como punto de corte "Sospecha muy alta o alta" vs "Sospecha media o baja" (Tabla 56).

Tabla 56. "Sospecha muy alta o alta" vs "Sospecha media o baja" en relación al resultado de la RT-PCR.

Score de sospecha inicial	RT-PCR		Total
	⊕	⊖	
Muy alta o alta	217	202	419
Media o baja	56	236	292
Total	273	438	711

Se = 79.5% IC 95%: (74.5 – 84.8)

Sp = 53.9% IC 95%: (50.0 – 59.9)

Índice J de Youden = 0.33 (Se+ Sp – 1)

Por último, establecimos el punto de corte “Sospecha muy alta o alta o media” vs “Baja” (Tabla 57).

Tabla 57. “Sospecha muy alta o alta o media” vs “Sospecha Baja” en relación al resultado de la RT-PCR.

Score de sospecha inicial	RT-PCR		Total	Se = 92.7% Sp = 23.1%	IC 95%: (87.9 – 95.0) IC 95%: (18.5 – 26.9)
	⊕	⊖			
Muy alta, alta o media	253	337	590	Índice J de Youden = 0.16 (Se+ Sp – 1)	
Baja	20	101	121		
Total	273	438	711		

Se seleccionó como punto óptimo el que proporcionó mayor valor en el índice de Youden (es el punto en el que se maximiza la Se y la Sp), que fue: “Sospecha muy alta o alta” vs “Sospecha media o baja”. En dicho punto, la Se es del 79.5% (IC95%: 74.5-84.8) y la Sp de 53.9% (IC95%: 50.0-59.9).

4.1.2. Uso del SSC para discriminar pacientes que se van a positivar de los que no, de entre los que tienen RT-PCR ⊖,

Se analizó si era posible que el SSC permitiera saber qué pacientes, de los que tenían RT-PCR negativa, se iban a positivar. Basándonos en los datos de la Figura 9 construimos la siguiente tabla (Tabla 58).

Tabla 58. SSC inicial en relación a la positividad del resultado de la RT-PCR.

Score de sospecha inicial	RT-PCR	
	⊖⊕	⊖⊖
Diagnóstico confirmado	2	4
Muy alta	17	48
Alta	15	104
Media	5	93
Baja	1	48
.	9	24
Total	49	321

De los 49 pacientes que tuvieron RT-PCR negativa y luego tuvieron alguna RT-PCR positiva, 2 tenían diagnóstico confirmado al ingreso. No se anotó la sospecha inicial al ingreso de 9 pacientes. De los 321 con al menos 2 RT-PCR negativas que no tuvieron ninguna RT-PCR positiva, 4 tuvieron diagnóstico confirmado y no se anotó la sospecha inicial de 24 pacientes. En base a estos datos, simplificamos la tabla anterior (Tabla 59).

Tabla 59. SSC inicial en relación a la positividad del resultado de la RT-PCR (simplificada).

Score de sospecha inicial	RT-PCR	
	$\ominus\oplus$	$\oplus\oplus$
Muy alta	17	48
Alta	15	104
Media	5	93
Baja	1	48
Total	38	293

Se construyó la tabla 2x2 para analizar si existía un punto de corte que permitiera discriminar a los pacientes que se positivizaban de los que no. Hicimos para ello, los mismos puntos de corte que establecimos en el apartado anterior y, calculamos, la Se, la Sp y el índice de Youden para cada uno de ellos.

Nota: se consideraron los límites exactos de la binomial para el cálculo de los IC al 95%.

El primer punto de corte que establecimos fue, "Sospecha muy alta" vs "Sospecha alta, media o baja" (Tabla 60).

Tabla 60. "Sospecha muy alta" "Sospecha alta, media o baja" vs positividad del resultado de la RT-PCR.

Score de sospecha inicial	RT-PCR		Total
	$\ominus\oplus$	$\oplus\oplus$	
Muy alta	17	48	65
Alta, media o baja	21	245	266
Total	38	293	331

Se = 44.7% IC 95% (28.6 – 61.7)

Sp = 83.6% IC 95% (78.6 – 87.4)

VPP = 26.2% IC 95% (16.3 – 38.5)

VPN = 92.1% IC 95% (87.9 – 95.4)

Índice J de Youden = 0.28 (Se+Sp-1)

Posteriormente, establecimos como punto de corte "Sospecha muy alta o alta" vs "Sospecha media o baja" (Tabla 61).

Tabla 61. "Sospecha muy alta o alta" "Sospecha media o baja" vs positividad del resultado de la RT-PCR.

Score de sospecha inicial	RT-PCR		Total
	$\ominus\oplus$	$\oplus\oplus$	
Muy alta o alta	32	152	184
Media o baja	6	141	147
Total	38	293	331

Se = 84.2% IC 95% (68.8 – 94.0)

Sp = 48.1% IC 95% (40.9 – 52.5)

VPP = 17.4% IC 95% (12.5 – 24.2)

VPN = 95.9% IC 95% (91.4 – 98.1)

Índice J de Youden = 0.32 (Se+Sp-1)

Por último, establecimos el punto de corte "Sospecha muy alta o alta o media" vs "Sospecha baja" (Tabla 62).

Tabla 62. "Sospecha muy alta, alta o media" "Sospecha baja" vs positivización del resultado de la RT-PCR.

Score de sospecha inicial	RT-PCR		Total
	$\ominus\oplus$	$\ominus\ominus$	
Muy alta, alta o media	37	245	282
Baja	1	48	49
Total	38	293	331

Se = 97.4% IC 95% (86.2 – 99.9)
 Sp = 16.4% IC 95% (12.5 – 20.7)
 VPP = 13.1% IC 95% (9.40 – 17.6)
 VPN = 98.0% IC 95% (89.2 – 99.6)

Índice J de Youden = 0.14 (Se+Sp-1)

El mejor punto de corte se obtuvo cuando clasificamos a los pacientes en "Sospecha Muy alta o Alta" vs "Sospecha Media o Baja". En cuyo caso, la Se=84.2% (IC 95%: 68.8–94.0) y la Sp=48.1% (IC 95% 40.9–52.5). El VPP=17.4% y el VPN=95.9%, lo que indica que al 4% de los pacientes les diremos que no se van a positivizar si el SSC inicial es medio o bajo y sí lo harán.

4.1.3. Estudio de la edad para mejorar la discriminación que hace el SSC entre pacientes que se van a positivizar de los que no.

A la vista de los resultados anteriores, estudiamos si había variables, por ej: la edad, que ayudasen a mejorar la discriminación entre los pacientes que se iban a positivizar de los que no. Para ello, seleccionamos el mejor punto de corte que habíamos encontrado para el SSC y desglosamos en función de los grupos de edad. Primero analizamos: "<75 años" vs ">75 años" y, después, "<65 años" vs ">65 años" (Tabla 63).

Tabla 63. Comparativa del SSC para discriminar la positivización de pacientes según grupos de edad (corte 75 años).

Score de sospecha clínica Muy alta o alta			Total	Score de sospecha clínica Media o baja			Total
Edad	RT-PCR $\ominus\oplus$	RT-PCR $\ominus\ominus$		Edad	RT-PCR $\ominus\oplus$	RT-PCR $\ominus\ominus$	
≥ 75 años	9	59	68	≥ 75 años	1	65	66
< 75 años	23	84	107	< 75 años	5	75	80
Total	32	143	175	Total	6	140	146

Se = 28.1% IC 95% (13.8 – 46.8)
 Sp = 58.7% IC 95% (51.4 – 68.1)

Se = 16.7% IC 95% (0.42 – 64.1)
 Sp = 53.6% IC 95% (44.9 – 62.0)

Se realizó la comparativa de ambas sensibilidades aplicando el test exacto de Fisher (Tabla 64) y se comprobó que no había diferencias estadísticamente significativas ($p=n.s.$) entre los pacientes con 75 años o más y los de menos de 75 años, es decir, la edad no mejoraba la discriminación que ya hacía el SSC entre los pacientes que se positivizaban y los que no.

Tabla 64. Test exacto de Fisher para comparar las sensibilidades obtenidas para "SSC Muy alta o alta" vs "SSC Media o baja", en pacientes que se positivizan, con el corte de edad en 75 años.

Edad	RT-PCR $\ominus\oplus$		Total
	SSC Muy alta o alta	SSC Media o baja	
≥ 75 años	9	1	10
< 75 años	23	5	28
Total	32	6	38

$p = n.s.$

Considerando el corte de edad en 65 años (Tabla 65).

Tabla 65. Comparativa del SSC para discriminar la positivización de pacientes según grupos de edad (corte 65 años).

Score de sospecha clínica Muy alta o alta			Total
Edad	RT-PCR $\ominus\oplus$	RT-PCR $\ominus\ominus$	
≥ 65 años	17	90	107
< 65 años	15	53	68
Total	32	143	175

Se = 53.1% IC 95% (34.7 – 70.9)

Sp = 37.1% IC 95% (29.9 – 46.5)

Score de sospecha clínica Media o baja			Total
Edad	RT-PCR $\ominus\oplus$	RT-PCR $\ominus\ominus$	
≥ 65 años	3	97	100
< 65 años	3	43	46
Total	6	140	146

Se = 50.0% IC 95% (11.8 – 88.2)

Sp = 30.7% IC 95% (23.3 – 39.1)

Se compararon las sensibilidades obtenidas para cada grupo de edad (Tabla 66) y obtuvimos que no existían diferencias estadísticamente significativas ($p=n.s.$).

Tabla 66. Test exacto de Fisher para comparar las sensibilidades obtenidas para “SSC Muy alta o alta” vs “SSC Media o baja”, en pacientes que se positivizan, con el corte de edad en 65 años.

Edad	RT-PCR $\ominus\oplus$		Total
	SSC Muy alta o alta	SSC Media o baja	
≥ 65 años	17	3	20
< 65 años	15	3	18
Total	32	6	38

$p = n.s.$

Por lo tanto, concluimos que la edad no nos ayuda a mejorar la discriminación que ya hacía el SSC para diferenciar entre los pacientes que se iban a positivizar de los que no.

4.1.4. Estudio de algunas de las variables que componen el SSC para mejorar la discriminación que hace el SSC entre pacientes que se van a positivizar y los que no.

Con los análisis anteriores supimos que el SSC podía ayudarnos a arrojar luz acerca de qué pacientes presentaban o no la enfermedad y cuáles se iban a positivizar.

Surgió la idea de que la presencia de síntomas, en las 48h previas a una RT-PCR \ominus , en los pacientes ingresados, podía mejorar la discriminación entre aquellos que se iban a positivizar en los 7 días siguientes tras haberle realizado una RT-PCR con un resultado \ominus .

Basándonos en los datos de la Figura 9, seleccionamos:

- Pac. sólo con RT-PCR \ominus : 433. RT-PCR \ominus en las 48h tras ingreso: 426 pac.
Más de 1 RT-PCR \ominus : 321 pac.
- Pac. con alguna RT-PCR \ominus y alguna RT-PCR \oplus : 288.
 $\ominus\oplus$: 60 pac. \rightarrow RT-PCR \ominus \leq 48h tras ingreso: 59 pac. \rightarrow RT-PCR \oplus \leq 7días: 49 pac.

Para este análisis seleccionamos los síntomas: fiebre, tos o disnea por ser los más frecuentemente detectados en los pacientes que padecían la enfermedad de la COVID-19.

Analizamos los síntomas de manera conjunta:

1. Consideramos si los pacientes habían tenido alguno de los síntomas: “Fiebre o Tos o Disnea” y obtuvimos la tabla 67:

Tabla 67. Fiebre o Tos o Disnea vs positivización de la RT-PCR.

FIEBRE o TOS o DISNEA	RT-PCR ⊕⊕	RT-PCR ⊖⊖	TOTAL
Sx	38	297	335
No Sx	11	24	35
TOTAL	49	321	370

Se=77.6% (IC 95%:63.38 – 88.23)

Sp=7.5% (IC 95%:4.92 – 11.30)

VPP=11.3% (IC 95%:7.98 – 15.49)

VPN= 68.6% (IC 95%: 50.71 – 83.15)

2.- Consideramos si los pacientes habían tenido “Fiebre y (Tos o Disnea)” y obtuvimos la tabla 68:

Tabla 68. Fiebre y (Tos o Disnea) vs positivización de la RT-PCR

FIEBRE y (TOS o DISNEA)	RT-PCR ⊕⊕	RT-PCR ⊖⊖	TOTAL
Sx	25	121	146
No Sx	24	200	224
TOTAL	49	321	370

Se=51.2% (IC 95%:36.34 – 65.58)

Sp=62.3% (IC 95%: 57.6 - 68.78)

VPP=17.1% (IC 95%:11.1 – 23.64)

VPN=89.3% (IC 95%: 84.6 – 92.7)

Analizamos los síntomas de manera individualizada (Tabla 69, 70 y 71):

Tabla 69. Fiebre vs positivización de la RT-PCR.

FIEBRE	RT-PCR ⊕⊕	RT-PCR ⊖⊖	TOTAL
Sx	27	174	201
No Sx	22	147	169
TOTAL	49	321	370

Se=55.1% (IC95%:40.23 – 69.33)

Sp=45.8% (IC95%:37.67 – 49.15)

VPP=13.4% (IC 95%:9.10 – 19.05)

VPN=87.0% (IC 95%:81.1 – 91.2)

Tabla 70. Disnea vs positivización de la RT-PCR.

DISNEA	RT-PCR ⊕⊕	RT-PCR ⊖⊖	TOTAL
Sx	26	190	216
No Sx	23	131	154
TOTAL	49	321	370

Se=59.2% (IC 95%: 44.21 – 73.0)

Sp=43.3% (IC 95%: 37.67 - 49.15)

VPP=13.7% (IC 95%: 9.10 – 19.05)

VPN=87.4% (IC 95%: 81.40 – 91.7)

Tabla 71. Tos vs positivización de la RT-PCR.

TOS	RT-PCR ⊖⊕	RT-PCR ⊖⊖	TOTAL	
Sx	29	182	211	Se=53.1% (IC 95%: 38.27 – 67.47)
No Sx	20	139	159	Sp=40.8% (IC 95%: 34.44 – 45.79)
TOTAL	49	321	370	VPP=12.0% (IC 95%: 7.85 – 17.35)
				VPN=85.1% (IC 95%: 78.6 – 89.8)

Ninguno de los 3 síntomas (Fiebre, Tos, Disnea), por sí solo, nos permitía discriminar entre los pacientes que se iban a positivizar de aquellos que no lo iban a hacer. Tampoco la combinación de síntomas, ni “Fiebre o Tos o Disnea” ni “Fiebre y (Tos o Disnea)” permitía realizar dicha discriminación. Los VPP para ambas combinaciones de síntomas fueron: VPP=11.3% (IC95%:7.98–15.49) y VPP=17.1% (IC 95%:11.1–23.64) respectivamente.

4.2. RISK SCORE

Se realizó una validación externa del Risk Score con todos los pacientes del HUD que tuvieron resultado positivo en la RT-PCR en la 1ª ola de la pandemia (n=450). Los resultados obtenidos se muestran en la tabla 72.

Finalmente, sólo pudimos aplicar el score a un total de 105 pacientes debido a la gran cantidad de datos faltantes en las variables que conforman dicho score.

Tabla 72. Resumen de la validación del Risk Score en nuestra población.

Score	Resultado	n	Variables	AUC (IC95%)
Risk Score	Éxito Críticos	105	PAS, FC, FR, Nivel de conciencia	0.62 (0.42 – 0.62)

n = nº de individuos que participan en la validación externa de la regla.

A la vista de los resultados pudimos concluir que este score no nos ayudaba a predecir qué pacientes, de los que ingresaban en el HUD con COVID-19 positivo, tenían mayor riesgo de sufrir un evento grave (ingreso en UCI o muerte).

DISCUSIÓN

DISCUSIÓN

- MODELOS PREDICTIVOS
 - Dificultades durante el desarrollo de la tesis
 - Uso de la herramienta PROBAST para la evaluación de riesgo de sesgo
 - Utilidad clínica
- PRUEBAS DIAGNÓSTICAS DE LABORATORIO
 - Prueba RT-PCR
 - Prueba rápida de antígenos PANBIO™
 - Prueba de determinación de anticuerpos (Serología)
- SCORES CREADOS POR LA UNIDAD DE MEDICINA INTERNA
 - Score de sospecha clínica (SSC)
 - Risk score
- LIMITACIONES
- ACCIONES FUTURAS

1. MODELOS PREDICTIVOS

Los objetivos 1 y 2 de esta tesis han sido proporcionar a los clínicos herramientas que permitieran predecir la presencia de infección por SARS-CoV-2 en pacientes con sospecha de infección, y modelos para predecir el pronóstico o el curso de la infección (ingreso en UCI o desenlace fatal) en los pacientes con la enfermedad.

1.1. DIFICULTADES DURANTE EL DESARROLLO DE LA TESIS

Consideramos que ha habido 3 cuestiones que han dificultado el desarrollo de esta tesis en lo que respecta al apartado de modelos predictivos. Uno de los grandes inconvenientes ha sido la falta de cumplimentación de algunas variables de las bases de datos clínicas de nuestra OSI para ciertos predictores que se empleaban en los modelos predictivos publicados en las fechas consultadas. Esto es debido a que, por un lado, antes de la creación de los formularios COVID-19 en el HUD y en AP no se consideraron todas las variables que posteriormente se fueron estableciendo como relevantes para estudiar la enfermedad. Y, por otro, a que no se han cumplimentado los formularios de la manera que habría sido deseable. Ha habido un elevado porcentaje de datos que no se han recogido o que se han registrado de forma incorrecta.

Para intentar minimizar la ausencia de datos en algunas de las variables extraídas de los sistemas de información clínica, se decidió revisar las historias clínicas de los pacientes, pero no fue posible recuperar gran cantidad de información porque no había sido registrada previamente.

Se sabe que éste es un problema habitual cuando se trabaja con bases de datos asistenciales y clínicas, aunque consideramos que, en este caso concreto, puede haberse visto agravado por la sobrecarga de trabajo que han sufrido los profesionales durante la pandemia y que eso puede haber propiciado que la recogida de datos se haya realizado de manera incompleta o errónea.

Debido a la falta de datos de variables relevantes para la validación de los modelos identificados, se decidió no validar aquellos modelos para los que no disponíamos de los predictores que se empleaban en ellos. Esto se hizo así porque sabemos que los estudios que tienen como objetivo validar externamente un modelo de predicción existente tienen un RoB alto cuando

faltan datos de predictores en el momento de la validación y los investigadores validan el modelo de todos modos omitiendo estos predictores faltantes. Este es un fallo común en los estudios de validación y produce efectivamente resultados de validación para otro modelo en lugar de para el modelo previsto desarrollado originalmente.

Otro inconveniente que hemos encontrado en la realización de este estudio es, que un 20% de los estudios elegibles no proporcionaban el modelo y, por lo tanto, no permitían su validación. Creemos que los revisores deberían establecer como criterio indispensable antes de publicar un modelo que éste sea reportado de manera que pueda ser reproducible. Entendemos que hacer una crítica en este sentido, teniendo en cuenta las circunstancias que se han vivido durante la pandemia, puede ser arriesgado y comprendemos que los criterios de publicación que se han considerado pueden haber sido, en determinados momentos, excepcionales debido a la necesidad que existía de compartir la poca información disponible sobre la enfermedad y la urgencia de actuar para paliar sus efectos. A pesar de esto, creemos que para futuras ocasiones se deberían exigir unos criterios mínimos para llevar a cabo una publicación de este tipo de modelos.

Un tercer inconveniente que puede haber influido en este apartado ha sido que no existen protocolos sobre cómo se deben derivar y/o validar los modelos predictivos. No existe consenso sobre muchas de las cuestiones que determinan el modelo final obtenido y esto dificulta en gran medida valorar si los estudios han sido derivados o validados correctamente.

Tras este análisis preliminar, en términos generales, consideramos que se han obtenido resultados alentadores tras las validaciones externas llevadas a cabo con los pacientes de nuestra OSI. Tanto para el ANDC score como para el CSS score, se obtuvieron buenas medidas de discriminación con valores de $AUC > 0.8$, aunque hay aspectos relevantes en el desarrollo y en la validación de los modelos analizados en esta tesis que creemos que pueden haber afectado al rendimiento obtenido para dichos modelos.

1.2. USO DE LA HERRAMIENTA PROBAST PARA LA EVALUACIÓN DE RIESGO DE SESGOS

A través del uso de la herramienta PROBAST, hemos evaluado el riesgo de sesgos en los 4 dominios: pacientes, predictores, outcome y análisis y la aplicabilidad de los modelos. En todos los estudios analizados se ha obtenido bajo riesgo de sesgo en los 3 primeros dominios, pero alto riesgo de sesgo en el cuarto (salvo el 4C score). Esto es debido a que ha habido cuestiones, dentro del procedimiento de derivación y/o validación de los modelos, que se han realizado de manera incorrecta.

Los autores del estudio 4C score son los únicos que han proporcionado la evaluación PROBAST, tanto para derivación como para validación, e informaron los apartados correspondientes a las cuestiones planteadas en dicha herramienta, pero no reportaron la aplicabilidad. La puntuación otorgada para el dominio *análisis* fue "+" (es decir, bajo riesgo de sesgos). Nosotros aceptamos dicha evaluación, aunque fue completada con los apartados *RoB* y *applicability* que no habían sido reportados. El estudio de Knight es el único, de todos los estudios seleccionados, que recibe una buena calificación en el dominio 4, aunque el equipo COVID PRECISE, cuando evaluó el estudio, concluyó que no quedaba claro cómo se pasaba de la puntuación total obtenida para cada paciente a la escala que permitía clasificar al paciente en un rango concreto.

El score ANDC también tuvo una valoración bastante buena en el dominio *análisis*, pero puntuamos como alto riesgo de sesgo el apartado 4.1, que hace referencia a si el número de eventos es adecuado. Tan sólo 20 pacientes de la muestra (n=176) presentaban el evento y consideramos que dicho número debería haber sido mayor.

Wynants reporta que el 94% de los estudios tienen un alto riesgo de sesgos en el análisis. Nosotros hemos obtenido también un elevado nº de estudios con riesgo de sesgo. Creemos que los sesgos que se han producido en los estudios validados, principalmente, los sesgos en el *análisis*, son debidos a que falta mucho conocimiento acerca del desarrollo y de la validación de los modelos predictivos. Una manera de intentar reducir estos sesgos sería incluyendo en los equipos de investigación a personas con la formación adecuada, como son los bioestadísticos, o a personas con suficientes conocimientos estadísticos.

De hecho, en la herramienta PROBAST se especifica que algunos de los aspectos considerados en el dominio de análisis requieren que el evaluador tenga conocimientos avanzados de la materia, por lo que se recomienda que dicho dominio sea evaluado al menos por 1 investigador con experiencia estadística en modelos predictivos.

Otro motivo por el que se ha otorgado una puntuación alta en RoB en los estudios evaluados es, porque la calidad de los informes en los artículos publicados que describen el desarrollo y/o la validación de modelos de predicción multivariable es deficiente. Y en ausencia de informes detallados y transparentes de los detalles clave del estudio, es difícil para la comunidad científica y sanitaria juzgar objetivamente las fortalezas y debilidades de un estudio de modelo de predicción.

A continuación, vamos a detallar algunos de los aspectos que consideramos más relevantes y en los que hemos encontrado la mayoría de las deficiencias de los estudios revisados:

1.2.1. Tamaño muestral.

Llama la atención que, a pesar de ser el tamaño muestral un aspecto relevante, sólo se ha hecho el cálculo previo del tamaño muestral mínimo requerido en dos estudios, PREDI-CO y Zhang score, pero lo hacen de manera incorrecta. En el resto de los estudios, la ausencia de este cálculo podría estar justificada, dadas las circunstancias tan excepcionales que se vivían en el momento en el que los scores fueron derivados.

Creemos que los autores habrán considerado que lo importante era tener en cuenta a todos los pacientes que se pudieron reclutar en dicho momento, y suponemos que en algunos casos habrá habido tamaños que fueran inadecuados, pero era el número de pacientes que tenían disponibles.

Nosotros hemos calculado cuál sería el tamaño muestral mínimo que se debería haber considerado para derivar cada uno de los modelos analizados (Tabla 30), y tras obtener los EPP, concluimos que parece que no han sido tamaños adecuados los que se han empleado en los siguientes estudios (ya que los EPP son especialmente bajos): score ANDC (EPP=3.86), Zhang (EPP=6.30), Pre (EPP=3.22) y HNC-LL (EPP=5.67).

En el caso del score ANDC, sólo 20 pacientes presentan el evento (éxito) en la población de derivación. Si se aplica la regla de los 10 pacientes por cada variable predictora en el grupo del outcome que menos frecuencia presenta (consideración que algunos autores emplean), sólo se deberían haber incluido 2 predictores en el modelo final, y no 4 que son los que se han incluido en el score.

Es sabido que se recomienda realizar al menos una validación externa con otros datos diferentes a los datos de derivación (seleccionando sujetos de otro periodo temporal) y, si es posible, en otro ámbito. Destacamos que, todos nuestros estudios, menos el que deriva el PRE score, han realizado validaciones externas, aunque creemos que los tamaños muestrales considerados son pequeños y, en la mayoría de los casos, no se cumple la condición descrita de tener un mínimo de 100 eventos en la muestra de validación. Llama especialmente la atención el reducido tamaño seleccionado para validar el NEWS2 (n=68).

Una limitación en la aplicabilidad de los modelos, que afecta al rendimiento predictivo, ocurre cuando la validación se lleva a cabo con un tamaño menor que el que se empleó para derivar el modelo. Esto ha sido muy habitual en los modelos predictivos validados para COVID-19, de hecho, todos los modelos analizados en nuestra RS tienen un tamaño menor en la validación que en la derivación, salvo el CSS.

Hay diferentes opiniones sobre qué tamaño muestral es el adecuado. A modo de resumen y, en función del objetivo, podemos establecer que para:

- La derivación de un modelo predictivo se debería seguir el criterio que establecen Riley et al^(42,43,45).
- La validación de un modelo predictivo: lo más importante no es el tamaño muestral sino el número de eventos del outcome incluidos en la muestra. Se establece que como mínimo deben considerarse 100 eventos, pero lo deseable serían 200.

No existe consenso sobre el mejor método de construcción de modelos, pero está ampliamente aceptada la importancia de tener un tamaño de muestra adecuado y datos de alta calidad. Por lo que hay que tener especial cuidado cuando construyamos modelos a partir de pequeños conjuntos de datos.

A medida que tengamos un mayor tamaño muestral obtendremos resultados más precisos, es decir, errores estándares más pequeños e intervalos de confianza más estrechos.

En los estudios de modelos de predicción, el tamaño de la muestra es importante, pero el número de participantes con el evento es más importante aún. El número de participantes con el resultado es una fuente potencial de sesgo ya que, no solo influye en la precisión, sino que también afecta el desempeño predictivo.

1.2.2. Selección de predictores.

Los predictores de los modelos finales incluidos en nuestra RS son marcadores objetivos, lo que disminuye la variación entre individuos, aumenta la confiabilidad y mejora el control de la calidad. La facilidad con la que se determinan estos predictores en la práctica clínica habitual y el bajo coste de la evaluación facilitan la realización de las predicciones de riesgo en la práctica diaria, lo cual es de suma importancia para el manejo de los pacientes COVID-19.

Es importante que los predictores estudiados sean claramente definidos y estandarizados, de modo que puedan ser reproducibles, generalizables y que estén disponibles para ser aplicados en la práctica clínica. Así mismo, deben ser medidos empleando técnicas habituales de la práctica habitual⁽⁹⁾.

A la hora de seleccionar los predictores candidatos, la mayoría de los autores consideran que se debería emplear el juicio clínico y la información proporcionada en estudios previos y no confiar sólo en los métodos estadísticos de selección de variables.

En los estudios incluidos en nuestra RS, los autores han seleccionado los predictores de diferentes formas, por ejemplo, para la derivación del 4C score se siguieron las recomendaciones de Wynants, que establece 3 criterios: seleccionar las variables que han demostrado estar relacionadas con los outcome de interés en casos de pacientes con neumonía y/o gripe, seleccionar biomarcadores identificados en la literatura como potenciales predictores en pacientes con COVID-19 y variables identificadas en, al menos, dos tercios de pacientes dentro de la cohorte de derivación.

Los autores del PREDI-CO y del HNC-LL score no dicen cómo seleccionan los predictores del modelo, sólo indican cuáles han seleccionado y hacen análisis univariante para seleccionar aquellos que sean significativos. Los autores del ANDC score, seleccionan según la recomendación de los clínicos y de la literatura existente. Para la derivación del score propuesto por Zhang, se seleccionan las variables basándose en la evidencia previa, significación clínica (análisis univariante), correlación de predictores y datos válidos. Y, por último, los autores del CSS score no dicen por qué seleccionan esos predictores, sólo dicen, que se quedan con los que han sido seleccionados mediante regresión de LASSO de entre los incluidos en el modelo.

Es habitual seleccionar los predictores realizando un análisis univariante (analizando la relación de cada uno de los predictores con el outcome). Aunque esta técnica es desaconsejada por diferentes autores ya que tiene algunas desventajas. Por ejemplo, si establecemos la significación estadística en $p=0.05$, podemos correr el riesgo de no seleccionar predictores que realmente sean relevantes^(31,192,193). Esto puede suceder ya que algunos predictores pueden ser importantes sólo después de ser ajustados por otros predictores en un modelo multivariante (por ejemplo, debido a un pequeño tamaño muestral).

Nosotros consideramos que el análisis univariante puede ser adecuado siempre y cuando se seleccione un valor de significación entre $p=0.20$ y $p=0.30$, en lugar de un valor 0.05 (que es el habitual).

Para la selección de los predictores finales (una vez hecha la selección previa al modelo) se suele emplear la técnica stepwise (por ej: en la regresión logística multivariante) aunque se desaconseja su uso, especialmente, cuando el número de eventos es pequeño, ya que la selección es inestable, las estimaciones de los coeficientes de regresión son demasiado extremas y el rendimiento del modelo final puede estar sobreestimado.

1.2.3. Missing data.

Hasta la fecha, se han publicado diferentes métodos óptimos para manejar los datos faltantes (missing data) en la etapa de derivación de los modelos de predicción clínica, pero existen muchas lagunas aún sobre los métodos que se deben usar en las etapas de validación e implementación externas. Tsvetanova et al⁽¹⁹⁴⁾, han llevado a cabo una RS de los estudios de modelos predictivos que se usan actualmente en el Reino Unido y han concluido que el enfoque más usado es el del análisis de casos completos (ACC). Ha sido así en un 40% de los estudios de derivación, en un 44% de los estudios de validación y en un 57% de los de implementación.

En los estudios que hemos incluido en nuestra RS se han empleado técnicas de imputación de datos para intentar reducir la pérdida de información, pero las técnicas empleadas se han informado de manera deficiente en la mayoría de los casos y las estrategias de manejo de datos faltantes fueron generalmente inconsistentes.

En 2 estudios (PREDI-CO score y NEWS2), se han considerado únicamente a los pacientes para los que se tenían datos completos (ACC), pero no se han analizado los datos faltantes para saber si eran completamente aleatorizados. Sólo se pueden excluir todos los casos con datos faltantes cuando éstos sean completamente aleatorizados y cuando el tamaño muestral sea suficiente para producir inferencias robustas. En caso contrario, pueden producirse sesgos de selección a menos que las pérdidas sean inferiores al 5%^(25,26).

En los estudios del PRE score y del HNC-LL score no se menciona esta cuestión en su reporte. El estudio del ANDC score también aplica imputación de datos faltantes y emplea la técnica random forest, que tiene la ventaja de permitir diferentes tipos de missing data, así como permitir la no linealidad de los datos. Aunque la técnica más recomendada⁽¹⁹⁵⁻¹⁹⁸⁾ es la de imputación múltiple (IM), que es adecuada cuando la pérdida de datos en la variable predictora ocurre por azar y no supera de un 20% a un 30% del total⁽³¹⁾. Esta técnica es la que se ha empleado en 3 (ANDC, CSS y Zhang) de los 8 estudios de la RS realizada.

Si se realiza la imputación de datos se recomienda presentar los resultados con y sin imputación, pero no se ha hecho en ninguno de los estudios analizados⁽³¹⁾.

1.2.4. Dicotomización de variables continuas.

En algunos de nuestros modelos finales analizados (PREDI-CO, ANDC, PRE y HNC-LL) se han incluido variables dicotomizadas. Los autores explican cómo se han obtenido los puntos de corte, salvo en el estudio que deriva el ANDC. Todos han empleado el índice de Youden para establecer el punto de corte, ya que es el método más habitual y el que proporciona mayor sensibilidad y mayor especificidad.

Sabemos que debe evitarse la dicotomización de predictores continuos^(37,199,200) porque implica pérdida de información y cuando se hace, generalmente, se determina un punto de corte arbitrario. Se suele argumentar que la dicotomización ayuda a la interpretación clínica y mantiene la simplicidad, pero un modelo de predicción que incluya predictores continuos dicotomizados puede tener una capacidad predictiva sustancialmente reducida y mayores sesgos^(37,38,199,200).

En ocasiones es inevitable el uso de este tipo de predictores, pero en muchos modelos se fuerza la dicotomización de los predictores aun cuando podrían ser considerados empleando su naturaleza continua.

1.2.5. Técnicas estadísticas.

En los estudios que se han publicado para predecir mortalidad en pacientes COVID-19, se han obtenido $AUC > 0.8$ cuando se han empleado técnicas de inteligencia artificial para derivar los modelos. De las técnicas de IA, la más empleada ha sido Machine Learning seguida de Deep Learning. Ambas técnicas son elegidas cuando hay un gran número de predictores en relación al número de eventos y cuando se dispone del tamaño muestral adecuado.

Las técnicas más habituales empleadas en medicina, para outcome binarios, son la regresión logística y la regresión de Cox. De hecho, en todos los estudios que hemos validado externamente con datos de nuestra OSI se han empleado técnicas tradicionales salvo en uno de ellos (4C score) que se han empleado regresión de LASSO y Machine Learning. En el caso del score ANDC y CSS se ha usado LASSO para reducir la dimensionalidad y seleccionar los predictores. Para posteriormente realizar una regresión logística multivariante.

1.2.6. Overfitting.

Se sabe que los métodos de penalización o shrinkage producen modelos de predicción clínica poco fiables cuando el tamaño muestral es pequeño.

En algunos casos, los métodos de penalización permiten corregir el sobreajuste (overfitting) si lo hubiera. Riley⁽²⁰¹⁾ propone usar estos métodos cuando se disponga de un conjunto de datos, para la derivación del modelo, suficientemente grande. De este modo, se minimiza el sobreajuste, se mejora la precisión de las estimaciones de los parámetros del modelo y proporciona una predicción más sólida para la práctica clínica.

Creemos que las diferencias observadas en cuanto a los valores de AUC obtenidos en las validaciones externas realizadas con los datos de nuestra OSI respecto a sus correspondientes AUC de derivación son debidas, en gran medida, al reducido tamaño muestral con el que hemos llevado a cabo las validaciones externas y a las diferencias en las prevalencias, aunque puede que los modelos estuviesen sobreajustados y no hayan sido corregidos.

Es habitual que la capacidad predictiva de un modelo sea inferior en la validación externa que la que se obtuvo en la etapa de derivación. Algunas de las razones son: diferencias en la definición de los predictores o de los outcome, diferencias en los pacientes seleccionados y diferencias en la prevalencia o incidencia del outcome, aunque una razón frecuente es el optimismo (overfitting) de los modelos⁽²⁰²⁾. El sobreajuste suele ser mayor cuando el tamaño muestral es pequeño, si disponemos de un gran número de predictores en relación al número de eventos que se hayan producido o si se han aplicado técnicas de selección de predictores (stepwise).

Una forma de reducir el sobreajuste es llevando a cabo una validación interna dentro de la etapa de derivación del modelo, Por ejemplo, empleando las técnicas bootstrapping o cross-validation. Sólo se podrá prescindir de realizar este tipo de técnicas cuando el tamaño muestral y el número de eventos por variable sea extremadamente grande pero no suele ser habitual.

Varias RS en diferentes ámbitos de la medicina han mostrado que apenas un 10% de los modelos desarrollados son validados externamente^(203,204). Y a pesar de que para realizar la validación interna del modelo no es necesario

un esfuerzo añadido en la recogida de más datos, tan sólo, entre el 30% y el 40% de los modelos de la literatura son validados internamente^(203,204).

En nuestro estudio, sólo ha habido 3 modelos (ANDC, PREDI-CO Y HNC-LL) que han mencionado haber realizado validación interna. En el resto de los estudios no se reporta información al respecto. Lo cual resulta muy llamativo porque es una etapa fundamental dentro del proceso de derivación del modelo de predicción. Por lo tanto, creemos que si no se reporta dicha información es porque no se ha llevado a cabo la validación interna. Y, al menos en esos modelos, consideramos que no se ha calculado el sobreajuste.

En el caso del 4C score podría estar justificada dicha ausencia debido al elevado número de eventos y al gran tamaño muestral considerado, además el modelo ha sido derivado empleando regresión de LASSO, que permite reducir el overfitting. En el caso del CSS score, se aplica LASSO para reducir la dimensionalidad y quizás, el sobreajuste (aunque no se haya medido) pero lo correcto sería que lo hubiesen estimado.

1.2.7. Test Hosmer-Lemeshow.

En los estudios analizados, la mayoría de ellos, proporcionan el gráfico de calibración salvo: PRE, HNC-LL y NEWS2. En el caso del HNC-LL sólo se reporta el test de Hosmer-Lemeshow, pero en los otros dos (PRE y NEWS2) no hay mención alguna a la calibración.

Parece haber una cierta tendencia a dar más importancia a los resultados obtenidos en la discriminación del modelo, pero a no considerar tan relevantes los resultados obtenidos en la calibración. Resulta incomprensible ya que, para evaluar las predicciones, lo adecuado es valorar la calibración y la discriminación.

La medida habitual empleada para valorar la calibración es el test de Hosmer-Lemeshow, a pesar de que es sabido que el test presenta serias limitaciones: tiene poder limitado para evaluar una mala calibración, es sensible al nº de grupos y también al tamaño de la muestra. Por lo tanto, lo más correcto es presentar el test con el gráfico de calibración (o la pendiente de calibración).

1.2.8. Fórmula completa de regresión o sistema de puntuación.

A pesar de que existen herramientas para que los investigadores puedan reportar de forma estandarizada los estudios de modelos predictivos, como es el caso de la herramienta TRIPOD (transparent reporting of multivariable prediction model for individual prognosis or diagnosis), ha habido estudios de los que fueron identificados mediante la RS, que no se han podido reproducir porque no se proporcionaba el modelo completo de la RPC (no proporcionaban la constante) o directamente no informaban el modelo. Existe una cierta costumbre de informar los OR de los predictores en lugar de proporcionar los coeficientes y, en esos casos, no se suele indicar el valor de la constante.

Hay autores que prefieren simplificar los cálculos y aplican un sistema de puntuación, proporcionando un valor para cada coeficiente (a partir de los coeficientes obtenidos en el modelo multivariante) y estos pesos se traducen en puntuaciones de una escala, de manera que se pueda obtener una puntuación total para el score.

En algunos estudios, por ej: PREDI-CO, se han cometido errores a la hora de asignar dichos pesos. El modo correcto de proceder es: una vez obtenidos los coeficientes del modelo multivariante, se debe coger como referencia el coeficiente de menor valor y dividir el coeficiente de cada predictor respecto a ese. Una vez obtenido, seleccionamos como puntuación final para cada predictor, la aproximación con 1 decimal.

Esto se suele hacer porque los clínicos prefieren trabajar con el sistema de puntuación a trabajar con la fórmula de regresión, ya que les resulta más sencillo porque es más manejable. El problema que hemos detectado es, que hay casos en los que el paso de la fórmula de regresión al sistema de puntuación se hace de manera incorrecta, por lo que se invalida el modelo derivado. Además, el sistema de puntuación suele tener menor capacidad predictiva que la obtenida para el modelo completo de regresión.

También hay que mencionar que, en ocasiones, el autor no reporta el modelo cuando proporciona una calculadora web o cuando se añade el nomograma en el estudio.

En algunos de los estudios evaluados en esta tesis, los autores han creado herramientas fiables que a su vez fuesen fáciles de manejar. Se han proporcionado calculadoras web (4C score y NEWS2) o nomogramas (ANDC y Zhang) que permitieran obtener el riesgo del outcome analizado (ej: fallecimiento o ingreso en UCI) de manera sencilla y rápida, pero al mismo tiempo han proporcionado la información necesaria para poder reproducir el modelo.

1.2.9. Aplicabilidad.

Otra cuestión relevante que se debe considerar es la aplicabilidad del modelo derivado. La herramienta PROBAST no sólo permite evaluar los riesgos de sesgos, también permite considerar las inquietudes respecto a la aplicabilidad de un estudio primario.

En los estudios evaluados en nuestra RS, no hemos encontrado problemas de aplicabilidad. Los pacientes, predictores y outcomes considerados en ellos han obtenido una buena calificación en dichos dominios.

Los problemas de aplicabilidad pueden surgir cuando la población, los predictores o los resultados del estudio difieren de los especificados en la pregunta de revisión. Por ejemplo, cuando los participantes en el estudio del modelo de predicción provienen de un entorno médico diferente al de la población objetivo definida en la pregunta de revisión (un modelo de predicción desarrollado en la atención secundaria puede tener una discriminación y calibración diferentes en la atención primaria porque los pacientes en entornos hospitalarios suelen tener una enfermedad más grave que los de atención primaria)^(64,205).

1.3. UTILIDAD CLÍNICA

Es fundamental que un modelo sea reportado de manera correcta y que esté libre de sesgos, pero no es suficiente para que dicho modelo se pueda aplicar en la práctica clínica. Reilly y Evans⁽¹⁶⁾ proponen que una RPC debe validarse externamente con éxito de manera consistente en múltiples contextos o poblaciones (validación "amplia"), antes de que los médicos puedan usar sus predicciones con confianza en pacientes futuros.

Desconocemos cuántos de los modelos que se han derivado para predecir mortalidad en pacientes COVID-19 han sido aplicados en la práctica habitual durante la pandemia. En general, pocos modelos son usados en la práctica clínica porque no han sido validados externamente.

La validación externa temporal, puede ser considerada como una validación intermedia entre la validación interna y la validación geográfica o de dominio⁽²⁵⁾. Diferencias entre la población de derivación y la de validación son generalmente pequeñas en estudios de validación temporal y más grandes en estudios de validación geográfica.

Si obtenemos buen rendimiento de una RPC en una validación temporal, indica que la RPC puede ser generalizada a nuevos pacientes, pero si obtenemos buen rendimiento de una RPC en una validación geográfica, se puede considerar una fuerte evidencia de generalización de dicha RPC⁽⁸⁰⁾.

Nosotros hemos realizado validaciones externas geográficas con los datos de pacientes de la 1ª ola de nuestra OSI y, de todos los scores validados, el que mejor capacidad predictiva proporcionó fue el ANDC score. Se llevó a cabo una búsqueda bibliográfica en diciembre 2020 para identificar estudios en los que se hubiesen realizado validaciones externas de dicho score, pero no se encontró ningún estudio. Por lo tanto, decidimos realizar una segunda validación externa con los pacientes de la segunda ola de nuestra OSI. El objetivo era comprobar si los resultados obtenidos en la 1ª validación se mantenían estables en el tiempo y se comprobó que el rendimiento del score en la 2ª validación era incluso un poco mejor que el obtenido en la primera (al igual que les sucedía a los autores en su estudio).

Los autores habían reportado muy buenos resultados de dicho score, tanto en la derivación (AUC=0.92; IC95%: 0.83-0.97) como en la validación externa realizada (AUC=0.97; IC95%: 0.94-1). Nosotros obtuvimos valores de AUC inferiores a los reportados por los autores, tal y como se esperaba, ya que es sabido que es habitual obtener un rendimiento menor en la validación externa, más aún cuando la validación externa es geográfica.

El score ANDC para predecir mortalidad en pacientes COVID-19 fue derivado en un único centro en Wuhan (China) y con un tamaño muestral insuficiente, pero ha demostrado buen rendimiento predictivo (tanto en la primera como en la segunda ola de la pandemia en pacientes del HUD) y está formado por un número reducido de variables, tan sólo 4. Dichas variables son objetivas y fáciles de obtener en la práctica clínica, por lo que creemos que es un score que se debería tener en cuenta, aunque sería deseable que se realizasen más validaciones externas geográficas.

Un modelo puede no mostrar una buena capacidad predictiva en la práctica porque existan deficiencias en los métodos de derivación o porque el conjunto de datos empleados en la validación sea muy diferente del que se empleó en la derivación. Consideramos que no hay deficiencias graves en la derivación del score ANDC que invaliden los resultados obtenidos y que los pacientes que se habían seleccionado para la derivación eran similares a los pacientes de las olas seleccionadas en el estudio.

Para poder implementar una RPC en la práctica habitual de un entorno hospitalario, es necesario realizar un estudio de utilidad clínica, que como comentamos al comienzo de esta tesis, se debe llevar a cabo mediante un ECA. El problema es que, dadas las circunstancias y debido a la situación que se ha vivido en los hospitales de todo el mundo, resultaba muy complicado poner en marcha este tipo de estudio en nuestro hospital en ese momento. De hecho, en ninguno de los scores que hemos validado se ha realizado un estudio de utilidad clínica.

La unidad de urgencias consideró que este score podría serles de utilidad para realizar el triaje de pacientes ya que el score les proporcionaba una información que a priori no tenían, es decir, les daba una clasificación del riesgo de morir del paciente al ingreso (riesgo bajo, moderado o alto). Por lo que, a la hora de decidir si el paciente era o no ingresado al acudir a urgencias, consideraban el score a la vez que otra serie de variables.

Una vez que se implementó el score dentro del algoritmo de triaje de urgencias, llevamos a cabo un análisis del rendimiento de dicho score. Los resultados obtenidos indican que el score ANDC, unido a los demás aspectos considerados en el algoritmo, ha ayudado en el triaje de los pacientes COVID-19 que han acudido a las urgencias del HUD durante el período de estudio.

En nuestros análisis ha quedado demostrado que no ha habido perjuicio para los pacientes, ya que los pacientes que fueron dados de alta y retornaron no lo hicieron porque se hubiese cometido el error de enviarlos de alta a su domicilio de manera equivocada, sino porque el momento en el que acudieron a urgencias (la primera vez) se encontraban en una fase temprana de la enfermedad y retornaron cuando la enfermedad había evolucionado.

Para que un modelo se considere útil⁽²⁵⁾, la puntuación de riesgo debe ser clínicamente creíble, precisa (bien calibrada con buena capacidad discriminativa), ser generalizable (estar validada externamente) y demostrar que es clínicamente eficaz, es decir, que proporciona información adicional útil para que los clínicos mejoren la toma de decisiones terapéuticas y, por lo tanto, el resultado del paciente^(16,210).

Cabe destacar que la consideración del score ANDC como una característica más del algoritmo de triaje, no supone más trabajo para el clínico ya que el valor del score ANDC es proporcionado automáticamente como un parámetro más del informe del laboratorio y el clínico sólo debe revisar dicho valor, así como la clasificación de riesgo otorgada.

El tipo de paciente COVID-19 que no ingresaron porque se empleó en el triaje el algoritmo modificado pero que habrían sido ingresados si se hubiese empleado el algoritmo establecido por el Ministerio de Sanidad, tenían una estancia promedio de 2.5 días en las dos primeras olas. Si consideramos el coste por día de estancia (aprox. 700 euros/día según datos de los costes publicados por el Ministerio de Sanidad en el informe de costes de hospitalización en el Sistema Nacional de Salud), podemos decir que la implementación del algoritmo modificado ha permitido ahorrar aproximadamente 128.000 euros.

Y, más importante aún, el nuevo algoritmo ha permitido realizar un uso más eficiente de los recursos sanitarios, ha evitado los ingresos innecesarios (mejorando la calidad del sistema) y podría servir para disminuir la sobrecarga de los hospitales y descongestionar las urgencias en momentos de mayor carga asistencial ya que, en el momento en el que se monitorizó el rendimiento del algoritmo modificado era un período de poca incidencia de la enfermedad y aún así se obtuvo una reducción de ingresos de casi el 13%. Consideramos que es un dato muy relevante y que podría ser superior si se aplica dicho algoritmo en momentos de incidencias superiores.

Por esto, consideramos que el ANDC score puede ayudar a los clínicos a estratificar a los pacientes según su riesgo de morir y permitirá a su vez mejorar el manejo clínico de los pacientes COVID-19 y el cuidado de los mismos.

2. PRUEBAS DIAGNÓSTICAS DE LABORATORIO

Tras analizar las pruebas diagnósticas en el entorno hospitalario, hemos comprobado que ninguna es fiable al 100% y que la que presenta mayor rendimiento diagnóstico es la RT-PCR.

2.1. PRUEBA RT-PCR.

La prueba RT-PCR ha sido y sigue siendo de gran utilidad en esta pandemia. Se ha empleado desde el comienzo como prueba diagnóstica para detectar la presencia del material génico específico del virus SARS-CoV-2 porque es la que presenta mayor sensibilidad y mayor especificidad. Los datos publicados sobre el rendimiento de la RT-PCR con exudado nasofaríngeo en el laboratorio mostraban una sensibilidad menor (entorno al 80%) que la obtenida en el entorno hospitalario ($Se=88.8\%$ con un $IC_{95\%}: 85.9-91.8$). Llama la atención que, en septiembre de 2020 y siendo la prueba de referencia para el diagnóstico de la enfermedad, no se hubiesen llevado a cabo estudios, en los hospitales de España, para poder conocer la sensibilidad de dicha prueba en los entornos hospitalarios y que fuésemos el único centro que proporcionó dicha información.

Nuestros análisis nos han permitido dar respuesta a las necesidades que iban surgiendo en nuestro hospital y han facilitado la toma de decisiones en algunos casos, por ej: decidir dónde ubicar a los pacientes que no tenían un diagnóstico COVID y, por lo tanto, debían permanecer en las zonas denominadas "grises" hasta tener claro si presentaban o no la enfermedad. Para arrojar luz sobre esta cuestión, se analizó qué porcentaje de los pacientes que ingresaban y tenían una RT-PCR negativa (realizada en las 48h tras el ingreso) se positivizaban en la 1ª semana (10.3%). Se consideró que este porcentaje era demasiado elevado y que no nos permitía ubicar a pacientes que aparentemente eran sanos en zonas libres de enfermedad (zonas limpias). Por este motivo, estudiamos cuántos de los pacientes que tenían dos RT-PCR negativas consecutivas tras el ingreso (con menos de 48h entre ambas) se positivizaban a la semana. Y dicho porcentaje disminuyó hasta el 2.1%. Estos resultados permitieron trasladar a los pacientes que tenían 2 RT-PCR negativas consecutivas tras el ingreso (con menos de 48h entre ambas) a una zona limpia sin asumir un riesgo demasiado elevado de generar un brote en zonas libres de COVID del hospital, y a su vez, permitió descongestionar las zonas de grises con una cierta premura.

2.2. PRUEBA RÁPIDA DE ANTÍGENOS PANBIO™ (AG RAPID TEST DEVICE).

Para el estudio del rendimiento diagnóstico del test rápido de antígenos, nos hemos basado únicamente en el *Ag Rapid Test Device (Panbio™)* porque era el test que el Ministerio había acordado emplear en España y, por lo tanto, era el que se iba a utilizar en nuestro hospital. El objetivo era determinar en qué contextos se podía aplicar esta prueba y se demostró que no era válida para realizar screening poblacionales ni para su uso habitual en la población general como medida de detección de pacientes contagiados por la enfermedad.

A pesar de que se ha autorizado la venta en farmacias de este test, consideramos que no tiene sentido emplear este tipo de prueba en pacientes asintomáticos ya que se ha demostrado su escaso rendimiento en estos pacientes. Sin embargo, sí se puede emplear para detectar casos positivos en zonas con una elevada prevalencia de contagios y/o en pacientes sintomáticos, aunque si se obtiene un resultado negativo en el test de antígenos y el paciente presenta síntomas, se debe realizar una RT-PCR para confirmar la enfermedad.

Cabe mencionar que los datos obtenidos con el estudio del ISCIII y los datos que proporcionó la empresa ABBOTT sobre la capacidad diagnóstica de la prueba rápida de antígenos Panbio™, están basados en pruebas que se han llevado a cabo en el laboratorio y con muestras seleccionadas, pero la prueba *Ag Rapid Test Device (Panbio™)* se ha desarrollado para que sea aplicada en el punto de atención médica (*point of care*). El rendimiento que reportan tanto la empresa ABBOTT ($Se=93.3\%$ e $IC95\%: 83.3-98.2$) como el ISCIII ($Se=90.5\%$; $IC95\%: 86-93.6$) es muy elevado y, al mismo tiempo, diferente al que se obtuvo mediante MA combinando los estudios obtenidos en la RS ($Se=0.83$; $IC95\%: 0.80-0.86$). Creemos que es debido a que se seleccionaron muestras de laboratorio con una alta prevalencia.

En el estudio de Alemany-Mitja, la sensibilidad reportada para los pacientes sintomáticos es del 93% y llama la atención que la prevalencia de la enfermedad en esa población sea del 94%, es decir, casi el 100% de los pacientes sintomáticos seleccionados están enfermo. Por otro lado, Billaud proporciona una sensibilidad muy similar tanto en sus pacientes sintomáticos como en los asintomáticos. Parece extraño que obtenga una $Se=54\%$ para pacientes sintomáticos teniendo una prevalencia del 50.3% en dicho grupo. Drevinek y Masía también ofrecen sensibilidades bajas, entorno al 67%

aunque las prevalencias son del 58% y del 25% respectivamente para el grupo de sintomáticos.

Por estos motivos, decidimos realizar el MA sin considerar los resultados de sensibilidades proporcionados por esos 4 estudios (Alemay, Billaud, Drevinek y Masiá), de ese modo, conseguimos reducir la heterogeneidad, de un $I^2=91\%$ a un $I^2=62.8\%$, aunque sigue siendo un valor elevado y hace que los resultados obtenidos en el análisis sean menos fiables. La heterogeneidad obtenida al combinar los estudios es fruto sin duda de las diferentes formas de seleccionar las muestras, del lugar de la toma de las muestras (AP u Hospitalaria) y/o de las diferentes prevalencias.

El único estudio que analiza el efecto de la duración de los síntomas en la sensibilidad, el de Masiá, indica que la sensibilidad observada en la prueba, se mantiene hasta el 7º día, a partir del cual la sensibilidad disminuye notablemente, hasta un 40%. Es por esto que, se recomienda emplear la prueba Ag rapid test device Panbio™ en pacientes sintomáticos y con menos de 7 días tras el comienzo de los síntomas, aunque este período parece ser menor tras la publicación de un reciente estudio⁽²¹¹⁾, publicado en diciembre de 2021, en el que se indica que la sensibilidad del test baja hasta un 32.5% a partir del cuarto día de los síntomas. Por lo que se recomienda emplear el test en pacientes con síntomas constantes y con menos de 4 días desde el comienzo de los síntomas. En dicho estudio, se indica que el test tiene un bajo rendimiento en el entorno hospitalario y que la Se es del 45.8% para detectar casos positivos cuando se emplea en pacientes con ingresos no planificados.

En pacientes asintomáticos, hay dos estudios que llaman poderosamente la atención y hacen que se pongan en duda los resultados obtenidos. Por un lado, el estudio de Alemany-Mitja, en el que la prevalencia de los contactos es del 87.7%. Estos valores de prevalencias no son los que hemos observado en la vida cotidiana, es decir, el porcentaje de pacientes que se ha contagiado por contacto en la población general ha sido de entre un 10-20%, muy lejos del valor proporcionado por dichos autores. En el estudio, no se indica si los contactos son sintomáticos o no y se ofrece una sensibilidad del 94.2%, mayor incluso que la obtenida para pacientes sintomáticos. Por dicho motivo, hemos seleccionado el grupo que ellos llaman *screening* para la combinación de estudios en pacientes asintomáticos, ya que las prevalencias eran más similares a las del resto de estudios (aunque seguían

siendo más elevadas que en el resto de estudios considerados). Por otro lado, el segundo estudio que nos hace cuestionarnos los resultados obtenidos en el MA de pacientes asintomáticos es el estudio de Merino-Amador, se obtiene una sensibilidad de la prueba del 100% en pacientes asintomáticos, mayor incluso que la que reportan para pacientes sintomáticos. Lo cual no parece muy creíble. Es por esto, que se excluyeron esos dos estudios para realizar la combinación mediante MA para pacientes asintomáticos.

Tras eliminar esos 2 estudios del análisis, la heterogeneidad (I^2) pasa de un 91% a un 0%. Por lo que parece que no hay porqué dudar de los resultados obtenidos y se concluye que la prueba rápida de antígenos no es válida para detectar la enfermedad del SARS-CoV-2 en pacientes asintomáticos.

2.3. PRUEBA DE DETERMINACIÓN DE ANTICUERPOS (SEROLOGÍA).

Tras la primera ola y el fin del confinamiento (mediados de mayo). Se decidió realizar en el HUD un estudio de serología, para determinar cuántas personas habían pasado la enfermedad y poder conocer así la prevalencia en la población de Gipuzkoa. El análisis arrojó unos resultados llamativos ya que, tan sólo un 2.8% de la población había desarrollado anticuerpos. La prevalencia promedio en España, para esas fechas, fue de aproximadamente un 5.6%.

En la prueba de determinación de anticuerpos (Serología) relativa a la RT-PCR, cuando se considera que la serología se ha realizado posterior a la RT-PCR, con un mínimo de 15 días entre ambas pruebas, se obtuvo una sensibilidad del 63% con un (IC95%: 58.8-66.9) en nuestro entorno hospitalario.

No es una prueba muy sensible, pero esto es debido a que depende del momento en el que se realice. Podemos obtener un resultado serológico negativo en los primeros días de infección, ya que los anticuerpos comienzan a estar presentes, aproximadamente, a los 15 días. Por este motivo, se estableció este período de tiempo para seleccionar las pruebas serológicas que se iban a tener en cuenta para realizar el análisis de datos (Tabla 52).

3. SCORES PROPIOS CREADOS EN EL HUD

3.1. SCORE DE SOSPECHA CLÍNICA (SSC).

El SSC creado en el HUD permitió, en un momento muy complicado de la pandemia, guiar el juicio clínico. De modo que, se decidía qué pacientes estaban infectados por la enfermedad (a los que asignaron una alta o muy alta sospecha de estar infectados) y quiénes no (a quiénes asignaron una sospecha media o baja de estar infectados).

Cuando se analizó el rendimiento de este score, se comprobó que la sensibilidad era del 80% con un (IC95%: 75%-85%). Lo que indica que el juicio clínico era bastante acertado.

Se estudió si este score podía ayudar a identificar a los pacientes que se iban a positivizar. Y se obtuvo una Se=84% con un (IC95%: 69%-94%), lo que indica que podíamos estar bastante confiados con la sospecha inicial asignada por los clínicos.

Se analizaron otras características que pudiesen ayudar a mejorar la discriminación de aquellos pacientes que se iban a positivizar en los próximos 7 días, tras la primera RT-PCRT negativa en las 48h tras el ingreso. Y tanto la edad como los síntomas (fiebre, tos y/o disnea) no arrojaron buenos resultados, es decir, no ayudaban a mejorar la discriminación ya analizada.

Estos resultados van en consonancia con lo que habíamos comentado al principio. Los síntomas de los pacientes que tenían la enfermedad y los que presentaban síntomas compatibles con ella pero que finalmente no estuvieron enfermos eran muy similares, por lo que era complicado determinar qué pacientes estaban enfermos y cuáles no basándonos en los síntomas.

La edad parece un factor determinante para el pronóstico y para el desenlace (al menos en las olas analizadas), pero no para predecir la infección.

3.2. RISK SCORE.

Los clínicos del HUD junto con las unidades de Epidemiología Clínica del HUD y del Hospital Ramón y Cajal, crearon hace varios años un score de alerta temprana para calcular el riesgo que tienen los pacientes de presentar eventos graves, ingresar en UCI o fallecer, en las 24h tras el ingreso. Este score ha ayudado durante este tiempo a priorizar la asistencia a los pacientes con mayor riesgo, por dicho motivo, los clínicos se plantearon que podía ayudar también a los pacientes COVID-19.

Al validar el Risk Score con nuestros pacientes, vimos que el rendimiento predictivo era muy bajo por lo que se descartó emplear dicho score en pacientes COVID-19.

Consideramos que los resultados obtenidos son debido a que el score está formado por variables (frecuencia respiratoria, frecuencia cardíaca, nivel de conciencia y PAS) que no han sido incluidas en los modelos finales o no han demostrado tener mayor peso en los modelos derivados para predecir mortalidad o ingreso en UCI.

Tan sólo las variables frecuencia respiratoria y nivel de conciencia aparecen en algunos de los scores validados con los pacientes de nuestra OSI. El score 4C incluye esos dos predictores y el score PREDI-CO incluye la frecuencia respiratoria pero no son las variables que más peso representaban en ambos modelos.

4. LIMITACIONES

- Una de las limitaciones de este estudio es que, al ser una enfermedad nueva y tan desconocida, puede que haya factores relevantes que aún no hayan sido descritos, aunque parece improbable dada la cantidad de estudios que se han llevado a cabo y que se siguen realizando.
- Por otro lado, la recogida de los datos que hemos empleado para llevar a cabo las validaciones externas se ha realizado en una situación clínica real y esto ha provocado que haya un elevado número de datos faltantes en las bases de datos clínicas que se han consultado para realizar el estudio, ya que, al comienzo de la pandemia, no se planteó realizar un estudio concreto ni se creó un sistema de recogida de datos adicional a lo que se hace en la práctica habitual. Como consecuencia, no se han podido validar muchos de los modelos que se identificaron en las RS llevadas a cabo y puede que se hayan excluido modelos que podrían proporcionar buenos resultados si fuesen validados, en nuestra población, con un tamaño de muestra adecuado.
- El modo de trabajar y las condiciones de trabajo eran muy diferentes al comienzo de la pandemia, así como el tipo de paciente. Creemos que esto puede hacer que los resultados obtenidos sean diferentes si se aplica a nuevos pacientes en el momento actual.

5. ACCIONES FUTURAS

Tras realizar esta tesis, se han abierto nuevas vías de estudio:

- Por un lado, creemos que es importante seguir revisando las publicaciones de los scores que se vayan produciendo para identificar aquellos que mejor rendimiento predictivo hayan obtenido y poder validarlos en nuestra población de estudio. Para ello, por un lado, se han creado alertas y, por el otro, continuamos revisando las publicaciones que realiza el grupo COVID-PRECISE (por si actualizan la Living Systematic Review).
- Por otro lado, creemos que hay que hacer un mayor esfuerzo en identificar estudios que se hayan centrado únicamente en pacientes críticos para predecir mortalidad. Con los datos que teníamos disponibles, no pudimos aplicar correctamente los scores que se habían derivado para este tipo de pacientes y es por ello que hemos mantenido varias reuniones con el jefe de servicio de la UCI del HUD para trabajar en esta línea.
- Estamos trabajando con la unidad de urgencias del HUD para realizar la publicación de un artículo de divulgación científica para dar a conocer los resultados obtenidos tras la implementación del score ANDC en el algoritmo de triaje empleado en dicha unidad.

CONCLUSIONES

RESUMEN

Podemos concluir que es importante considerar la experiencia y el juicio clínico, ya que han demostrado su validez en esta pandemia. Así mismo, es crucial tener en cuenta a los científicos, epidemiólogos y bioestadísticos. La ciencia ha demostrado sobradamente su aportación en esta crisis sin precedentes y resulta inaudito que haya sectores que tomen decisiones al margen (o incluso en contra) de los conocimientos aportados por estos profesionales.

Consideramos que se debería hacer más hincapié en los centros de trabajo en la importancia que tiene la recogida rigurosa, exhaustiva y completa de las bases de datos clínicas. Por otro lado, como la recogida de los datos es muy costosa, creemos que se debería hacer de manera automatizada para facilitar dicha tarea a los clínicos.

Esta crisis mundial ha puesto de manifiesto la importancia de publicar protocolos para derivar y/o validar modelos de predicción cuando se realiza un estudio prospectivo y creemos que esta necesidad va reconociéndose^(212,213). Consideramos que es indispensable un marco teórico en el que, mediante una guía metodológica, se ayude a los investigadores a poder derivar de manera adecuada los modelos predictivos, así como a llevar a cabo correctamente las validaciones y los estudios de utilidad clínica.

Existen muchas lagunas sobre como derivar y validar de manera adecuada los modelos predictivos, y esto repercute en los resultados que se obtienen, por lo que, sería deseable acotar, en la medida de lo posible, el modo de proceder para realizar los estudios y así obtener modelos que puedan ser implementados para ayudar a los clínicos en la toma de decisiones clínicas.

Se han publicado muchísimos artículos durante la pandemia (demasiados), y el elevado número de publicaciones ha hecho que fuese muy costoso seleccionar los "buenos", es decir, separar *la paja del grano*. Consideramos que se ha hecho un gran esfuerzo por desarrollar modelos predictivos, pero muchos de ellos de baja calidad metodológica y, por lo tanto, científica. Y debido al gran riesgo de sesgos evaluados en ellos, a la falta de validaciones externas y a la escasez de estudios de utilidad clínica, la implantación de dichos modelos en la práctica clínica no es recomendable.

Por todo ello, alentamos a los investigadores a desarrollar protocolos, especialmente para estudios de derivación de modelos, e incluso a registrarlos en registros que incluyan estudios observacionales (como ClinicalTrials.gov)^(214,215), ya que este tipo de estudios se encuentran todavía en una etapa relativamente temprana en comparación con la de los estudios de intervención aleatorios y los estudios de precisión de las pruebas de diagnóstico.

Por último, sería deseable que, antes de permitir la publicación de los artículos de modelos predictivos, se establecieran una serie de requisitos exigibles por los revisores. Actualmente, la información necesaria para evaluar el sesgo y la aplicabilidad a menudo no se informa o, no al menos, de una manera completa y correcta. Esperamos que la adherencia, tanto de las revistas como de los autores, a la guía de reporte de modelos predictivos TRIPOD^(29,127) ayude a reducir este problema.

CONCLUSIONES

MODELOS PREDICTIVOS DE PRONÓSTICO

1. El 94% de los modelos predictivos de pronóstico tienen un alto riesgo de sesgos en el dominio *Análisis*.
2. Los modelos no se han derivado calculando el mínimo tamaño muestral necesario.
3. Un reducido número de estudios siguen las recomendaciones establecidas por TRIPOD para el reporte de modelos predictivos.
4. Ningún modelo ha hecho un estudio de utilidad clínica, por lo que no deberían ser implementado en la práctica clínica.
5. La capacidad predictiva del score ANDC, en la validación externa con pacientes de la 1ª ola, es AUC=0.82 (IC95%: 0.77-0.87). Y, en la validación externa con pacientes de la 2ª ola, AUC=0.84 (IC95%: 0.81-0.87).
6. El score ANDC mejora el triaje de pacientes en urgencias.

PRUEBAS DIAGNÓSTICAS DE LABORATORIO

Prueba RT-PCR.

7. La sensibilidad de la 1ª RT-PCR en el HUD es 88.8% (IC95%: 85.9%-91.8%) y la sensibilidad de la 2ª RT-PCR es 89.8% (IC95%: 77.8%-96.6%).
8. El 10.3% de los pacientes que tienen una RT-PCR negativa en las 48h tras el ingreso, se positivarán en los próximos 7 días. Y el 2.1% de los pacientes que tienen dos RT-PCR negativas consecutivas (tras el ingreso), con un máximo de 48h entre ambas, se positivarán en los próximos 7 días.

Prueba de antígenos *Ag test rapid Device (Panbio™)*.

9. Se recomienda usar la prueba rápida de antígenos en pacientes sintomáticos (sensibilidad relativa=83%; IC95%: 80%-86%, I²=62.8% y sensibilidad absoluta=74%) o en zonas de alta prevalencia de la enfermedad.
10. No se recomienda emplear la prueba rápida de antígenos en pacientes asintomáticos (sensibilidad relativa=47%; IC95%: 41%-53%, I²=0% y sensibilidad absoluta=42%) o para realizar screening poblacionales.

Prueba de detección de anticuerpos (Serología).

11. La sensibilidad de la prueba de detección de anticuerpos en nuestra OSI, relativa a la RT-PCR, cuando se considera que la Serología se ha realizado posterior a la RT-PCR con un mínimo de 15 días entre ambas pruebas es del 63% (IC95%: 58.8%-66.9%).

SCORES CREADOS EN EL HUD

Score de Sospecha Clínica (SSC).

12. La sensibilidad del SSC para discriminar entre pacientes COVID-19/NoCOVID-19 de los que presentan síntomas compatibles con la enfermedad es del 80% (IC95%: 75%-85%).
13. La sensibilidad del SSC para discriminar quiénes se van a positivar en los 7 días siguientes tras el ingreso, en pacientes que han obtenido una primera RT-PCR negativa, es del 84% (IC95%: 69%-94%).
14. La edad o los principales síntomas (Fiebre, tos o disnea) no mejoran esta última discriminación.

Risk Score.

15. El Risk score no es un score válido para predecir mortalidad o ingreso en UCI, en las 24h tras el ingreso, en pacientes COVID-19.

REFERENCIAS
BIBLIOGRÁFICAS

1. Google. Enfermedad por el nuevo coronavirus [Internet]. Mountain View: Google; c.2022 [Consulta: 23-12-2021]. Disponible en: https://www.google.com/search?q=contagios+coronavirus+espa%C3%B1a&rlz=1C1GCEU_esES819ES819&oq=contagios+coronavirus+&aqs=chrome.1.69i57j0i7.10182j0j15&sourceid=chrome&ie=UTF-8
2. Arroyo J. Coronavirus: infectados en España y la evolución del brote desde el origen. Redacción Médica [Internet]. 23 mar 2020 [Consulta:23-12-2021]; Disponible en: <https://www.redaccionmedica.com/secciones/sanidad-hoy/coronavirus-infectados-espana-y-evolucion-covid19-desde-origen-4148>
3. Pereda R, Rioja-Andueza I. La crisis del coronavirus: mapas y gráficos de su evolución en Euskadi. Eldiario.es [Internet]. 18 ene 2022 [Consulta:23-12-2021]. Disponible en: https://www.eldiario.es/euskadi/euskadi/sociedad/crisis-coronavirus-graficos-expansion-euskadi_1_1031552.html
4. Lo Benjamin WY, Fukuda H, Nishimura Y, Farrokhya F, Thabane L, Levine Mitchell AH. Systematic review of clinical prediction tools and prognostic factors in aneurysmal subarachnoid hemorrhage. *Surg Neurol Int.* 2015;6(1):135.
5. LaValley MP, Lo GH, Price LL, Driban JB, Eaton CB, McAlindon TE. Development of a clinical prediction algorithm for knee osteoarthritis structural progression in a cohort study: value of adding measurement of subchondral bone density. *Arthritis Res Ther.* 2017;19:95.
6. Moons KGM, Kengne AP, Woodward M, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart Br Card Soc.* mayo de 2012;98(9):683-90.
7. Wynants L, Collins GS, Van Calster B. Key steps and common pitfalls in developing and validating risk models. *BJOG Int J Obstet Gynaecol.* febrero de 2017;124(3):423-32.
8. Stiell IG, Wells GA. Methodologic standards for the development of clinical decision rules in emergency medicine. *Ann Emerg Med.* abril de 1999;33(4):437-47.
9. 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.
10. Hopper AD, Cross SS, Hurlstone DP, McAlindon ME, Lobo AJ, Hadjivassiliou M, et al. Pre-endoscopy serological testing for coeliac disease: evaluation of a clinical decision tool. *BMJ.* 5 de abril de 2007;334(7596):729.
11. Kheterpal S, Tremper KK, Heung M, Rosenberg AL, Englesbe M, Shanks AM, et al. Development and Validation of an Acute Kidney Injury Risk Index for Patients Undergoing General Surgery: Results from a National Data Set. *Anesthesiology.* 1 de marzo de 2009;110(3):505-15.
12. Steyerberg EW, Mushkudiani N, Perel P, Butcher I, Lu J, McHugh GS, et al. Predicting outcome after traumatic brain injury: development and international validation of prognostic scores based on admission characteristics. *PLoS Med.* 5 de agosto de 2008;5(8):e165; discussion e165.
13. Ferro JM, Bacelar-Nicolau H, Rodrigues T, Bacelar-Nicolau L, Canhão P, Crassard I, et al. Risk score to predict the outcome of patients with cerebral vein and dural sinus thrombosis. *Cerebrovasc Dis Basel Switz.* 2009;28(1):39-44.
14. Hemingway H, Croft P, Perel P, Hayden JA, Abrams K, Timmis A, et al. Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. *BMJ.* 5 de febrero de 2013;346:e5595.
15. Laupacis A, Sekar N, Stiell IG. Clinical prediction rules. A review and suggested modifications of methodological standards. *JAMA.* 12 de febrero de 1997;277(6):488-94.
16. Reilly BM, Evans AT. Translating Clinical Research into Clinical Practice: Impact of Using Prediction Rules To Make Decisions. *Ann Intern Med.* 7 de febrero de 2006;144(3):201-9.

17. Pace NL, Eberhart LHJ, Kranke PR. Quantifying prognosis with risk predictions. *Eur J Anaesthesiol.* enero de 2012;29(1):7-16.
18. 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.
19. 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.
20. Biesheuvel CJ, Vergouwe Y, Oudega R, Hoes AW, Grobbee DE, Moons KG. Advantages of the nested case-control design in diagnostic research. *BMC Med Res Methodol.* 21 de julio de 2008;8(1):48.
21. Rutjes AWS, Reitsma JB, Di Nisio M, Smidt N, van Rijn JC, Bossuyt PMM. Evidence of bias and variation in diagnostic accuracy studies. *Can Med Assoc J.* 14 de febrero de 2006;174(4):469-76.
22. Lijmer JG. Empirical Evidence of Design-Related Bias in Studies of Diagnostic Tests. *JAMA.* 15 de septiembre de 1999;282(11):1061.
23. Sanderson J, Thompson SG, White IR, Aspelund T, Pennells L. Derivation and assessment of risk prediction models using case-cohort data. *BMC Med Res Methodol.* 13 de septiembre de 2013;13:113.
24. Wasson JH, Sox HC, Neff RK, Goldman L. Clinical prediction rules. Applications and methodological standards. *N Engl J Med.* 26 de septiembre de 1985;313(13):793-9.
25. Steyerberg EW, Moons KGM, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med.* 2013;10(2):e1001381.
26. Royston P, Moons KGM, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. *BMJ.* 31 de marzo de 2009;338:b604.
27. Battle CE, Hutchings H, Evans PA. Expert opinion of the risk factors for morbidity and mortality in blunt chest wall trauma: results of a national postal questionnaire survey of Emergency Departments in the United Kingdom. *Injury.* enero de 2013;44(1):56-9.
28. de Salis I, Whiting P, Sterne JAC, Hay AD. Using qualitative research to inform development of a diagnostic algorithm for UTI in children. *Fam Pract.* junio de 2013;30(3):325-31.
29. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med.* 6 de enero de 2015;162(1):W1-73.
30. Wallace E, Smith SM, Perera-Salazar R, Vaucher P, McCowan C, Collins G, et al. Framework for the impact analysis and implementation of Clinical Prediction Rules (CPRs). *BMC Med Inform Decis Mak.* 14 de octubre de 2011;11:62.
31. Steyerberg E. *Clinical prediction models: a practical approach to development, validation, and updating.* USA: Springer; 2009. p. 459.
32. Hosmer DW, Lemeshow S, Sturdivant RX. *Applied Logistic Regression.* New Jersey: John Wiley & Sons; 2013. p. 528.
33. Amat R. Selección de predictores y mejor modelo lineal múltiple: subset selection, ridge regression, lasso regression y dimension reduction, available under a Attribution 4.0 International (CC BY 4.0), 2016. Disponible: https://www.cienciadedatos.net/documentos/31_seleccion_de_predictores_subset_selection_ridge_lasso_dimension_reduction
34. Máxima Formación [Internet]. Granada: Máxima Formación; c2022. Selección paso a paso e importancia de los predictores; 17-10-2021 [Consulta: 23-12-2021]; [3 págs.]. Disponible: <https://www.maximaformacion.es/blog-dat/seleccion-paso-a-paso-e-importancia-de-los-predictores/>

35. Little RJA. *Statistical Analysis with Missing Data*. 2da ed. New Jersey: Published by John Wiley & Sons, Inc.; 2002. 371.
36. Sauerbrei W, Royston P, Binder H. Selection of important variables and determination of functional form for continuous predictors in multivariable model building. *Stat Med*. 2007 Dec 30;26(30):5512-28.
37. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med*. 15 de enero de 2006;25(1):127-41.
38. Collins GS, Ogundimu EO, Cook JA, Manach YL, Altman DG. Quantifying the impact of different approaches for handling continuous predictors on the performance of a prognostic model. *Stat Med*. 15 de octubre de 2016;35(23):4124-35.
39. Steyerberg EW, Uno H, Ioannidis JPA, van Calster B, Collaborators. Poor performance of clinical prediction models: the harm of commonly applied methods. *J Clin Epidemiol*. junio de 2018;98:133-43.
40. Van Smeden M, de Groot JAH, Moons KGM, Collins GS, Altman DG, Eijkemans MJC, et al. No rationale for 1 variable per 10 events criterion for binary logistic regression analysis. *BMC Med Res Methodol*. 24 de noviembre de 2016;16(1):163.
41. Van Smeden M, Moons KG, de Groot JA, Collins GS, Altman DG, Eijkemans MJ, et al. Sample size for binary logistic prediction models: Beyond events per variable criteria. *Stat Methods Med Res*. agosto de 2019;28(8):2455-74.
42. Riley RD, Snell KI, Ensor J, Burke DL, Harrell FE, Moons KG, et al. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. *Stat Med*. 30 de marzo de 2019;38(7):1276-96.
43. Riley RD, Snell KIE, Ensor J, Burke DL, Harrell FE, Moons KGM, et al. Minimum sample size for developing a multivariable prediction model: Part I - Continuous outcomes. *Stat Med*. 30 de marzo de 2019;38(7):1262-75.
44. Courvoisier DS, Combescure C, Agoritsas T, Gayet-Ageron A, Perneger TV. Performance of logistic regression modeling: beyond the number of events per variable, the role of data structure. *J Clin Epidemiol*. septiembre de 2011;64(9):993-1000.
45. Riley RD, Ensor J, Snell KIE, Harrell FE, Martin GP, Reitsma JB, et al. Calculating the sample size required for developing a clinical prediction model. *BMJ*. 18 de marzo de 2020;368:m441.
46. Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. *J Clin Epidemiol*. diciembre de 1995;48(12):1503-10.
47. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol*. diciembre de 1996;49(12):1373-9.
48. Concato J, Peduzzi P, Holford TR, Feinstein AR. Importance of events per independent variable in proportional hazards analysis. I. Background, goals, and general strategy. *J Clin Epidemiol*. diciembre de 1995;48(12):1495-501.
49. Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. *Am J Epidemiol*. 15 de marzo de 2007;165(6):710-8.
50. Moons KGM, Grobbee DE. When should we remain blind and when should our eyes remain open in diagnostic studies? *J Clin Epidemiol*. julio de 2002;55(7):633-6.
51. Van Doorn S, Brakenhoff TB, Moons KGM, Rutten FH, Hoes AW, Groenwold RHH, et al. The effects of misclassification in routine healthcare databases on the accuracy of prognostic prediction models: a case study of the CHA2DS2-VASc score in atrial fibrillation. *Diagn Progn Res*. 2017;1:18.

52. Wang L, Shaw Pa, Mathelier Hm, Kimmel Se, French B. Evaluating risk-prediction models using data from electronic health records. *Ann Appl Stat.* marzo de 2016;10(1):286-304.
53. Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. *Stat Med.* 30 de enero de 2016;35(2):214-26.
54. Vergouwe Y, Steyerberg EW, Eijkemans MJC, Habbema JDF. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. *J Clin Epidemiol.* mayo de 2005;58(5):475-83.
55. 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.* junio de 2016;74:167-76.
56. Snell KI, Archer L, Ensor J, Bonnett LJ, Debray TP, Phillips B, et al. External validation of clinical prediction models: simulation-based sample size calculations were more reliable than rules-of-thumb. *J Clin Epidemiol.* julio de 2021; 135: 79-89.
57. Sarasin FP, Reymond JM, Griffith JL, Beshansky JR, Schifferli JA, Unger PF, et al. Impact of the acute cardiac ischemia time-insensitive predictive instrument (ACI-TIPI) on the speed of triage decision making for emergency department patients presenting with chest pain: a controlled clinical trial. *J Gen Intern Med.* 1994;9(4):187-194.
58. Stiell IG, McDowell I, Nair RC, Aeta H, Greenberg G, McKnight RD, et al. Use of radiography in acute ankle injuries: physicians' attitudes and practice. *CMAJ Can Med Assoc J.* 1 de diciembre de 1992;147(11):1671-8.
59. Stiell IG, McKnight RD, Greenberg GH, McDowell I, Nair RC, Wells GA, et al. Implementation of the Ottawa ankle rules. *JAMA.* 16 de marzo de 1994;271(11):827-32.
60. 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.
61. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Ann Intern Med.* 16 de marzo de 1999;130(6):515-24.
62. McGinn TG, Guyatt GH, Wyer PC, Naylor CD, Stiell IG, Richardson WS. Users' guides to the medical literature: XXII: how to use articles about clinical decision rules. Evidence-Based Medicine Working Group. *JAMA.* 2000 Jul 5;284(1):79-84. doi: 10.1001/jama.284.1.79. PMID: 10872017.
63. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. *BMJ.* 2009 May 28;338:b605. doi: 10.1136/bmj.b605. PMID: 19477892.
64. Moons KGM, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ.* 4 de junio de 2009;338:b606.
65. 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.
66. Ochoa Sangrador C. Diseño y evaluación de Reglas de Predicción Clínica. *Evid Pediatr.* 2009;5:28.
67. Harrell FE. S-Plus Software. En: Harrell FE, editor. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis.* New York, NY: Springer; 2001. p. 105-20.
68. Harrell FE. Resampling, Validating, Describing, and Simplifying the Model. En: Harrell FE, editor. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis.* New York, NY: Springer; 2001. p. 87-103.
69. Efron B. Bootstrap Methods: Another Look at the Jackknife. En: Kotz S, Johnson NL, editores. *Breakthroughs in Statistics: Methodology and Distribution.* New York, NY: Springer; 1992. p. 569-93.

70. Efron B, Tibshirani RJ. An Introduction to the Bootstrap [Internet]. Boston, MA: Springer US; 1993. p. 456.
71. Félix BMF. Validación interna de modelos predictivos de regresión logística. Comando validation (Stata). :54.
72. Brunelli A. A synopsis of resampling techniques. *J Thorac Dis.* diciembre de 2014;6(12):1879-82.
73. Steyerberg EW, Harrell FE, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol.* agosto de 2001;54(8):774-81.
74. Steyerberg EW. Validation in prediction research: the waste by data splitting. *J Clin Epidemiol.* 1 de noviembre de 2018;103:131-3.
75. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Ann Intern Med.* 1999;130(6):515–524. doi: 10.7326/0003-4819-130-6-199903160-00016
76. Isaza Jaramillo S, Jaimes Barragán F. Ronda clínica y epidemiológica: aproximación a los modelos de predicción clínica. *Iatreia.* 2017;92-9.
77. Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: validating a prognostic model. *BMJ.* 28 de mayo de 2009;338:b605.
78. Toll DB, Janssen KJM, Vergouwe Y, Moons KGM. Validation, updating and impact of clinical prediction rules: a review. *J Clin Epidemiol.* noviembre de 2008;61(11):1085-94.
79. Janssen KJM, Moons KGM, Kalkman CJ, Grobbee DE, Vergouwe Y. Updating methods improved the performance of a clinical prediction model in new patients. *J Clin Epidemiol.* enero de 2008;61(1):76-86.
80. Cowley LE. Methodological standards for the development and evaluation of clinical prediction rules: a review of the literature. 2019;23.
81. Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med.* 29 de febrero de 2000;19(4):453-73.
82. Ferrero F. [Clinical prediction rules]. *Arch Argent Pediatría.* 1 de febrero de 2010;108:6-7.
83. Steyerberg EW, Van Calster B, Pencina MJ. Medidas del rendimiento de modelos de predicción y marcadores pronósticos: evaluación de las predicciones y clasificaciones. *Rev Esp Cardiol.* septiembre de 2011;64(9):788-94.
84. Gerds TA, Cai T, Schumacher M. The performance of risk prediction models. *Biom J Biom Z.* agosto de 2008;50(4):457-79.
85. Brier GW. Verification of forecasts expressed in terms of probability. *Mon Weather Rev.* 1 de enero de 1950;78(1):1-3.
86. Pepe MS. *The Statistical Evaluation of Medical Tests for Classification and Prediction*: 31. Oxford New York; 2004. 320 p.
87. Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. *Ann Intern Med.* 1 de enero de 2019;170(1):W1-33.
88. Copas JB. Regression, Prediction and Shrinkage. *J R Stat Soc Ser B Methodol.* 1983;45(3):311-54.
89. Van Houwelingen JC, Le Cessie S. Predictive value of statistical models. *Stat Med.* noviembre de 1990;9(11):1303-25.
90. Austin PC, Steyerberg EW. Graphical assessment of internal and external calibration of logistic regression models by using loess smoothers. *Stat Med.* 10 de febrero de 2014;33(3):517-35.
91. Parikh R, Mathai A, Parikh S, Chandra Sekhar G, Thomas R. Understanding and using sensitivity, specificity and predictive values. *Indian J Ophthalmol.* febrero de 2008;56(1):45-50.

92. Carter JV, Pan J, Rai SN, Galandiuk S. ROC-ing along: Evaluation and interpretation of receiver operating characteristic curves. *Surgery*. junio de 2016;159(6):1638-45.
93. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. abril de 1982;143(1):29-36.
94. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation*. 20 de febrero de 2007;115(7):928-35.
95. Cook NR. Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve. *Clin Chem*. enero de 2008;54(1):17-23.
96. Pencina MJ, D'Agostino RB, D'Agostino RB, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med*. 30 de enero de 2008;27(2):157-72; discussion 207-212.
97. Leening MJG, Steyerberg EW, Van Calster B, D'Agostino Sr. RB, Pencina MJ. Net reclassification improvement and integrated discrimination improvement require calibrated models: relevance from a marker and model perspective. *Stat Med*. 2014;33(19):3415-8.
98. Leening MJG, Vedder MM, Wittelman JCM, Pencina MJ, Steyerberg EW. Net reclassification improvement: computation, interpretation, and controversies: a literature review and clinician's guide. *Ann Intern Med*. 21 de enero de 2014;160(2):122-31.
99. Pepe MS, Fan J, Feng Z, Gerds T, Hilden J. The Net Reclassification Index (NRI): a Misleading Measure of Prediction Improvement Even with Independent Test Data Sets. *Stat Biosci*. 1 de octubre de 2015;7(2):282-95.
100. Burch PM, Glaab WE, Holder DJ, Phillips JA, Sauer J-M, Walker EG. Net Reclassification Index and Integrated Discrimination Index Are Not Appropriate for Testing Whether a Biomarker Improves Predictive Performance. *Toxicol Sci Off J Soc Toxicol*. 1 de marzo de 2017;156(1):11-3.
101. Hilden J, Gerds TA. A note on the evaluation of novel biomarkers: do not rely on integrated discrimination improvement and net reclassification index. *Stat Med*. 30 de agosto de 2014;33(19):3405-14.
102. Antolini L, Tassistro E, Valsecchi MG, Bernasconi DP. Graphical representations and summary indicators to assess the performance of risk predictors. *Biom J Biom Z*. noviembre de 2019;61(6):1417-29.
103. Peirce CS. The numerical measure of the success of predictions. *Science*. 1884 Nov 14;4(93):453-4. doi: 10.1126/science.ns-4.93.453-a. PMID: 17795531.
104. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ*. 25 de enero de 2016;i6.
105. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Mak Int J Soc Med Decis Mak*. diciembre de 2006;26(6):565-74.
106. Van Calster B, Vickers AJ, Pencina MJ, Baker SG, Timmerman D, Steyerberg EW. Evaluation of markers and risk prediction models: overview of relationships between NRI and decision-analytic measures. *Med Decis Mak Int J Soc Med Decis Mak*. mayo de 2013;33(4):490-501.
107. Vergara IA, Norambuena T, Ferrada E, Slater AW, Melo F. StAR: a simple tool for the statistical comparison of ROC curves. *BMC Bioinformatics*. 5 de junio de 2008;9(1):265.
108. Hanley JA. The use of the «binormal» model for parametric ROC analysis of quantitative diagnostic tests. *Stat Med*. 30 de julio de 1996;15(14):1575-85.
109. Metz CE. Basic principles of ROC analysis. *Semin Nucl Med*. octubre de 1978;8(4):283-98.
110. Zou KH, Hall WJ, Shapiro DE. Smooth non-parametric receiver operating characteristic (ROC) curves for continuous diagnostic tests. *Stat Med*. 15 de octubre de 1997;16(19):2143-56.

111. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. septiembre de 1988;44(3):837-45.
112. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*. septiembre de 1983;148(3):839-43.
113. Uchechukwu O, Ibezimako J. Comparison of Two or More Correlated AUCS in Paired Sample Design [Internet]. *EA Journals*. 2018; 6(2): 39-55.
114. Demler OV, Pencina MJ, D'Agostino RB. Misuse of DeLong test to compare AUCs for nested models. *Stat Med*. 15 de octubre de 2012;31(23):2577-87.
115. Simon R, Altman DG. Statistical aspects of prognostic factor studies in oncology. *Br J Cancer*. junio de 1994;69(6):979-85.
116. Bouwmeester W, Zuithoff NPA, Mallett S, Geerlings MI, Vergouwe Y, Steyerberg EW, et al. Reporting and methods in clinical prediction research: a systematic review. *PLoS Med*. 2012;9(5):1-12.
117. Mallett S, Royston P, Dutton S, Waters R, Altman DG. Reporting methods in studies developing prognostic models in cancer: a review. *BMC Med*. 30 de marzo de 2010;8(1):20.
118. Collins GS, de Groot JA, Dutton S, Omar O, Shanyinde M, Tajar A, et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC Med Res Methodol*. 19 de marzo de 2014;14:40.
119. Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, et al. GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias). *J Clin Epidemiol*. abril de 2011;64(4):407-15.
120. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *The BMJ*. 18 de octubre de 2011;343:d5928.
121. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*. 12 de octubre de 2016;355:i4919.
122. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 18 de octubre de 2011;155(8):529-36.
123. Whiting P, Savović J, Higgins JPT, Caldwell DM, Reeves BC, Shea B, et al. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *J Clin Epidemiol*. enero de 2016;69:225-34.
124. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, Reitsma JB, Kleijnen J, Mallett S; PROBAST Group†. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann Intern Med*. 2019 Jan 1;170(1):51-58. doi: 10.7326/M18-1376. PMID: 30596875.
125. CASPe: Programa de habilidades en lectura crítica español= Critical appraisal skills programme español [Internet]. Alicante: CASPe; c2022 [Consulta: 03-01-2022]. Disponible en: <https://www.redcaspe.org/>
126. Equator Network: Enhancing the Quality and Transparency Of health Research [Internet]; c2022. Resources in Spanish=Recursos en español [Consulta: 04-01-2022]. Disponible en: <https://www.equator-network.org/library/spanish-resources-recursos-en-espanol/>
127. G S Collins, J B Reitsma, D G Altman, K G M Moons, Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD Statement, *British*

- Journal of Surgery*, Volume 102, Issue 3, February 2015, Pages 148–158, <https://doi.org/10.1002/bjs.9736>
128. COVID Precise [Internet]. [Consulta:3-1-2022]. Disponible en: <https://www.covprecise.org/project/>
 129. COVID Precise [Internet]. [Consulta: 3-1-2022]. Disponible en: <https://www.covprecise.org/living-review/>
 130. Wynants L, Calster BV, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ*. 2020 Apr 7; 369:m1328.
 131. Jehi L, Ji X, Milinovich A, Erzurum S, Merlino A, Gordon S, et al. Development and validation of a model for individualized prediction of hospitalization risk in 4,536 patients with COVID-19. *PloS One*. 2020;15(8):e0237419.
 132. Jehi L, Ji X, Milinovich A, Erzurum S, Rubin BP, Gordon S, et al. Individualizing Risk Prediction for Positive Coronavirus Disease 2019 Testing: Results From 11,672 Patients. *Chest*. octubre de 2020;158(4):1364-75.
 133. Knight SR, Ho A, Pius R, Buchan I, Carson G, Drake TM et al. Risk stratification of patients admitted to hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: development and validation of the 4C Mortality Score. *BMJ*. 2020 Sep 9;370:m3339. doi: 10.1136/bmj.m3339. Erratum in: *BMJ*. 2020 Nov 13;371:m4334. PMID: 32907855; PMCID: PMC7116472.
 134. Cochrane.org [Internet]. [Consulta:3-1-2022]. Disponible en: <https://www.cochrane.org/welcome>.
 135. Search | Cochrane Library [Internet]. [Consulta:3-1-2022]. Disponible en: <https://www.cochranelibrary.com/es/search?cookiesEnabled>.
 136. Struyf T, Deeks JJ, Dinnes J, Takwoingi Y, Davenport C, Leeftang MM, Spijker R, Hooft L, Emperador D, Dittich S, Domen J, Horn SRA, Van den Bruel A; Cochrane COVID-19 Diagnostic Test Accuracy Group. Signs and symptoms to determine if a patient presenting in primary care or hospital outpatient settings has COVID-19 disease. *Cochrane Database Syst Rev*. 2020 Jul 7;7(7):CD013665.
 137. World Health Organization [Internet]. Geneva: WHO; c2022. Coronavirus disease (COVID-19) pandemic; [2022] [Consulta: 3-1-2022]; [aproximadamente 4 pantallas]. Disponible en: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>
 138. 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.
 139. Van Loon N, Verbrugge M, Cartuyvels R, Ramaekers D. Diagnosis of COVID-19 Based on Symptomatic Analysis of Hospital Healthcare Workers in Belgium: Observational Study in a Large Belgian Tertiary Care Center During Early COVID-19 Outbreak. *J Occup Environ Med*. 2021;63(1):27-31.
 140. Wong HYF, Lam HYS, Fong AH-T, Leung ST, Chin TW-Y, Lo CSY, et al. Frequency and Distribution of Chest Radiographic Findings in Patients Positive for COVID-19. *Radiology*. 1 de agosto de 2020;296(2):E72-8.
 141. Bernheim A, Mei X, Huang M, Yang Y, Fayad ZA, Zhang N, et al. Chest CT Findings in Coronavirus Disease-19 (COVID-19): Relationship to Duration of Infection. *Radiology*. 1 de junio de 2020;295(3):200463.
 142. Lu W, Zhang S, Chen B, Chen J, Xian J, Lin Y, et al. A Clinical Study of Noninvasive Assessment of Lung Lesions in Patients with Coronavirus Disease-19 (COVID-19) by Bedside Ultrasound. *Ultraschall Med Stuttg Ger* 1980. junio de 2020;41(3):300-7.

143. Park D-H, Park K, Park J, Park H-H, Chae H, Lim J, et al. Screening of sepsis using leukocyte cell population data from the Coulter automatic blood cell analyzer DxH800. *Int J Lab Hematol.* agosto de 2011;33(4):391-9.
144. Jung Y-J, Kim J-H, Park Y-J, Kahng J, Lee H, Lee K-Y, et al. Evaluation of cell population data on the UniCel DxH 800 Coulter Cellular Analysis system as a screening for viral infection in children. *Int J Lab Hematol.* junio de 2012;34(3):283-9.
145. Crouser ED, Parrillo JE, Seymour C, Angus DC, Bicking K, Tejjidor L, et al. Improved Early Detection of Sepsis in the ED With a Novel Monocyte Distribution Width Biomarker. *Chest.* septiembre de 2017;152(3):518-26.
146. Van de Geijn G-JM, Denker S, Meuleman-van Waning V, Koeleman HGM, Birnie E, Braunstahl G-J, et al. Evaluation of new laboratory tests to discriminate bacterial from nonbacterial chronic obstructive pulmonary disease exacerbations. *Int J Lab Hematol.* diciembre de 2016;38(6):616-28.
147. Royal College of Physicians. *National Early Warning Score (NEWS) 2: Standardising the assessment of acute-illness severity in the NHS.* Updated report of a working party. London: RCP, 2017.
148. R Paterson, DC MacLeod, D Thetford, A Beattie, C Graham, S Lam, D Bell. *Clinical Medicine* May 2006, 6 (3) 281-284; DOI: 10.7861/clinmedicine.6-3-281.
149. Health Standards Organization [Internet]. Ottawa: HSO; c2021. *Leading Practices: The Hamilton Early Warning Score: Improved safety through attention to abnormal vital signs; 2015* [Consulta: 3-1-2022]; [aproximadamente 2 pantallas]. Disponible en: <https://healthstandards.org/leading-practice/the-hamilton-early-warning-score-improved-safety-through-attention-to-abnormal-vital-signs/>
150. Burch VC, Tarr G, Morroni C. Modified early warning score predicts the need for hospital admission and inhospital mortality. *Emerg Med J.* 1 de octubre de 2008;25(10):674-8.
151. MuLBSTA Score to Predict Mortality in Viral Pneumonia [Internet]. [Pittsburgh]: Foundation Internet Services; c1998-2020 [Consulta: 3-1-2022]. Disponible en: https://www.msmanuals.com/medical-calculators/MuLBSTA_Score.htm
152. CURB-65 Score for Pneumonia Severity [Internet]. [New York]: MDCalc; c2005-2022 [Consulta: 3-1-2022]. Disponible en: <https://www.mdcalc.com/curb-65-score-pneumonia-severity>
153. Sequential Organ Failure Assessment (SOFA) Score [Internet]. [New York]: MDCalc; c2005-2022 [Consulta:3-1-2022]. Disponible en: <https://www.mdcalc.com/sequential-organ-failure-assessment-sofa-score>
154. qSOFA: quick Sepsis Related Organ Failure Assessment [Internet]. [Pittsburgh]: UPMC, University of Pittsburgh; [2022] [Consulta: 3-1-2022]. Disponible en: <https://qsofa.org/>
155. Gupta RK, Marks M, Samuels THA, Luintel A, Rampling T, Chowdhury H, et al. Systematic evaluation and external validation of 22 prognostic models among hospitalised adults with COVID-19: an observational cohort study. *Eur Respir J.* diciembre de 2020;56(6):2003498.
156. Gidari A., De Socio G.V., Sabbatini S., Francisci D. Predictive value of National Early Warning Score 2 (NEWS2) for intensive care unit admission in patients with SARS-CoV-2 infection. *Infect Dis.* 2020;52(10):698-704.
157. Kucirka LM, Lauer SA, Laeyendecker O, Boon D, Lessler J. Variation in False-Negative Rate of Reverse Transcriptase Polymerase Chain Reaction–Based SARS-CoV-2 Tests by Time Since Exposure. *Ann Intern Med.* 13 de mayo de 2020;173(4):262-7.
158. Ministerio de Sanidad, Servicios Sociales e Igualdad [Internet]. Madrid: Ministerio de Sanidad, Servicios Sociales e Igualdad; [2022]. *Compromiso por la Calidad de las Sociedades Científicas*

- en España; [2022] [Consulta:3-1-2022]; [aproximadamente 5 pantallas]. Disponible en: https://www.sanidad.gob.es/organizacion/sns/planCalidadSNS/cal_sssc.htm
159. Zou L, Ruan F, Huang M, Liang L, Huang H, Hong Z, et al. SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients. *N Engl J Med*. 19 de marzo de 2020;382(12):1177-9.
 160. Pan Y, Zhang D, Yang P, Poon LLM, Wang Q. Viral load of SARS-CoV-2 in clinical samples. *Lancet Infect Dis*. abril de 2020;20(4):411-2.
 161. To KK-W, Tsang OT-Y, Leung W-S, Tam AR, Wu T-C, Lung DC, et al. Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study. *Lancet Infect Dis*. mayo de 2020;20(5):565-74.
 162. Wölfel R, Corman VM, Guggemos W, Seilmaier M, Zange S, Müller MA, et al. Virological assessment of hospitalized patients with COVID-2019. *Nature*. mayo de 2020;581(7809):465-9.
 163. European Centre for Disease Prevention and Control [Internet]. Stockholm: ECDC; c2022. Rapid risk assessment: Coronavirus disease 2019 (COVID-19) pandemic: increased transmission in the EU/EEA and the UK – eighth update; 8 de abril de 2020 [Consulta:3-1-2022]; 39 p. Disponible en: <https://www.ecdc.europa.eu/en/publications-data/covid-19-rapid-risk-assessment-coronavirus-disease-2019-pandemic-eighth-update>
 164. World Health Organization. Coronavirus disease 2019 (COVID-19): situation report, 73 [Internet]. Geneva: World Health Organization; 2 de abril 2020 [Consulta:3-1-2022]. 13 p. Disponible en: <https://apps.who.int/iris/handle/10665/331686>
 165. Böger B, Fachi MM, Vilhena RO, Cobre AF, Tonin FS, Pontarolo R. Systematic review with meta-analysis of the accuracy of diagnostic tests for COVID-19. *Am J Infect Control*. 2021 Jan;49(1):21-29. doi: 10.1016/j.ajic.2020.07.011. Epub 2020 Jul 10. PMID: 32659413; PMCID: PMC7350782.
 166. Kim H, Hong H, Yoon SH. Diagnostic Performance of CT and Reverse Transcriptase Polymerase Chain Reaction for Coronavirus Disease 2019: A Meta-Analysis. *Radiology*. septiembre de 2020;296(3):E145-55.
 167. Yang Y, Yang M, Shen C, Wang F, Yuan J, Li J, et al. Evaluating the accuracy of different respiratory specimens in the laboratory diagnosis and monitoring the viral shedding of 2019-nCoV infections. *Infectious Diseases (except HIV/AIDS)*; 2020 feb. <https://doi.org/10.1101/2020.02.11.20021493>.
 168. Maroto-Vela M. SARS-CoV-2: Problemas e incertidumbres. *ANALES RANM* [Internet]. Real Academia Nacional de Medicina de España; An RANM · Año 2020 · numero 137(02):98-103. DOI: <http://dx.doi.org/10.32440/ar.2020.137.02.rev01>.
 169. European Centre for Disease Prevention and Control [Internet]. Stockholm: ECDC; c2022. Case definition for coronavirus disease 2019 (COVID-19), as of 3 December 2020; 3 de diciembre de 2020 [Consulta:3-1-2022]; [aproximadamente 2 pantallas]. Disponible en: <https://www.ecdc.europa.eu/en/covid-19/surveillance/case-definition>
 170. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. septiembre de 1986;7(3):177-88.
 171. Borenstein M, Higgins JPT, Hedges LV, Rothstein HR. Basics of meta-analysis: I² is not an absolute measure of heterogeneity. *Res Synth Methods*. marzo de 2017;8(1):5-18.
 172. Bartoletti M, Giannella M, Scudeller L, Tedeschi S, Rinaldi M, Bussini L et al.; PREDICO study group. Development and validation of a prediction model for severe respiratory failure in hospitalized patients with SARS-CoV-2 infection: a multicentre cohort study (PREDI-CO study). *Clin Microbiol Infect*. 2020 Nov;26(11):1545-1553.

173. Weng Z, Chen Q, Li S, Li H, Zhang Q, Lu S, et al. ANDC: an early warning score to predict mortality risk for patients with Coronavirus Disease 2019. *J Transl Med.* 2020;18(1):328.
174. Zhang S, Guo M, Duan L, Wu F, Hu G, Wang Z, et al. Development and validation of a risk factor-based system to predict short-term survival in adult hospitalized patients with COVID-19: a multicenter, retrospective, cohort study. *Crit Care Lond Engl.* 2020;24(1):438.
175. Shang Y, Liu T, Wei Y, Li J, Shao L, Liu M, et al. Scoring systems for predicting mortality for severe patients with COVID-19. *EClinicalMedicine.* 2020;24(101733727):100426.
176. Liu Q, Song NC, Zheng ZK, Li JS, Li SK. Laboratory findings and a combined multifactorial approach to predict death in critically ill patients with COVID-19: a retrospective study. *Epidemiol Infect.* 2020;148(epi, 8703737):e129.
177. Xiao L-S, Zhang W-F, Gong M-C, Zhang Y-P, Chen L-Y, Zhu H-B, et al. Development and validation of the HNC-LL score for predicting the severity of coronavirus disease 2019. *EBioMedicine.* 2020;102880-102880.
178. Linares M, Pérez-Tanoira R, Carrero A, Romanyk J, Pérez-García F, Gómez-Herruz P, et al. Panbio antigen rapid test is reliable to diagnose SARS-CoV-2 infection in the first 7 days after the onset of symptoms. *J Clin Virol.* 1 de diciembre de 2020;133:104659.
179. Albert E, Torres I, Bueno F, Huntley D, Molla E, Fernández-Fuentes MÁ et al. Field evaluation of a rapid antigen test (Panbio™ COVID-19 Ag Rapid Test Device) for COVID-19 diagnosis in primary healthcare centres. *Clin Microbiol Infect.* 2021 Mar;27(3):472.e7-472.e10. doi: 10.1016/j.cmi.2020.11.004. Epub 2020 Nov 13. PMID: 33189872; PMCID: PMC7662075.
180. Gremmels H, Winkel BMF, Schuurman R, Rosingh A, Rigter NAM, Rodriguez O, Ubijaan J, Wensing AMJ, Bonten MJM, Hofstra LM. Real-life validation of the Panbio™ COVID-19 antigen rapid test (Abbott) in community-dwelling subjects with symptoms of potential SARS-CoV-2 infection. *EClinicalMedicine.* 2021 Jan;31:100677. doi: 10.1016/j.eclinm.2020.100677. Epub 2020 Dec 5. PMID: 33521610; PMCID: PMC7832943.
181. Fenollar F, Bouam A, Ballouche M, Fuster L, Prudent E, Colson P, et al. Evaluation of the Panbio Covid-19 rapid antigen detection test device for the screening of patients with Covid-19. *J Clin Microbiol.* 2 de noviembre de 2020;
182. Alemany A, Baró B, Ouchi D, Rodó P, Ubals M, Corbacho-Monné M, et al. Analytical and clinical performance of the panbio COVID-19 antigen-detecting rapid diagnostic test. *J Infect.* 7 de enero de 2021;
183. Berger A, Nsoga MTN, Perez-Rodriguez FJ, Aad YA, Sattonnet-Roche P, Gayet-Ageron A, et al. Diagnostic accuracy of two commercial SARS-CoV-2 antigen-detecting rapid tests at the point of care in community-based testing centers. *PLOS ONE.* 31 de marzo de 2021;16(3):e0248921.
184. Mar Masiá, Marta Fernández-González, Manuel Sánchez, Mar Carvajal, José Alberto García et al. Nasopharyngeal Panbio COVID-19 Antigen Performed at Point-of-Care Has a High Sensitivity in Symptomatic and Asymptomatic Patients With Higher Risk for Transmission and Older Age, *Open Forum Infectious Diseases*, Volume 8, Issue 3, March 2021, ofab059, <https://doi.org/10.1093/ofid/ofab059>
185. Dřevínek P, Hurych J, Kepka Z, Briksi A, Kulich M, Zajac M, et al. The sensitivity of SARS-CoV-2 antigen tests in the view of large-scale testing. *Epidemiol Mikrobiol Imunol Cas Spolecnosti Epidemiol Mikrobiol Ceske Lek Spolecnosti JE Purkyne.* 2021;70(3):156-60.
186. Bulilete O, Lorente P, Leiva A, Carandell E, Oliver A, Rojo E, et al. Evaluation of the Panbio™ rapid antigen test for SARS-CoV-2 in primary health care centers and test sites. 2020 nov. p. 2020.11.13.20231316.
187. Merino P, Guinea J, Muñoz-Gallego I, González-Donapetry P, Galán JC, Antona N, et al. Multicenter evaluation of the Panbio™ COVID-19 rapid antigen-detection test for the diagnosis

- of SARS-CoV-2 infection. *Clin Microbiol Infect Off Publ Eur Soc Clin Microbiol Infect Dis.* 16 de febrero de 2021;S1198-743X(21)00076-8.
188. Abdulrahman A, Mustafa F, AlAwadhi AI, Alansari Q, AlAlawi B, AlQahtani M. Comparison of SARS-COV-2 nasal antigen test to nasopharyngeal RT-PCR in mildly symptomatic patients. :16.
 189. Billaud G, Gaymard A, Lina B. Laboratoire de Virologie des HCL CNR des virus des infections respiratoires. Evaluation du Test Antigénique ABBOTT SARS-COV2 ABBOTT. Lyon, France: SFM (French Society of Microbiology); 2020.
 190. Schwob JM, Miauton A, Petrovic D, Perdrix J, Senn N, Jaton K, et al. Title: Antigen rapid tests, nasopharyngeal PCR and saliva PCR to detect SARS-CoV-2: a prospective comparative clinical trial. :16.
 191. Torres I, Poujois S, Albert E, Colomina J, Navarro D. Evaluation of a rapid antigen test (Panbio™ COVID-19 Ag rapid test device) for SARS-CoV-2 detection in asymptomatic close contacts of COVID-19 patients. *Clin Microbiol Infect Off Publ Eur Soc Clin Microbiol Infect Dis.* abril de 2021;27(4):636.e1-636.e4.
 192. Harrell FE. Binary Logistic Regression. En: Harrell FE, editor. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis.* New York, NY: Springer; 2001. p. 215-67.
 193. Sun GW, Shook TL, Kay GL. Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. *J Clin Epidemiol.* agosto de 1996;49(8):907-16.
 194. Tsvetanova A, Sperrin M, Peek N, Buchan I, Hyland S, Martin GP. Missing data was handled inconsistently in UK prediction models: a review of method used. *J Clin Epidemiol.* 1 de diciembre de 2021;140:149-58.
 195. Little R, Rubin D. *Statistical Analysis with Missing Data*, 2nd Edition. Wiley. 2019.
 196. Janssen KJM, Donders ART, Harrell FE, Vergouwe Y, Chen Q, Grobbee DE, et al. Missing covariate data in medical research: to impute is better than to ignore. *J Clin Epidemiol.* julio de 2010;63(7):721-7.
 197. Marshall A, Altman DG, Royston P, Holder RL. Comparison of techniques for handling missing covariate data within prognostic modelling studies: a simulation study. *BMC Med Res Methodol.* 19 de enero de 2010;10(1):7.
 198. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ.* 29 de junio de 2009;338:b2393.
 199. Altman DG, Lausen B, Sauerbrei W, Schumacher M. Dangers of using «optimal» cutpoints in the evaluation of prognostic factors. *J Natl Cancer Inst.* 1 de junio de 1994;86(11):829-35.
 200. Altman DG, Royston P. The cost of dichotomising continuous variables. *BMJ.* 4 de mayo de 2006;332(7549):1080.
 201. Riley RD, Snell KIE, Martin GP, Whittle R, Archer L, Sperrin M, et al. Penalization and shrinkage methods produced unreliable clinical prediction models especially when sample size was small. *J Clin Epidemiol.* abril de 2021;132:88-96.
 202. Moons KGM, Donders ART, Steyerberg EW, Harrell FE. Penalized maximum likelihood estimation to directly adjust diagnostic and prognostic prediction models for overoptimism: a clinical example. *J Clin Epidemiol.* diciembre de 2004;57(12):1262-70.
 203. Counsell C, Dennis M. Systematic review of prognostic models in patients with acute stroke. *Cerebrovasc Dis Basel Switz.* 2001;12(3):159-70.
 204. Perel P, Edwards P, Wentz R, Roberts I. Systematic review of prognostic models in traumatic brain injury. *BMC Med Inform Decis Mak.* 2006 Nov 14;6:38. doi: 10.1186/1472-6947-6-38. PMID: 17105661; PMCID: PMC1657003.

205. Oudega R, Hoes AW, Moons KGM. The Wells rule does not adequately rule out deep venous thrombosis in primary care patients. *Ann Intern Med*. 19 de julio de 2005;143(2):100-7.
206. Wolfisberg S, Gregoriano C, Struja T, Kutz A, Koch D, Bernasconi L, Hammerer-Lercher A, Mohr C, Haubitz S, Conen A, Fux CA, Mueller B, Schuetz P. Call, chosen, HA₂T₂, ANDC: validation of four severity scores in COVID-19 patients. *Infection*. 2021 Nov 19:1–9. doi: 10.1007/s15010-021-01728-0. Epub ahead of print. PMID: 34799814; PMCID: PMC8604199.
207. Cyprian FS, Suleman M, Abdelhafez I, Doudin A, Masud Danjuma IM, Mir FA, et al. Complement C5a and Clinical Markers as Predictors of COVID-19 Disease Severity and Mortality in a Multi-Ethnic Population. *Front Immunol*. 2021;12:5247.
208. Bilge M, Akilli IK, Karaayvaz EB, Yesilova A, Kart Yasar K. Comparison of systemic immune-inflammation index (SII), early warning score (ANDC) and prognostic nutritional index (PNI) in hospitalized patients with malignancy, and their influence on mortality from COVID-19. *Infect Agent Cancer*. 15 de septiembre de 2021;16(1):60.
209. Lombardi Y, Azoyan L, Szychowiak P, Bellamine A, Lemaitre G, Bernaux M, et al. External validation of prognostic scores for COVID-19: a multicenter cohort study of patients hospitalized in Greater Paris University Hospitals. *Intensive Care Med*. diciembre de 2021;47(12):1426-39.
210. Wyatt J, Altman D. Commentary: Prognostic models: Clinically useful or quickly forgotten? *BMJ*. 9 de diciembre de 1995;311:1539-41.
211. Roger S, Lefeuvre C, Pivert A, Ducancelle A, Savary D, Bouthry É, Le Guillou-Guillemette H. What is the true place of the SARS-CoV-2 rapid point-of-care antigen test in the hospital setting? Lessons learned from real life. *J Med Virol*. 2021 Dec 6. doi: 10.1002/jmv.27505. Epub ahead of print. PMID: 34873718.
212. Canadian CT Head and C-Spine (CCC) Study Group. Canadian C-Spine Rule study for alert and stable trauma patients: I. Background and rationale. *CJEM*. marzo de 2002;4(2):84-90.
213. Canadian CT Head and C-Spine (CCC) Study Group. Canadian C-Spine Rule study for alert and stable trauma patients: II. Study objectives and methodology. *CJEM*. mayo de 2002;4(3):185-93.
214. Williams RJ, Tse T, Harlan WR, Zarin DA. Registration of observational studies: is it time? *CMAJ Can Med Assoc J J Assoc Medicale Can*. 19 de octubre de 2010;182(15):1638-42.
215. Hemingway H, Riley RD, Altman DG. Ten steps towards improving prognosis research. *BMJ*. 30 de diciembre de 2009;339:b4184.

ANEXOS

1. BÚSQUEDAS BIBLIOGRÁFICAS

Búsqueda bibliográfica para modelos predictivos de diagnóstico realizada en EMBASE.

Database: Embase <1974 to 2021 Week 12>

Search Strategy:

- 1 exp Coronavirus/ (23501)
- 2 exp Coronavirus Infections/ (24562)
- 3 (coronavirus* or corona virus* or OC43 or NL63 or 229E or HKU1 or HCoV* or nCoV* or covid* or sars-cov* or sarscov* or Sars-coronavirus* or Severe Acute Respiratory Syndrome Coronavirus*).mp. (146492)
- 4 (or/1-3) and 20190101:20301231.(dc). [this set is the sensitive/broad part of the search] (125533)
- 5 4 not (SARS or SARS-CoV or MERS or MERS-CoV or Middle East respiratory syndrome or camel* or dromedar* or equine or coronary or coronal or covidence* or covidien or influenza virus or HIV or bovine or calves or TGEV or feline or porcine or BCoV or PED or PEDV or PDCoV or FIPV or FCoV or SADS-CoV or canine or CCov or zoonotic or avian influenza or H1N1 or H5N1 or H5N6 or IBV or murine corona*).mp. [line 5 removes noise in the search results] (77970)
- 6 ((pneumonia or covid* or coronavirus* or corona virus* or nCoV* or 2019-nCoV or sars*) .mp. or exp pneumonia/ and Wuhan.mp. (5047)
- 7 (coronavirus disease 2019 or 2019-nCoV or nCoV19 or nCoV-19 or 2019-novel CoV or severe acute respiratory syndrome coronavirus 2 or sars-cov2 or sars-cov-2 or sarscov2 or sarscov-2 or Sars-coronavirus2 or Sars-coronavirus-2 or SARS-like coronavirus* or coronavirus-19 or covid19 or covid-19 or ((novel or new or nouveau) adj2 (CoV or nCoV or covid or coronavirus* or corona virus or Pandemi*2)) or ((covid or covid19 or covid-19) and pandemic*2) or (coronavirus* and pneumonia)).mp. (123135)
- 8 (coronavirus disease 2019 or severe acute respiratory syndrome coronavirus 2).sh,dj. (101137)
- 9 ("630575119" or "630830186" or "630941329" or "631043694" or "631260659" or "631272428" or "631272880" or "631286076" or "631290163" or "631308782" or "631324397" or "631352500" or "631416440" or "631431802" or "631452886" or "631456079" or "631457551" or "631462438" or "631462876" or "631465538" or "631465685" or "631469310" or "2004499662" or "2004505338" or "2005280837" or "2005387675" or "2005408544" or "2005484987" or "2005549151").an. [Articles not captured by this search when created in April 2020, pending further indexing by NLM/Elsevier] (16)
- 10 (or/6-9) and 20191201:20301231.(dc). [Lines 5 to 8 are specific to Covid-19] (120049)
- 11 5 or 10 (121877)
- 12 models theoretical.mp. or theoretical model/ (86500)
- 13 diagnosis/ or diagnosis.mp. (5610438)
- 14 predictive value.mp. or predictive value/ (254301)
- 15 (predict* adj3 (model or models)).m_titl. (20920)
- 16 ((diagnostic or diagnosis) adj3 (score or tool or tools)).m_titl. (8680)
- 17 ((diagnostic or diagnosis) adj2 nomogram*).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word] (82)
- 18 12 and 13 (4009)
- 19 (Signs and symptoms).m_titl. (3538)
- 20 14 or 15 or 16 or 17 or 18 or 19 (288143)
- 21 11 and 20 (1333)
- 22 20200701:20200905.(dc). (342032)
- 23 21 and 22 (178)
- 24 limit 23 to (conference abstracts or embase) (141)

Búsqueda bibliográfica para modelos predictivos de diagnóstico realizada en MEDLINE.

Database: Ovid MEDLINE(R) ALL / PubMed(R) <1946 to Present>

Search Strategy:

```

1  exp Coronavirus/ (64584)
2  exp Coronavirus Infections/ (77644)
3  (coronavirus* or corona virus* or OC43 or NL63 or 229E or HKU1 or HCoV* or ncov* or covid* or sars-cov* or
sarscov* or Sars-coronavirus* or Severe Acute Respiratory Syndrome Coronavirus*).mp. (133749)
4  (or/1-3) and ((20191* or 202*).dp. or 20190101:20301231.(ep.)) [this set is the sensitive/broad part of the search]
(120765)
5  4 not (SARS or SARS-CoV or MERS or MERS-CoV or Middle East respiratory syndrome or camel* or dromedar*
or equine or coronary or coronal or covidence* or covidien or influenza virus or HIV or bovine or calves or TGEV or
feline or porcine or BCoV or PED or PEDV or PDCoV or FIPV or FCoV or SADS-CoV or canine or CCov or zoonotic
or avian influenza or H1N1 or H5N1 or H5N6 or IBV or murine corona*).mp. [line 5 removes noise in the search
results] (45085)
6  ((pneumonia or covid* or coronavirus* or corona virus* or ncov* or 2019-ncov or sars*).mp. or exp pneumonia/
and Wuhan.mp. (4774)
7  (2019-ncov or ncov19 or ncov-19 or 2019-novel CoV or sars-cov2 or sars-cov-2 or sarscov2 or sarscov-2 or
Sars-coronavirus2 or Sars-coronavirus-2 or SARS-like coronavirus* or coronavirus-19 or covid19 or covid-19 or covid
2019 or ((novel or new or nouveau) adj2 (CoV or nCoV or covid or coronavirus* or corona virus or Pandemi*2)) or
((covid or covid19 or covid-19) and pandemic*2) or (coronavirus* and pneumonia)).mp. (119131)
8  covid-19.mp. or COVID-19.rx.px.ox. or severe acute respiratory syndrome coronavirus 2.os. (112704)
9  ("32240632" or "32236488" or "32268021" or "32267941" or "32169616" or "32267649" or "32267499" or
"32267344" or "32248853" or "32246156" or "32243118" or "32240583" or "32237674" or "32234725" or "32173381"
or "32227595" or "32185863" or "32221979" or "32213260" or "32205350" or "32202721" or "32197097" or
"32196032" or "32188729" or "32176889" or "32088947" or "32277065" or "32273472" or "32273444" or "32145185"
or "31917786" or "32267384" or "32265186" or "32253187" or "32265567" or "32231286" or "32105468" or
"32179788" or "32152361" or "32152148" or "32140676" or "32053580" or "32029604" or "32127714" or "32047315"
or "32020111" or "32267950" or "32249952" or "32172715").ui. [Articles not captured by this search when created in
April 2020, pending further indexing by NLM] (49)
10 or/6-9 [Lines 6 to 9 are specific to Covid-19] (119196)
11 5 or 10 (121464)
12 11 and 20191201:20301231.(dt). (119314)
13 (predict* adj3 (model or models)).m_titl. (15403)
14 ((diagnostic or diagnosis) adj3 (score or tool or tools)).m_titl. (6024)
15 ((diagnostic or diagnosis) adj2 nomogram*).mp. [mp=title, abstract, original title, name of substance word,
subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word,
protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms] (65)
16 (Signs and symptoms).m_titl. (2823)
17 models theoretical.mp. or Models, Theoretical/ (155350)
18 Diagnosis/ or diagnosis.mp. (3699512)
19 17 and 18 (6709)
20 predictive value of tests.mp. or "Predictive Value of Tests"/ (209588)
21 13 or 14 or 15 or 16 or 19 or 20 (237565)
22 12 and 21 (715)
23 20200701:20200905.(dt). (292076)
24 20200701:20200905.ep. (4413)
25 23 or 24 (292897)
26 22 and 25 (174)

```

Búsqueda bibliográfica para modelos predictivos de pronóstico realizada en EMBASE.

Database: Embase <1974 to 2021 Week 06>

Search Strategy:

- 1 exp Coronavirus/ (23061)
- 2 exp Coronavirus Infections/ (24604)
- 3 (coronavirus* or corona virus* or OC43 or NL63 or 229E or HKU1 or HCoV* or nCoV* or covid* or sars-cov* or sarscov* or Sars-coronavirus* or Severe Acute Respiratory Syndrome Coronavirus*).mp. (128661)
- 4 (or/1-3) and 20190101:20301231.(dc). [this set is the sensitive/broad part of the search] (107607)
- 5 4 not (SARS or SARS-CoV or MERS or MERS-CoV or Middle East respiratory syndrome or camel* or dromedar* or equine or coronary or coronal or cvidence* or covidien or influenza virus or HIV or bovine or calves or TGEV or feline or porcine or BCoV or PED or PEDV or PDCoV or FIPV or FCoV or SADS-CoV or canine or CCov or zoonotic or avian influenza or H1N1 or H5N1 or H5N6 or IBV or murine corona*).mp. [line 5 removes noise in the search results] (67211)
- 6 ((pneumonia or covid* or coronavirus* or corona virus* or nCoV* or 2019-nCoV or sars*).mp. or exp pneumonia/) and Wuhan.mp. (4641)
- 7 (coronavirus disease 2019 or 2019-nCoV or nCoV19 or nCoV-19 or 2019-novel CoV or severe acute respiratory syndrome coronavirus 2 or sars-cov2 or sars-cov-2 or sarscov2 or sarscov-2 or Sars-coronavirus2 or Sars-coronavirus-2 or SARS-like coronavirus* or coronavirus-19 or covid19 or covid-19 or covid 2019 or ((novel or new or nouveau) adj2 (CoV or nCoV or covid or coronavirus* or corona virus or Pandemi*2)) or ((covid or covid19 or covid-19) and pandemic*2) or (coronavirus* and pneumonia)).mp. (105564)
- 8 (coronavirus disease 2019 or severe acute respiratory syndrome coronavirus 2).sh,dj. (87210)
- 9 ("630575119" or "630830186" or "630941329" or "631043694" or "631260659" or "631272428" or "631272880" or "631286076" or "631290163" or "631308782" or "631324397" or "631352500" or "631416440" or "631431802" or "631452886" or "631456079" or "631457551" or "631462438" or "631462876" or "631465538" or "631465685" or "631469310" or "2004499662" or "2004505338" or "2005280837" or "2005387675" or "2005408544" or "2005484987" or "2005549151").an. [Articles not captured by this search when created in April 2020, pending further indexing by NLM/Elsevier] (16)
- 10 (or/6-9) and 20191201:20301231.(dc). [Lines 5 to 8 are specific to Covid-19] (102468)
- 11 5 or 10 (104175)
- 12 models theoretical.mp. or theoretical model/ (86276)
- 13 prognosis/ or prognosis.mp. (993431)
- 14 predictive value.mp. or predictive value/ (250769)
- 15 (predict* adj3 (model or models)).mp. (160369)
- 16 ((prognostic or prognosis) adj3 (tool or tools)).mp. (9183)
- 17 ((prognostic or prognosis) adj2 nomogram*).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word] (1117)
- 18 12 and 13 (574)
- 19 (risk adj3 (score or tool)).mp. (65552)
- 20 14 or 15 or 16 or 17 or 18 or 19 (464894)
- 21 11 and 20 (1672)
- 22 death.mp. or death/ (1317060)
- 23 mortality/ or mortality.mp. (1560414)
- 24 critical illness.mp. or critical illness/ (37398)
- 25 Intensive Care Units.mp. or intensive care unit/ (182174)
- 26 (intensive adj3 (unit or units)).mp. (252342)
- 27 22 or 23 or 24 or 25 or 26 (2714492)
- 28 21 and 27 (611)
- 29 20200701:20200901.(dc). (334234)
- 30 28 and 29 (96)
- 31 limit 30 to (conference abstracts or embase) (75)

Búsqueda bibliográfica para modelos predictivos de pronóstico realizada en MEDLINE

Database: Ovid MEDLINE(R) ALL / PubMed(R) <1946 to Present>

Search Strategy:

- 1 exp Coronavirus/ (45624)
- 2 exp Coronavirus Infections/ (49830)
- 3 (coronavirus* or corona virus* or OC43 or NL63 or 229E or HKU1 or HCoV* or ncov* or covid* or sars-cov* or sarscov* or Sars-coronavirus* or Severe Acute Respiratory Syndrome Coronavirus*).mp. (121017)
- 4 (or/1-3) and ((20191* or 202*) .dp. or 20190101:20301231.(ep).) [this set is the sensitive/broad part of the search] (108027)
- 5 4 not (SARS or SARS-CoV or MERS or MERS-CoV or Middle East respiratory syndrome or camel* or dromedar* or equine or coronary or coronal or covidence* or covidien or influenza virus or HIV or bovine or calves or TGEV or feline or porcine or BCoV or PED or PEDV or PDCoV or FIPV or FCoV or SADS-CoV or canine or CCov or zoonotic or avian influenza or H1N1 or H5N1 or H5N6 or IBV or murine corona*).mp. [line 5 removes noise in the search results] (62020)
- 6 ((pneumonia or covid* or coronavirus* or corona virus* or ncov* or 2019-ncov or sars*).mp. or exp pneumonia/) and Wuhan.mp. (4516)
- 7 (2019-ncov or ncov19 or ncov-19 or 2019-novel CoV or sars-cov2 or sars-cov-2 or sarscov2 or sarscov-2 or Sars-coronavirus2 or Sars-coronavirus-2 or SARS-like coronavirus* or coronavirus-19 or covid19 or covid-19 or covid 2019 or ((novel or new or nouveau) adj2 (CoV or nCoV or covid or coronavirus* or corona virus or Pandemi*2)) or ((covid or covid19 or covid-19) and pandemic*2) or (coronavirus* and pneumonia)).mp. (106488)
- 8 covid-19.mp. or COVID-19.rx.px.ox. or severe acute respiratory syndrome coronavirus 2.os. (100453)
- 9 ("32240632" or "32236488" or "32268021" or "32267941" or "32169616" or "32267649" or "32267499" or "32267344" or "32248853" or "32246156" or "32243118" or "32240583" or "32237674" or "32234725" or "32173381" or "32227595" or "32185863" or "32221979" or "32213260" or "32205350" or "32202721" or "32197097" or "32196032" or "32188729" or "32176889" or "32088947" or "32277065" or "32273472" or "32273444" or "32145185" or "31917786" or "32267384" or "32265186" or "32253187" or "32265567" or "32231286" or "32105468" or "32179788" or "32152361" or "32152148" or "32140676" or "32053580" or "32029604" or "32127714" or "32047315" or "32020111" or "32267950" or "32249952" or "32172715").ui. [Articles not captured by this search when created in April 2020, pending further indexing by NLM] (50)
- 10 or/6-9 [Lines 6 to 9 are specific to Covid-19] (106565)
- 11 5 or 10 (108794)
- 12 11 and 20191201:20301231.(dt). (106647)
- 13 (predict* adj3 (model or models)).m_titl. (15255)
- 14 ((prognostic or prognosis) adj3 (tool or tools)).m_titl. (672)
- 15 ((prognostic or prognosis) adj2 nomogram*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms] (847)
- 16 (risk adj3 (score or tool or tools)).m_titl. (6206)
- 17 models theoretical.mp. or Models, Theoretical/ (154712)
- 18 prognosis.mp. or Prognosis/ (787380)
- 19 17 and 18 (1661)
- 20 predictive value of tests.mp. or "Predictive Value of Tests"/ (208634)
- 21 13 or 14 or 15 or 16 or 19 or 20 (229842)
- 22 12 and 21 (559)
- 23 20200701:20200905.(dt). (306941)
- 24 20200701:20200905.ep. (4659)
- 25 23 or 24 (307737)
- 26 22 and 25 (157)

2. ALGORITMO DE TRIAJE DISEÑADO POR EL MINISTERIO DE SANIDAD, SERVICIOS SOCIALES E IGUALDAD DE ESPAÑA.

