

Lima, julio de 2021

SERIE NOTA TECNICA COVID-19 N° 20-2021

Uso de las escalas o modelos de pronóstico 4C y NEWS2 en pacientes hospitalizados con COVID-19

Recomendaciones desarrolladas por el Grupo de Trabajo designado por el MINSA



INSTITUTO NACIONAL DE SALUD

UNIDAD DE ANÁLISIS Y GENERACIÓN DE EVIDENCIAS EN SALUD PÚBLICA

SERIE NOTA TECNICA COVID-19 N° 20-2021: Uso de las escalas o modelos de pronóstico 4C y NEWS2 en pacientes hospitalizados con COVID-19

PATOLOGÍA: COVID-19

Intervención: Escalas o modelos de pronóstico 4C y NEWS2

Rol de la Intervención: Pronóstico

FECHA: 12 de julio 2021

ANTECEDENTES

- El presente informe se efectúa en atención a la solicitud de la Jefatura del Instituto Nacional de Salud.
- El objetivo del presente informe es:
 - Describir el proceso para la elaboración de recomendaciones por el grupo de trabajo designado por el Ministerio de Salud, en adelante denominado grupo de trabajo.
 - Trasladar las Recomendaciones efectuadas por dicho grupo de trabajo en atención al uso de las escalas o modelos pronóstico 4C y NEWS2 para pacientes hospitalizados con COVID-19, según las preguntas PICO (P: Población, I: Intervención, C: Comparador, O: Outcome o desenlaces).
- La metodología considerada para arribar a la recomendación fue el Marco de Evidencia a la Decisión/Recomendación desarrollado por el Grupo de Trabajo GRADE(1,2).

ANALISIS

a) Formulación de la pregunta

¿Deberían usarse las escalas o modelos de pronóstico 4C versus NEWS2 versus no usarlas en pacientes adultos con COVID-19, hospitalizados?

Contexto de una decisión para elegir área de derivación

P	Paciente con COVID-19 presunto o confirmado hospitalizado
I	Escala de mortalidad
C	Escala de deterioro 4C
O	Escala NEWS2

b) Identificación de la evidencia para la pregunta PICO

Se siguieron las orientaciones establecidas en el documento interno de UNAGESP: Orientaciones para el soporte metodológico otorgado al grupo de trabajo designado por el MINSa. SE describe a continuación los resultados del proceso:

- Se identificó una guía de práctica clínica (GPC) con una recomendación vinculada a la pregunta PICO: Manejo clínico de la COVID-19: orientaciones evolutivas, 25 de enero de 2021, elaborada por la Organización Mundial de la Salud (OMS)(3), y que utilizó la metodología GRADE para la evaluación de la certeza de la evidencia y el Marco de Evidencia para la decisión (EtD). No obstante, la recomendación emitida se efectuó de manera global para una serie de escalas o modelos pronósticos, incluidos en una revisión sistemática con búsqueda bibliográfica hasta el 02 de julio de 2020. Dado que las intervenciones de interés de la pregunta PICO eran 3 escalas o modelos pronósticos, se procedió a identificar los estudios sobre éstas, que fueron incluidos en la revisión sistemática.
- Asimismo, se efectuó una búsqueda manual de estudios primarios en el mapa de evidencias COVID-19 de The Norwegian Institute of Public Health (NIPH)(4), Topic: Prognosis y en la plataforma L-OVE de Epistemonikos(5), Filtros: Prognosis/Predictor of outcome/Prediction rules/Primary studies/Reporting data, ambos con búsqueda el 02 de julio de 2021 y en LILACS, el 5 de julio de 2021. Se utilizaron los términos de búsqueda: NEWS2, ISARIC, 4C.
- Los criterios de selección de los estudios fueron: estudios de validación externa con diseño cohorte, con al menos 100 eventos para los desenlaces de interés y estudios de impacto

incluyendo ensayos clínicos aleatorios, cohorte ó estudios transversales que evalúen impacto en la toma de decisiones.

- Se consideró la evaluación de riesgo de sesgo para los estudios seleccionados a partir de la revisión sistemática identificada y de no estar disponible, ésta se realizó con la herramienta PROBAST(6) para estudios de modelos pronósticos.
- La certeza de la evidencia fue realizada según el enfoque GRADE que toma en cuenta los siguientes criterios: diseño del estudio, riesgo de sesgo, inconsistencia en los resultados, ausencia de evidencia directa, imprecisión, sesgo de publicación, tamaño de efecto, gradiente dosis-respuesta, y efecto de los potenciales factores de confusión residual (los tres últimos aplicables en estudios observacionales) (7,8). De acuerdo a estos criterios la evidencia se puede clasificar en 4 niveles:

Certeza o calidad de la evidencia	Interpretación
Alta ⊕ ⊕ ⊕ ⊕	Es muy poco probable que nuevos estudios cambien la confianza/certeza que se tiene en el resultado estimado.
Moderada ⊕ ⊕ ⊕ ○	Es probable que nuevos estudios tengan un impacto importante en la confianza que se tiene en el resultado estimado y que estos puedan modificar este resultado.
Baja ⊕ ⊕ ○ ○	Es muy probable que nuevos estudios tengan un impacto importante en la confianza que se tiene en el resultado estimado y es probable que puedan modificar el resultado.
Muy baja ⊕ ○ ○ ○	Cualquier resultado estimado es muy incierto.

- Los resultados fueron presentados utilizando la Tablas de Resumen de Hallazgos (SOF, por sus siglas en inglés) construidas a partir del software en línea GRADEpro (<https://grade.pro.org/>) (9)

c) Metodología considerada para la elaboración de las recomendaciones: Marco de Evidencia a la Decisión/Recomendación (EtD: Evidence to decisión framework)

- Los marcos EtD(1,2) son una herramienta del Enfoque GRADE, que tiene como finalidad fomentar el uso de la evidencia de una manera estructurada y transparente para informar decisiones relacionadas al manejo clínico de una enfermedad, salud pública, políticas del sistema de salud o en situaciones como el contexto actual de pandemia acerca de acciones con repercusión socio-económica entre otras.
- Los siguientes criterios del marco EtD fueron seleccionados para la discusión y juicio por el grupo de trabajo: Efectos deseables, Efectos indeseables, Certeza de la evidencia, Valores y preferencias de los pacientes, Balance de efectos, Recursos necesarios, Equidad, Aceptabilidad y Factibilidad. En caso de no haber consenso en la valoración del juicio, se efectuó una votación, determinándose la valoración por mayoría simple. La perspectiva fue del sistema de salud.
- Elaboración de las Recomendaciones: La metodología EtD considera determinar la fuerza y dirección de una recomendación(10). Ambas, como resultado del juicio acerca del balance beneficio-riesgo, calidad global de la evidencia, confianza en los valores y preferencias de los pacientes, uso de recursos, equidad en salud, aceptabilidad, y factibilidad. Se efectuó la valoración de estos criterios para cada escala o modelo pronóstico versus no usarla, en vista de la ausencia de estudios de impacto que comparen entre ellas. En este sentido, existirán recomendaciones “a favor de la intervención” o “en contra de la intervención” (a favor de la alternativa u opción). Asimismo, las recomendaciones pueden ser fuertes o condicionales.
 - Una recomendación será fuerte si existe una clara diferencia entre los efectos deseables e indeseables de la intervención, la certeza global de la evidencia alta o moderada, todos o casi todos los pacientes informados toman la misma decisión, el costo de la intervención está plenamente justificado, existe un impacto favorable en la equidad en salud, la intervención es aceptable para los usuarios interesados (pacientes y personal de salud) y la implementación de dicha intervención es viable.

- Una recomendación será condicional si alguna de las consideraciones siguientes está presente: Exista poca diferencia entre los efectos deseables e indeseables de la intervención, la calidad de la evidencia es baja o muy baja, existe variabilidad o incertidumbre respecto de lo que decidirán los pacientes informados o el costo de la intervención pudiera no estar justificado en algunas circunstancias.
- Las implicancias de la fuerza de la recomendación son las siguientes:

Recomendación fuerte	La mayoría de los pacientes deberían recibir la intervención recomendada.
Recomendación condicional	Si bien la intervención o alternativa recomendada es apropiada para la mayoría de los pacientes, puede no resultar aplicable a todos ellos. La decisión debiera ser individualizada considerando además un enfoque de decisiones compartidas con el paciente.

d) Diálogo Deliberativo para la valoración de los criterios del Marco EtD y elaboración de las recomendaciones

El Diálogo deliberativo se llevó a cabo el día 09 de julio de 2020, reunión virtual a través de la herramienta Zoom, con la participación de:

- 1) Profesionales del Grupo de trabajo designado por el Ministerio de Salud y Representantes del Ministerio de Salud, en su calidad de panel de expertos, habilitados para emitir los juicios para cada criterio, votar en caso de ser necesario y elaborar la recomendación.
- 2) Representantes de la Unidad de Análisis y Generación de Evidencias en Salud Pública (UNAGESP) del INS, quienes efectuaron la identificación de la evidencia presentada ante los expertos, en calidad de facilitadores y conductores de los aspectos metodológicos de la reunión.

Los participantes fueron los siguientes:

Grupo de trabajo designado por MINSAs	Representantes del MINSAs e INS	Facilitadores (Metodólogos)
Dos representantes de la sociedad Peruana de Medicina Intensiva (SOPEMI) Un representante de la Sociedad Peruana de Neumología.	Representante del DVMSP Representante de la DGIESP	Diana Gonzales Fabiola Huaroto Catherine Bonilla José Carlos Sánchez Lucía Villar

Luego de la discusión y la valoración de los criterios del marco EtD, finalmente se elaboró una recomendación.

RECOMENDACIÓN

1. En pacientes hospitalizados con COVID-19, se sugiere el uso de la escala NEWS2 en adición al criterio clínico, para apoyar la toma de decisión de ingreso a un área crítica

(Recomendación condicional a favor de la intervención, basada en calidad de evidencia Muy Baja)

Consideraciones adicionales:

- Considerar la aplicación de la escala en condiciones de sobre demanda alta a extrema de los servicios de salud y dependiendo de la experticia de los profesionales de salud a cargo de la atención del paciente.

JUSTIFICACIÓN DE LA RECOMENDACIÓN

La Tabla de resumen de los juicios es la siguiente:

Escala de mortalidad 4C		Modelo de deterioro 4C		Escala NEWS2	
Beneficios	Pequeño	Beneficios	Moderado	Beneficios	Moderado
Daños	Pequeño	Daños	Varía	Daños	Pequeño
Certeza de la evidencia	Muy baja	Certeza de la evidencia	Muy baja	Certeza de la evidencia	Muy baja
Valores y preferencias	Posiblemente hay incertidumbre y variabilidad importantes	Valores y preferencias	Posiblemente hay incertidumbre y variabilidad importantes	Valores y preferencias	Posiblemente hay incertidumbre y variabilidad importantes
Balance de efectos	No favorece ni al uso ni al no uso	Balance de efectos	Se desconoce	Balance de efectos	Probablemente favorece al uso
Uso de recursos	Costos y ahorros insignificantes	Uso de recursos	Costos moderados	Uso de recursos	Ahorros
Equidad	Probablemente incrementa la equidad	Equidad	Probablemente incrementa la equidad	Equidad	Probablemente incrementa la equidad
Aceptabilidad	Probablemente Si	Aceptabilidad	Probablemente No	Aceptabilidad	Probablemente Si
Factibilidad	Probablemente Si	Factibilidad	Probablemente No	Factibilidad	Probablemente Si

Resumen de la evidencia identificada:

- Respecto a la escala de mortalidad 4C se seleccionaron 3 estudios, el primero (Knight et al.(11)) correspondió al reporte de la derivación (N= 35463) y validación interna de la escala (N=22361), a partir de una cohorte de pacientes hospitalizados con COVID-19 en 260 hospitales de Reino Unido, para el periodo del 6 de febrero al 29 de junio de 2020, con mediana de edad entre 73 a 76 años y con una tasa de mortalidad entre 30% a 32%. El modelo final incluyó 8 predictores: Edad; sexo; No de comorbilidades, FR, SO₂ en aire ambiental, Escala de coma de Glasgow, urea, Proteína C-reactiva y el desenlace a pronosticar fue mortalidad intrahospitalaria. Las variables continuas del modelo fueron categorizadas y los autores construyeron la escala de mortalidad considerando 4 grupos de riesgo. La validación interna mostró una buena capacidad predictiva para la escala: AUC: 0.767 (IC 95%: 0.760-0.773) y calibración adecuada. El riesgo de sesgo fue valorado por Wynants et al.(12) como No es claro (dominio Análisis, de la herramienta PROBAST) debido a que los predictores continuos se dicotomizaron, y que si bien el modelo se convirtió en una puntuación total, no estaba claro cómo se tradujeron las puntuaciones a los riesgos de mortalidad previstos que se utilizaron para evaluar la calibración. Se identificaron dos estudios de validación externa con >100 eventos, realizados en Francia y Canadá que corroboraron la capacidad de discriminación de la escala para ambos entornos, aunque no se realizó la evaluación de la calibración(13,14). No se identificaron estudios de validación externa en Perú o en otro país de Latinoamérica. No se identificaron estudios que hayan valorado el impacto clínico del uso de esta escala.
- El modelo de deterioro 4C(15) fue desarrollado y validado (validación interna) en Reino Unido, a partir de la misma cohorte de pacientes que la escala de mortalidad 4C, pero considerando un mayor tamaño muestral y periodo de enrolamiento: Del 6 de febrero al 26 de agosto de 2020. El modelo final incluyó 11 predictores (medidos en el momento de la hospitalización): Edad; sexo, infección intrahospitalaria por COVID-19, infiltrado pulmonar, con soporte oxigenatorio, FR, SO₂ al ingreso, Escala de coma de Glasgow, urea, Proteína C-reactiva, recuento absoluto de linfocitos. El desenlace a predecir fue deterioro clínico intrahospitalario, definido como inicio de soporte ventilatorio (ventilación invasiva o no invasiva), admisión a UCI y muerte. La mediana de edad fue de 75 años, y el evento se presentó en el 43% de los

pacientes. El modelo mostró una capacidad de discriminación y calibración aceptable, con una AUC de 0.77 (IC 95%: 0.76-0.78). Los autores recomendaron la aplicación conjunta de la escala de mortalidad 4C y del modelo de deterioro 4C, debido a la discrodancia entre ambas en pacientes más jóvenes y que la capacidad de discriminación del modelo de deterioro disminuye a mayor tiempo para el horizonte temporal y se incrementa en el caso de la escala de mortalidad. Una calculadora en línea está disponible en la web <https://isaric4c.net/risk/>. El riesgo de sesgo fue valorado como Bajo. No se identificaron estudios de validación externa. No se identificaron estudios que hayan valorado el impacto clínico del uso de este modelo.

- La escala NEWS2 fue originalmente diseñada por The Royal College of Physicians para predecir la mortalidad del paciente hospitalizado dentro de las 24 horas siguientes a un conjunto de observaciones de parámetros clínicos. Evalúa 7 predictores: Frecuencia respiratoria, la saturación de oxígeno, el requerimiento de oxígeno suplementario, la frecuencia cardíaca, presión arterial, nivel de conciencia y la temperatura. Está recomendada para la evaluación inicial del paciente hospitalizado y la monitorización seriada(16). Respecto a su uso en pacientes con COVID-19, los autores recomiendan que en pacientes con COVID 19, se tenga además en cuenta el aumento en el requerimiento de oxígeno, para desencadenar una alerta y monitorización continua de al menos cada hora, ya que la escala no contempla este aumento como parámetro para la puntuación(17). Se identificaron 6 estudios de validación externa con >100 eventos para alguno de los desenlaces de interés que incluyeron mortalidad a las 72 horas, deterioro clínico o COVID-19 severo a las 24h o 3 días y deterioro clínico o ingreso a UCI a los 14 días o 30 días(11,15,18–21). Las definiciones para desenlaces que valoran progresión o deterioro clínico variaron entre los estudios. Los estudios fueron realizados con datos de Reino Unido, Noruega, China, donde incluyeron a pacientes entre 58 a 80 años de edad. Los resultados de la evaluación del rendimiento mostraron resultados discordantes según el desenlace y estudio. Se observó que para desenlaces con horizonte temporal cercano (1 a 3 días) por ejemplo mortalidad y deterioro clínico, la capacidad de discriminación era aceptable, a diferencia de desenlaces con mayor tiempo de horizonte temporal de predicción, donde la AUC obtenida fue < de 0.70. El riesgo de sesgo global fue calificado como alto riesgo debido a exclusiones del análisis debido a datos faltantes,

Riesgo de sesgo global utilizando PROBAST					
Estudio	Participantes	Predictores	Desenlace	Análisis	Global
Bradley2020(18)	Bajo	Bajo	Bajo	Alto ¹	Alto
Baker 2021(19)	No es claro ²	Bajo	Bajo	Alto ³	Alto
Gupta 2020(21)	Bajo	Bajo	Bajo	Bajo	Bajo
Carr 2021(20)	Bajo	Bajo	Bajo	No es claro ⁴	No es claro
Gupta 2021(15)	Bajo	Bajo	Bajo	No es claro ⁵	No es claro
Knight 2020(11)	Bajo	Bajo	Bajo	No es claro ⁵	No es claro

1. Se excluyeron 100 participantes del análisis debido a datos ausentes, no manejo adecuado de datos faltantes, no se reportan los IC 95% para los índices C, eventos <100 para el desenlace mortalidad a las 72h

2. Se incluyeron reingresos dentro de los 28 días de la hospitalización inicial. No es claro cómo fueron manejados estos datos para el análisis, respecto al rendimiento de NEWS2 considerando el puntaje al ingreso

3. No se evaluó la calibración como parte del rendimiento. Se señala que se excluyeron participantes que no tuvieran los valores para el cálculo de los 3 scores evaluados, pero no hay información de la cantidad de excluidos.

4. No se reporta la calibración para el desenlace a los 14 días para la escala NEWS2

5. El análisis no fue efectuado en la totalidad de la cohorte de validación y se desconoce si el número de eventos analizado fue <100

Certeza de la evidencia: la certeza de la evidencia para todas las intervenciones fue valorada como **Muy Baja**, debido a la ausencia de estudios que evalúen el impacto clínico de las escalas o modelos pronóstico (ausencia de ensayos clínicos que comparen el uso vs no uso frente a desenlaces clínicos), por lo que todos los estudios iniciaron con valoración de certeza Baja y la confianza disminuyó debido a evidencia indirecta: ausencia de estudios de validación externa en entornos que puedan ser aplicables al contexto nacional y a riesgo de sesgo para los estudios que evaluaron la escala NEWS2. El juicio fue determinado por consenso.

Valores y preferencias de los pacientes: No se recopiló estudios sobre los valores y preferencias en pacientes con COVID-19. El grupo de trabajo consideró que **es probable que haya incertidumbre o variabilidad importante** en los valores y preferencias de los pacientes, considerando las limitaciones de la evidencia identificada para estas escalas o modelos, y que el ingreso a UCI, se ha vuelto difícil de alcanzar y los pacientes están al tanto de como se toma la decisión, por lo que existe incertidumbre en si optarían o no para que esos modelos se apliquen para determinar la atención que se les brinda. El juicio fue determinado por consenso.

Balance de los efectos: Para valorar este criterio el grupo de trabajo consideró que la evidencia del desempeño de las escalas 4C procede de población de edad avanzada, en un periodo previo a la vacunación y que podría no representar el contexto actual de la población que acudiría para hospitalización, añadiéndose, el impacto de las variantes del SARS-CoV-2 existentes y afectación de población adulta más joven o con menor severidad al ingreso, en comparación a la cohorte de las escalas del grupo ISARIC4C. A ello, se suma la falta de estudios de validación externa de la escala de deterioro 4C. Por otro lado, se tiene experiencia de uso clínico previo para la escala NEWS2, con validación para predecir pronóstico en pacientes con deterioro agudo, en población general y su capacidad discriminatoria mejora para el pronóstico a corto plazo, aunque resulta necesario la realización de estudios de validación locales. Tomando en cuenta los potenciales beneficios y daños, el grupo de trabajo valoró que el balance de efectos **probablemente favorece al uso de la escala NEWS2**; se desconoce para el modelo de deterioro 4C y No favorece ni al uso ni al no uso de la escala de mortalidad 4C. Los juicios fueron determinados por mayoría simple, luego de una votación.

Este juicio fue emitido para el uso de la escala NEWS2 en adición al criterio clínico.

Uso de recursos: No se recopilaron estudios económicos en COVID-19, para valorar este criterio. El grupo de trabajo consideró que únicamente **el uso de la escala NEWS2 supondría ahorros** aunque no hubo acuerdo respecto a la magnitud de este ahorro. Para las otras dos intervenciones, su uso implicaría una mayor inversión en la capacitación del personal de salud, además de considerar las pruebas de laboratorio requeridas entre las que se incluyen: proteína C reactiva, que podría no estar disponibles de manera oportuna para la aplicación de la escala de mortalidad y modelo de deterioro 4C y con ello retrasar el ingreso a UCI, además de la necesidad de acceder a una página web para el cálculo, principalmente de el modelo de deterioro 4C. Para la escala NEWS2 hay experiencia de uso clínico previo e involucra parámetros fisiológicos y de estado de conciencia que forman parte de la evaluación clínica. Los juicios fueron determinados por mayoría simple, luego de una votación.

Equidad en salud: Por consenso, el grupo de trabajo valoró que el uso de cualquiera de las escalas o modelo pronóstico probablemente incrementa la equidad en salud.

Aceptabilidad: Para este criterio, el grupo de trabajo tomó en cuenta el número de parámetros o variables predictoras de cada modelo, el tiempo que tomaría obtener la puntuación (incluye resultados a tiempo para pruebas de laboratorio), más aún en momentos de sobredemanda del sistema de salud, la necesidad de acceder a páginas web para su cálculo (puntos de red) y la disponibilidad de los resultados de las pruebas de laboratorio para aplicar una escala, si ésta lo amerita. En función a ello, la aplicación de una escala o modelo podría ser aceptado o no por los profesionales de salud. Se identificaron limitaciones para el modelo de deterioro 4C debido a tener el mayor número de predictores y el requisito obligatorio de acceder a la calculadora en línea para la determinación. El grupo de trabajo consideró que el uso de la escala de mortalidad y la escala NEWS2 probablemente si sea aceptable por los profesionales de salud y pacientes.

Factibilidad: El panel consideró que la aplicación de la escala NEWS2 y la escala de mortalidad 4C eran factibles de implementar en pacientes hospitalizados. Sin embargo, el uso de la escala de deterioro clínico se consideró no factible, considerando las barreras señaladas en el criterio previo.

Al revisar los juicios emitidos para las 3 intervenciones, el grupo de trabajo emitió una recomendación a favor del uso de la escala NEWS2 en pacientes hospitalizados, con COVID-19, sustentado en la valoración obtenida para el balance de efectos, uso de recursos, equidad en salud, aceptabilidad y factibilidad.

El grupo de trabajo también identificó:

- Necesidades de investigación: estudios de validación de la escala NEWS2 para pronóstico de pacientes hospitalizados con COVID-19 locales, así como estudios que evalúen el impacto de su uso clínico.
- Consideraciones de implementación: La capacitación de los profesionales de salud respecto a la aplicación de la escala. El acceso a la herramienta así como la monitorización de su aplicación: seguimiento de cómo se llena, control de calidad del registro de información.

REFERENCIAS:

1. Alonso-Coello P, Oxman AD, Moberg J, Brignardello-Petersen R, Akl EA, Davoli M, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 2: Clinical practice guidelines. *BMJ*. 30 de 2016;353:i2089.
2. Alonso-Coello P, Schünemann HJ, Moberg J, Brignardello-Petersen R, Akl EA, Davoli M, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction. *BMJ*. 28 de 2016;353:i2016.
3. Organización Mundial de la Salud. Manejo clínico de la COVID-19: orientaciones evolutivas, 25 de enero de 2021 [Internet]. Ginebra: Organización Mundial de la Salud; 2021. Disponible en: <https://apps.who.int/iris/handle/10665/340629>
4. Map of COVID-19 evidence [Internet]. Norwegian Institute of Public Health. [citado el 12 de julio de 2021]. Disponible en: <https://www.fhi.no/en/qk/systematic-reviews-hta/map/>
5. Living Overview of the Evidence (L-OVE) [Internet]. [citado el 12 de julio de 2021]. Disponible en: https://app.iloveevidence.com/loves/5e6fdb9669c00e4ac072701d?question_domain=5b1dcdbde611de7ae84e8f15&population=5e7fce7e3d05156b5f5e032a&intervention=5dee1cdd19a11a02f2d7c653&classification=primary-study&reported-data=yes
6. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann Intern Med*. el 1 de enero de 2019;170(1):51–8.
7. Guyatt GH, Oxman AD, Sultan S, Glasziou P, Akl EA, Alonso-Coello P, et al. GRADE guidelines: 9. Rating up the quality of evidence. *J Clin Epidemiol*. diciembre de 2011;64(12):1311–6.
8. Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol*. abril de 2011;64(4):401–6.
9. GRADEpro GDT: GRADEpro Guideline Development Tool [Internet]. McMaster University, 2015 (developed by Evidence Prime, Inc.); 2015. Disponible en: grade.pro.org
10. Andrews JC, Schünemann HJ, Oxman AD, Pottie K, Meerpohl JJ, Coello PA, et al. GRADE guidelines: 15. Going from evidence to recommendation—determinants of a recommendation's direction and strength. *J Clin Epidemiol*. julio de 2013;66(7):726–35.

11. 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* [Internet]. el 9 de septiembre de 2020 [citado el 2 de julio de 2021];370:m3339. Disponible en: <https://www.bmj.com/content/370/bmj.m3339>
12. 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* [Internet]. el 7 de abril de 2020 [citado el 5 de julio de 2021];369:m1328. Disponible en: <https://www.bmj.com/content/369/bmj.m1328>
13. Lassau N, Ammari S, Chouzenoux E, Gortais H, Herent P, Devilder M, et al. Integrating deep learning CT-scan model, biological and clinical variables to predict severity of COVID-19 patients. *Nat Commun* [Internet]. el 27 de enero de 2021 [citado el 5 de julio de 2021];12(1):634. Disponible en: <https://www.nature.com/articles/s41467-020-20657-4>
14. External Validation of the 4C Mortality Score among COVID-19 Patients Visiting the Emergency Department or admitted to Hospital in Ontario, Canada [Internet]. 2021 [citado el 5 de julio de 2021]. Disponible en: <https://www.researchsquare.com>
15. Gupta RK, Harrison EM, Ho A, Docherty AB, Knight SR, Smeden M van, et al. Development and validation of the ISARIC 4C Deterioration model for adults hospitalised with COVID-19: a prospective cohort study. *Lancet Respir Med* [Internet]. el 1 de abril de 2021 [citado el 2 de julio de 2021];9(4):349–59. Disponible en: [https://www.thelancet.com/journals/lanres/article/PIIS2213-2600\(20\)30559-2/abstract](https://www.thelancet.com/journals/lanres/article/PIIS2213-2600(20)30559-2/abstract)
16. National Early Warning Score (NEWS) 2 [Internet]. RCP London. 2017 [citado el 12 de julio de 2021]. Disponible en: <https://www.rcplondon.ac.uk/projects/outputs/national-early-warning-score-news-2>
17. NEWS2 and deterioration in COVID-19 [Internet]. RCP London. 2020 [citado el 6 de julio de 2021]. Disponible en: <https://www.rcplondon.ac.uk/news/news2-and-deterioration-covid-19>
18. Bradley P, Frost F, Tharmaratnam K, Wootton DG. Utility of established prognostic scores in COVID-19 hospital admissions: multicentre prospective evaluation of CURB-65, NEWS2 and qSOFA. *BMJ Open Respir Res* [Internet]. el 1 de diciembre de 2020 [citado el 2 de julio de 2021];7(1):e000729. Disponible en: <https://bmjopenrespres.bmj.com/content/7/1/e000729>
19. Baker KF, Hanrath AT, Loeff IS van der, Kay LJ, Back J, Duncan CJ. National Early Warning Score 2 (NEWS2) to identify inpatient COVID-19 deterioration: a retrospective analysis. *Clin Med* [Internet]. el 1 de marzo de 2021 [citado el 2 de julio de 2021];21(2):84–9. Disponible en: <https://www.rcpjournals.org/content/clinmedicine/21/2/84>
20. Carr E, Bendayan R, Bean D, Stammers M, Wang W, Zhang H, et al. Evaluation and improvement of the National Early Warning Score (NEWS2) for COVID-19: a multi-hospital study. *BMC Med* [Internet]. el 21 de enero de 2021 [citado el 2 de julio de 2021];19(1):23. Disponible en: <https://doi.org/10.1186/s12916-020-01893-3>
21. 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* [Internet]. el 1 de diciembre de 2020 [citado el 6 de julio de 2021];56(6). Disponible en: <https://erj.ersjournals.com/content/56/6/2003498>

Elaborado por

Fabiola Huaroto¹

Revisado por:

Nora Reyes Puma¹

Diana Gonzales Zurita¹

¹ Unidad de Análisis y Generación de Evidencias en Salud Pública (UNAGESP), Centro Nacional de Salud Pública, Instituto Nacional de Salud.

Repositorio general de documentos técnicos UNAGESP:

<https://web.ins.gob.pe/es/salud-publica/publicaciones-unagesp>

NOTA: El Instituto Nacional de Salud es un Organismo Público Ejecutor del Ministerio de Salud del Perú dedicado a la investigación de los problemas prioritarios de salud y de desarrollo tecnológico. El Instituto Nacional de Salud tiene como mandato el proponer políticas y normas, promover, desarrollar y difundir la investigación científica-tecnológica y brindar servicios de salud en los campos de salud pública, control de enfermedades transmisibles y no transmisibles, alimentación y nutrición, producción de biológicos, control de calidad de alimentos, productos farmacéuticos y afines, salud ocupacional, protección del medio ambiente y salud intercultural, para contribuir a mejorar la calidad de vida de la población. A través de su Unidad de Análisis y Generación de Evidencias en Salud Pública (UNAGESP) participa en el proceso de elaboración de documentos técnicos, basados en la mejor evidencia disponible, que sirvan como sustento para la aplicación de intervenciones en Salud Pública, la determinación de Políticas Públicas Sanitarias y la Evaluación de Tecnologías Sanitarias.

ANEXO 01: Tablas de Resumen de hallazgos

Resumen de los resultados :

Escala de mortalidad 4C comparado con No usarla en pacientes con COVID-19 hospitalizados

Paciente o población : pacientes con COVID-19 hospitalizados

Intervención: Escala de mortalidad 4C

Comparación: No usarla

Desenlaces	Impacto	Nº de participantes (estudios)	Certeza de la evidencia (GRADE)
Mortalidad Intrahospitalaria	AUC: 0.77 (IC 95%: 0.76-0.77) Buena calibración	(1 estudio observacional)	⊕○○○ MUY BAJA ^a

Explicaciones

a. No se evaluado el impacto del uso vs no uso en los desenlaces clínicos. No hay estudios de validación externa en Perú o Latinoamérica, o entorno similar a partir del cual se pueda generalizar los resultados.

Resumen de los resultados :

Modelo de deterioro 4C comparado con No usarlo en pacientes con COVID-19 hospitalizados

Paciente o población : pacientes con COVID-19 hospitalizados

Intervención: Modelo de deterioro 4C

Comparación: No usarlo

Desenlaces	Impacto	Nº de participantes (estudios)	Certeza de la evidencia (GRADE)
Deterioro clínico (intrahospitalario)	AUC de 0.77 (IC 95%: 0.76-0.78) Calibración buena	(1 estudio observacional)	⊕○○○ MUY BAJA ^a

Explicaciones

a. No se evaluado el impacto del uso vs no uso en los desenlaces clínicos. No hay estudios de validación externa

Resumen de los resultados :

Escala NEWS2 comparado con No usarla en pacientes con COVID-19 hospitalizados

Paciente o población : pacientes con COVID-19 hospitalizados

Intervención: Escala NEWS2

Comparación: No usarla

Desenlaces	Impacto	Nº de participantes (estudios)	Certeza de la evidencia (GRADE)
Mortalidad (72h)	<u>Mortalidad 72 horas</u> Bradley 2020: AUC de 0.78 Calibración no reportada	(1 estudio observacional)	⊕○○○ MUY BAJA a,b
Mortalidad (intrahospitalaria o 30 días)	<ul style="list-style-type: none"> <u>Mortalidad intrahospitalaria</u> Knight 2020: AUC de 0.65 (IC 95%: 0.65-0.66) <u>Mortalidad 30 días</u> : Bradley 2020: AUC de 0.67, calibración pobre 	(2 estudios observacionales)	⊕○○○ MUY BAJA b,c
Deterioro clínico / COVID-19 severo (Deterioro clínico) evaluado con : 16 3 días	<ul style="list-style-type: none"> <u>Deterioro clínico (1 día)</u> Gupta 2020: AUC de 0.78 (0.73, 0.83) Calibración: Moderada Gupta 2021 (4C): AUC de 0.69 (0.68-0.70) <u>COVID-19 severo (3 días)</u> Carr 2021: AUC de 0.687 a 0.772 Calibración Pobre 	(3 estudios observacionales)	⊕○○○ MUY BAJA b,d
Deterioro clínico /COVID-19 severo / Ingreso a UCI evaluado con : Intrahospitalario, 14 o 30 días	<ul style="list-style-type: none"> <u>COVID-19 severo (14 días)</u> Car 2021: AUC entre 0.697 a 0.759 Calibración No reportada <u>Ingreso a UCI (30 días)</u> Bradley 2020: AUC de 0.65 Calibración No reportada <u>Deterioro clínico (Intrahospitalario)</u> Baker 2021: AUC de 0.71 (0.65–0.77)Calibración No evaluada 	(3 estudios observacionales)	⊕○○○ MUY BAJA b,e

Explicaciones

- Se excluyeron 100 participantes del análisis debido a datos ausentes, no manejo adecuado de datos faltantes, no se reportan el IC 95% para la AUC, eventos <100 para el desenlace mortalidad a las 72h. La calibración no fue reportada
- No se han comparado uso vs no uso del score en los desenlaces clínicos. Se espera que el entorno de donde se toman los datos puede ser distinto al local, no incluye datos de Latinoamérica en la validación
- En uno de los estudios se excluyeron 100 participantes del análisis debido a datos ausentes, no manejo adecuado de datos faltantes, no se reportan los IC 95% para los índices C, La calibración fue pobre o no reportada
- El rendimiento, respecto a la determinación varió y sólo en 1 estudio fue buena. La calibración fue pobre a moderada entre los estudios
- (Baker 2020): se incluyeron reingresos y no es claro cómo fueron manejados estos datos para el análisis, no se evaluó la calibración como parte del rendimiento. Se señala la exclusión de participantes debido a datos faltantes, pero no hay información de la cantidad de excluidos. En otro estudio (Bradley 2020): Se excluyeron 100 participantes del análisis debido a datos ausentes, no manejo adecuado; no se reportan los IC 95% para los índices C. No se dispone de resultados de la calibración

Grados de evidencia del GRADE Working Group

Alta certeza: Estamos muy seguros de que el verdadero efecto se acerca al de la estimación del efecto

Certeza moderada: Tenemos una confianza moderada en la estimación del efecto: es probable que el efecto real esté cerca de la estimación del efecto, pero existe la posibilidad de que sea sustancialmente diferente

Certeza baja: Nuestra confianza en la estimación del efecto es limitada: el efecto real puede ser sustancialmente diferente de la estimación del efecto.

Certeza muy baja: Tenemos muy poca confianza en la estimación del efecto: Es probable que el efecto real sea sustancialmente diferente de la estimación del efecto