



Validez de apariencia, fiabilidad y consistencia interna de un modelo predictivo de mortalidad en pacientes con enfermedad renal crónica

Apparent validity, reliability and internal consistency of a predictive model of mortality in patients with end-stage renal disease

Sergio Orlando Escalona González^{1*} <https://orcid.org/0000-0003-4261-6842>

Zoraida Caridad González Milán^{1,2} <https://orcid.org/0000-0002-4092-9389>

Yailé Caballero Mota³ <https://orcid.org/0000-0002-6725-5812>

Yanela Rodríguez Álvarez³ <https://orcid.org/0000-0003-2232-2265>

Diamela Hernández Navarro¹ <https://orcid.org/0009-0006-0343-3663>

Andrés Villar Bahamonde⁴ <https://orcid.org/0009-0002-9653-2169>

Raymar Molina Vega^{1,2} <https://orcid.org/0000-0002-6235-3560>

¹Universidad de Ciencias Médicas de Las Tunas. Facultad de Ciencias Médicas "Dr. Zoilo Enrique Marinello Vidaurreta". Las Tunas, Cuba.

²Hospital General Docente "Dr. Ernesto Guevara de la Serna". Las Tunas, Cuba.

³Universidad de Camagüey. Camagüey, Cuba.

⁴Universidad de Ciencias Médicas de Pinar del Río. Pinar del Río, Cuba.

*Autor para la correspondencia. Correo electrónico: sernephron96@gmail.com

RESUMEN

Introducción: A pesar de que existen varios modelos predictivos de mortalidad en la enfermedad renal crónica, la evidencia de métricas de validez, fiabilidad y consistencia interna es limitada, heterogénea y afecta la aplicabilidad y generalización.





Objetivo: Evaluar la validez de apariencia, fiabilidad y consistencia interna de un modelo predictivo de mortalidad en la enfermedad renal crónica.

Métodos: Se realizó un estudio descriptivo, de validación de apariencia, fiabilidad y consistencia interna. Se tomó una muestra por intencionalidad que incluyó a 10 expertos a los que se le aplicó un cuestionario sobre un modelo predictivo de mortalidad en pacientes con enfermedad renal crónica, mediante el método de consenso de expertos descrito por Delphi. Se determinó el coeficiente alfa de Cronbach y se calculó para cada uno de los predictores del modelo predictivo. Se utilizó el método de división por mitades de Spearman-Brown para el análisis de la consistencia interna.

Resultados: Los expertos evidenciaron un coeficiente $K > 0,8$ respecto a la predicción de mortalidad en la enfermedad renal crónica. La fiabilidad fue alta ($\alpha = 0,824$), lo que indica una reproducibilidad sobresaliente en las estimaciones. Si los predictores: enfermedad cardiovascular, albúmina < 30 g/L y sepsis son eliminados del modelo predictivo, disminuye el resultado en el alfa global. El coeficiente de Spearman-Brown = 0,842 traduce que existe adecuada correlación y coherencia entre los predictores para medir el constructo.

Conclusiones: El modelo predictivo demostró una sólida validez aparente, alta fiabilidad y adecuada consistencia interna, por lo que puede usarse como herramienta de apoyo clínico.

Palabras clave: consistencia interna; enfermedad renal crónica; fiabilidad; modelo predictivo; validación.

ABSTRACT

Introduction: Although several predictive models of mortality in chronic kidney disease exist, the evidence for validity, reliability and internal consistency metrics is limited and heterogeneous, affecting applicability and generalizability.

Objective: To evaluate apparent validity, reliability and internal consistency of a predictive model of mortality in chronic kidney disease.

Methods: A descriptive study was conducted to validate the criteria, reliability and internal consistency. A purposive sample of 10 experts was administered a questionnaire about a predictive



model of mortality in patients with chronic kidney disease, using the expert consensus method described by Delphi. The overall Cronbach's alpha coefficient was determined and it was calculated for each of the predictors of the predictive model. The Spearman-Brown split-half method was used for the internal consistency analysis.

Results: Experts demonstrated a K coefficient > 0.8 for predicting mortality in chronic kidney disease. Reliability was high ($\alpha = 0.824$), indicating outstanding reproducibility of the estimates. If the predictors cardiovascular disease, albumin < 30 g/L and sepsis are removed from the predictive model, the overall alpha score decreases. The Spearman-Brown coefficient = 0.842 indicates adequate correlation and consistency among the predictors for measuring the construct.

Conclusions: The predictive model demonstrated strong face validity, high reliability and adequate internal consistency and can therefore be used as a clinical support tool.

Keywords: chronic kidney disease; internal consistency; predictive model; reliability; validation.

Recibido: 24/02/2026

Aprobado: 19/05/2026

INTRODUCCIÓN

La enfermedad renal crónica (ERC) representa un gran reto para los sistemas de salud pública a nivel global, con una prevalencia que ha aumentado significativamente en los últimos años. Se estima que aproximadamente entre el 10 % y 15 % de la población mundial padece la enfermedad. Se espera que para el año 2040, la ERC ocupe la quinta causa de años de vida perdidos y que la tasa bruta de pacientes con necesidad de terapias de reemplazo renal (TRR) supere los 350 pacientes por millón de habitantes. También reviste especial importancia aquellos casos que progresan hacia fases avanzadas de forma silenciosa y mueren prematuramente por complicaciones asociadas.⁽¹⁾



El abordaje de la ERC impone cargas económicas considerables tanto a países desarrollados como a aquellos en vías de desarrollo. El gasto público asociado a ERC está en aumento considerable y representa un mayor porcentaje del presupuesto sanitario total, sobre todo los costos destinados a las TRR y al tratamiento de complicaciones y comorbilidades que influyen directamente en la productividad, discapacidad e impacto socioeconómico en pacientes y familiares.⁽²⁾

Los modelos predictivos (MP) tradicionales implementados en la estratificación del riesgo de mortalidad en pacientes con ERC en HD han evolucionado a medida que se perfeccionaron las técnicas de análisis de datos, la metodología y el conocimiento sobre la fisiopatología del daño renal. Existen varios MP; pero la mayoría presentan limitaciones en términos de interpretabilidad, aplicabilidad, reproducibilidad, generalización, discriminación, calibración y si han sido validados de forma confiable. Varios de estos modelos subestiman la variabilidad entre poblaciones de desarrollo y validación, lo que ocasiona pérdida de calibración al aplicarlos en otros contextos.⁽³⁾

La validación de los MP es un componente indispensable que asegura si una herramienta predictiva es fiable, si se puede generalizar y si es de utilidad en la práctica clínica. La robustez de un MP, que ha sido medido por el rendimiento en una muestra de entrenamiento, puede mostrar tendencias al sobreajuste y falsa bondad de criterio. Por lo que en ausencia de pruebas de validación rigurosas, los MP pueden inducir a errores en predicción con repercusión en la toma de decisiones sistemáticas.⁽⁴⁾

La validez de apariencia es una evaluación preliminar y de carácter no estadístico sobre la percepción superficial o la apariencia de una herramienta de medición, en este caso un MP. Además, informa si las variables incluidas en el instrumento miden de forma aparente o lógica el constructo que se pretende medir. El objetivo de esta forma de validez es evaluar si para un grupo de jueces expertos, los ítems que componen al MP son relevantes para el dominio de interés.⁽⁵⁾ Generalmente, la información obtenida de la validez de apariencia, se utiliza para descartar ítems de menor importancia, para refinar la eficacia de herramientas pronósticas.

La fiabilidad o pruebas de reproducibilidad hacen referencia a si una herramienta funciona de forma similar bajo diferentes condiciones, dependientes del mismo instrumento, del tiempo de aplicación y del personal clínico que hace la medición.⁽⁶⁾ En la predicción de mortalidad en la ERC, se necesita



un MP fiable que genere respuestas consistentes y estables, características que son esenciales para la credibilidad clínica y posterior generalización.

Por otra parte, la consistencia interna de un MP mide el grado de correlación y coherencia entre los predictores, es decir, si hay homogeneidad entre ellos.⁽⁷⁾ En el caso de los MP, la consistencia interna verifica que la combinación de variables incluidas contribuya con homogeneidad a la predicción de la variable dependiente en estudio y no introduzca ruido o redundancias que puedan generar sesgos en las estimaciones.

Teniendo en cuenta las inconsistencias metodológicas identificadas en la literatura, que podrían afectar la generalización y reproducibilidad de un MP, se decidió realizar la presente investigación con el objetivo de evaluar la validez de apariencia, fiabilidad y consistencia interna de un modelo predictivo de mortalidad en la enfermedad renal crónica.

MÉTODOS

Contexto y clasificación de la investigación

Se realizó un estudio descriptivo, de validación de apariencia, fiabilidad y consistencia interna de un MP de mortalidad en la enfermedad renal crónica. El MP está basado en el diseño de una red neuronal artificial (RNA), desarrollado por *Escalona-González SO* y otros.⁽⁸⁾ Los predictores que incluyó la RNA fueron: enfermedad cardiovascular, albúmina < 30 g/L, sepsis, inicio tardío en HD, urea > 15 mmol/L, hemoglobina < 10 g/L, edad > 65 años, potasio > 5,5 mmol/L, catéter venoso central, ganancia de peso interdialítica excesiva, filtrado glomerular < 8 mL/min/1,73m² y dislipidemia.

Validez de apariencia

Para evaluar esta propiedad se empleó el método de criterio de expertos, para determinar su aplicabilidad y utilidad práctica. Se contemplaron de manera inicial 10 posibles candidatos a expertos, los cuales debían reunir tres o más de los siguientes requisitos: 1) Doctor en Ciencias en determinada área, 2) Máster en Ciencias, 3) Profesor Titular o Auxiliar, 4) Investigador Titular o





Auxiliar, 5) Especialista de Segundo Grado en Nefrología y 6) Experiencia profesional mayor de 10 años.

El método que se utilizó para determinar los expertos fue la obtención del coeficiente de competencia K; se calculó por la siguiente fórmula: $K = (K_c + K_a)/2$.⁽⁹⁾ Para ello se elaboró una encuesta en la cual se le solicitó la autoevaluación del nivel de conocimiento acerca de la predicción de mortalidad en pacientes con ERC en HD, además de completar una tabla patrón que se le aplicó a los candidatos a expertos, en la cual reflejaron el grado de influencia de los argumentos mediante los cuales habían asimilado sus conocimientos (archivo complementario).

El coeficiente de conocimiento o información (K_c) del experto acerca del tema, se calculó sobre la base de la valoración del propio experto, en una escala de 0 a 10 y multiplicado por 0,1 de modo que el valor 0 indica absoluto desconocimiento de la problemática que se evalúa y el valor 1 indica pleno conocimiento de la referida problemática. El experto marcó una cruz en la casilla que estimó pertinente.⁽⁹⁾

El coeficiente de argumentación o fundamentación (K_a) de los criterios de los expertos, se determinó como el resultado de la suma de los puntos alcanzados a partir de una tabla patrón, que se les presentó en la encuesta sin cifras, orientando que marcaran con una (x) sobre cuál de las fuentes influyó más en sus conocimientos sobre el tema, de acuerdo con los niveles establecidos. Después de calculado el coeficiente K, se consideraron los siguientes intervalos para definir la competencia del experto: nivel bajo ($K < 0,5$), nivel medio ($0,5 \leq K < 0,8$) y nivel alto ($0,8 \leq K < 1$). Se seleccionaron como expertos los 10 integrantes porque presentaron coeficiente $K > 0,8$.^(10,11)

A cada experto se le solicitó que diera su opinión en relación a los predictores incluidos en el MP y su implicación con la mortalidad. Se les pidió que evaluaran esa relación en alguna de las siguientes categorías: MA (muy adecuado), BA (bastante adecuado), A (adecuado), PA (poco adecuado) e I (inadecuado). Luego, cada una de esas respuestas cualitativas se equipararon con puntajes; MA (5 puntos), BA (4 puntos), A (3 puntos), PA (2 puntos) e I (1 punto); se promediaron las respuestas para cada predictor dado por los expertos y se redondeó al número entero superior; se volvió a equiparar el puntaje dado con las categorías cualitativas de las variables.





Se determinó la confiabilidad ítems-test según la respuesta dada por los expertos al calcular el coeficiente alfa de Cronbach.⁽¹²⁾ Además, se determinó el coeficiente de correlación intraclase (CCI).

El alfa de Cronbach expresa el grado en que los ítems del instrumento miden el mismo constructo y toma valores entre 0 y 1, aunque también puede mostrar valores negativos, lo que indicaría que en el MP hay predictores que miden lo opuesto al resto. Se consideró la confiabilidad ítems-test según el criterio de experto teniendo en cuenta el valor del coeficiente alfa de Cronbach: $< 0,5$ (no aceptable), $0,5 \leq K < 0,6$ (pobre), $0,6 \leq K < 0,7$ (débil), $0,7 \leq K < 0,8$ (aceptable), $0,8 \leq K < 0,9$ (buena) y $\geq 0,9$ (excelente).⁽¹³⁾

Fiabilidad

Se determinó el coeficiente alfa de Cronbach y se calculó para cada uno de los predictores del MP (correlación ítem-dominio). Valores por encima de 0,7 indican que los ítems del MP estiman de manera coherente el mismo constructo. El valor máximo aceptado es 0,9. Valores por encima indican que hay variables duplicadas o con redundancia.⁽⁶⁾

Consistencia interna

Se utilizó el método de división por mitades de Spearman-Brown. Valores $> 0,7$ indican que existe adecuada correlación entre los predictores.⁽¹⁴⁾

Para todos los análisis, se utilizó el *software* SPSS versión 23 para Windows.

RESULTADOS

La validez de apariencia se analizó a través de la consulta de expertos. Se definieron 10 expertos, de los cuáles el 80 % eran doctores en ciencias en determinada área y 90 % tenían maestría. El 100 % eran profesores titulares o auxiliares, el 90 % ostentaba la especialidad en segundo grado, el 80 % tenía categoría de investigador titular o auxiliar y la mediana de experiencia laboral fue de 22 años (tabla 1).


Tabla 1 – Relación de expertos según formación académica

Formación académica	No	%
Doctor en Ciencias en determinada área	8	80
Máster en Ciencias	9	90
Especialista de segundo grado en Nefrología	9	90
Profesor Titular o Auxiliar	10	100
Investigador Titular o Auxiliar	8	80

En la tabla 2 se evidencia que todos los expertos mostraron un coeficiente $K > 0,8$ luego del llenado de la encuesta que investigaba el nivel de competencia con respecto a la predicción de mortalidad en los pacientes con ERC en HD.

Tabla 2 – Coeficiente K de cada uno de los expertos

Experto	Kc	Ka	Coeficiente K	Nivel de competencia
1	1,0	1,0	1,0	Alto
2	1,0	0,9	0,95	Alto
3	0,8	1,0	0,9	Alto
4	1,0	1,0	1,0	Alto
5	0,8	0,9	0,85	Alto
6	0,8	0,8	0,8	Alto
7	0,9	1,0	0,95	Alto
8	0,8	1,0	0,9	Alto
9	0,7	1,0	0,85	Alto
10	1,0	0,8	0,9	Alto

La tabla 3 muestra que ocho de los 12 predictores incluidos en el MP se consideraron como muy adecuados. Los otros cuatro predictores se consideraron como bastante adecuados. Las variables señaladas como muy adecuadas en relación con el constructo que pretende medir el MP fueron: enfermedad cardiovascular, sepsis, urea > 15 mmol/L, hemoglobina < 10 g/L, catéter venoso central, potasio $> 5,5$ mmol/L, inicio tardío de hemodiálisis y albúmina < 30 g/L.





Tabla 3 – Resultado de la validez de apariencia de cada predictor según el criterio de los expertos

Variables	Promedio del puntaje	Evaluación
Enfermedad cardiovascular	4,9 (5)	Muy adecuado
Sepsis	4,9 (5)	Muy adecuado
Urea > 15 mmol/L	4,8 (5)	Muy adecuado
Hemoglobina < 10 g/L	4,7 (5)	Muy adecuado
Catéter venoso central	4,7 (5)	Muy adecuado
Potasio > 5,5 mmol/L	4,5 (5)	Muy adecuado
Inicio tardío de hemodiálisis	4,5 (5)	Muy adecuado
Albúmina < 30 g/L	4,5 (5)	Muy adecuado
Ganancia de peso interdialítica excesiva	4,3 (4)	Bastante adecuado
Filtrado glomerular < 8 mL/min/1,73m ²	4 (4)	Bastante adecuado
Dislipidemia	3,8 (4)	Bastante adecuado
Edad > 65 años	3,6 (4)	Bastante adecuado

El valor del coeficiente alfa de Cronbach obtenido fue de 0,824 y el estandarizado con los ítems fue de 0,829. Estos valores demuestran una buena fiabilidad. La relación entre cada uno de los predictores que conformaron el MP y su relación con el alfa de Cronbach se muestra en la tabla 4. Se aprecia que si la enfermedad cardiovascular, la albúmina < 30 g/L y la sepsis son eliminadas del modelo, disminuye el resultado en el alfa global.



Tabla 4 – Relación entre cada uno de los predictores que integraron el modelo predictivo con el α de Cronbach

Variables	Correlación total de elementos corregida	α de Cronbach si el elemento se suprime
Edad > 65 años	0,613	0,821
Enfermedad cardiovascular	0,727	0,720
Dislipidemia	0,648	0,833
Hemoglobina < 10 g/L	0,537	0,808
Albúmina < 30 g/L	0,628	0,689
Urea > 15 mmol/L	0,654	0,819
Filtrado glomerular < 8 mL/min/1,73m ²	0,616	0,822
Potasio > 5,5 mmol/L	0,714	0,808
Inicio tardío de hemodiálisis	0,602	0,823
Catéter venoso central	0,633	0,821
Ganancia de peso interdialítica excesiva	0,683	0,816
Sepsis	0,660	0,713

En la tabla 5, el coeficiente de Spearman-Brown = 0,842 traduce una adecuada correlación entre los predictores para medir el constructo, aceptable fiabilidad y consistencia interna. Existe convergencia de evidencias teniendo en cuenta la semejanza del valor con el alfa de Cronbach.

Tabla 5 – Estadísticas de fiabilidad y consistencia interna

Alfa de Cronbach	Parte 1	Valor	0,836
		N elementos	6
	Parte 2	Valor	0,830
		N elementos	6
N total de elementos			12
Correlación entre formularios			0,718
Coeficiente de Spearman-Brown		Longitud igual	0,842
		Longitud desigual	0,842
Coeficiente de dos mitades de Guttman			0,824



El análisis de fiabilidad por mitades reveló consistencia interna excelente en ambas partes del MP (mitad 1: $\alpha = 0,836$; mitad 2: $\alpha = 0,830$), con diferencia mínima entre estas, lo que muestra equilibrio. La correlación entre mitades fue alta (correlación entre formularios = 0,718). El coeficiente de Guttman de 0,824 corroboró los hallazgos anteriores. El MP mantuvo robustez métrica independientemente de la división en subconjuntos equivalentes.

Adicionalmente, se calculó el CCI bajo un modelo de efectos mixtos de dos vías con definición de consistencia. El CCI para medidas promedio fue de 0,891 (IC 95 %: 0,865-0,930), lo que evidencia una alta consistencia entre los ítems del instrumento. La prueba Hotelling's T^2 confirmó diferencias significativas entre las medias de los ítems ($T^2 = 245,73$; $p < 0,001$), lo que refleja la variabilidad esperada en la medición del constructo. La prueba F asociada al CCI fue estadísticamente significativa ($F = 6,83$; $p < 0,001$), por lo que se rechaza la hipótesis nula de ausencia de consistencia entre las medidas. Se confirmó que los ítems del MP presentaban una concordancia superior al azar en la medición del constructo.

DISCUSIÓN

Uno de los resultados más evaluados en los pacientes con ERC, tratados con HD, es la mortalidad. En este contexto, el uso de técnicas de la inteligencia artificial (IA) ha crecido con rapidez en los últimos años para estudiar y modelar las complejidades, las no linealidades y la heterogeneidad que suelen caracterizar la fisiopatología renal, cuando se utilizan enfoques estadísticos convencionales para apoyar la toma de decisiones, a menudo, de utilidad limitada.⁽¹⁵⁾ A pesar de los beneficios y garantías de diseñar un MP basado en IA, específicamente en una RNA, algoritmo con gran precisión; la evaluación de métricas que representen los diferentes tipos de validez, fiabilidad y consistencia interna es relevante.

Evangelidis N y otros⁽¹⁶⁾ realizaron una encuesta Delphi que sirvió para generar una lista priorizada basada en consenso de resultados principales para ensayos en HD, entre estos resultados estaba la mortalidad por todas las causas. En el estudio, que participaron 979 profesionales de la salud de 73





países, cinco de las variables incluidas fueron detectadas y priorizadas como las más adecuadas. Los criterios de expertos generan mayor confiabilidad de los constructos que se evalúan y perfeccionan las herramientas creadas con ese fin.

La fiabilidad y la consistencia interna son dos conceptos relacionados que se refieren a la estabilidad y consistencia de las medidas o pruebas utilizadas en una investigación. La primera se refiere a la capacidad de una medida para reproducir resultados consistentes y reproducibles, mientras que la segunda permite estimar la fiabilidad de un MP a través de un conjunto de ítems que se espera que midan el mismo constructo o una única dimensión teórica del constructo latente, o sea, aporta información sobre la homogeneidad de los ítems que conforman una herramienta.⁽¹⁷⁾ El coeficiente alfa de Cronbach es una técnica basada en el promedio de las correlaciones entre los ítems. Mediante este coeficiente se puede evaluar cuánto mejoraría o empeoraría la fiabilidad de una prueba si se excluyera un determinado ítem. El valor del alfa de Cronbach cuánto más cerca se encuentre a 1, mayor es la consistencia interna de los ítems analizados, es decir, se asume que se mide una misma dimensión.⁽¹⁸⁾

Teniendo en cuenta el valor obtenido del alfa de Cronbach se evidenció una fiabilidad aceptable. Respecto a las limitaciones de este indicador hay que tener en cuenta que lo afecta el número de los ítems, el número de alternativas de respuesta y la proporción de la varianza del test. Es por ello que se recomienda tener al menos tres ítems y cuánto mayor sea el número de ítems mayor podría ser la fiabilidad del modelo.⁽¹⁹⁾ El análisis de fiabilidad realizado reveló correlaciones ítem-total corregidas que oscilaron entre 0,537 y 0,727; superando el punto de corte de 0,30 recomendado⁽¹⁸⁾ para instrumentos de *screening* clínicos.

Las correlaciones múltiples al cuadrado, que indican la proporción de varianza del ítem explicada por los demás ítems, variaron entre 0,258 y 0,528; lo que evidenció una adecuada interrelación entre los componentes del modelo.

Fitriani A y otros⁽²⁰⁾ realizaron un estudio transversal para identificar los factores dominantes y sus relaciones en la adherencia a la HD con vistas a disminuir la mortalidad. La herramienta creada mostró un alfa de Cronbach de 0,838 que tradujo buena confiabilidad. Si bien se obtuvo mayor



fiabilidad en esta investigación, presentó limitaciones relacionadas con el manejo de datos y el diseño metodológico mostró sesgos.

La consistencia interna mide el grado de correlación y coherencia entre los predictores de un MP, es decir si hay homogeneidad entre ellos. En la presente investigación se utilizó el método de división de dos mitades de Spearman-Brown. Mediante esta técnica se divide el número de ítems en dos partes para correlacionar ambas y analizar si existe buena correlación.⁽¹⁴⁾ El resultado obtenido de 0,842 demuestra que al separar el MP en dos partes, representadas por la misma cantidad de predictores, existía buena correlación.

Para la predicción de mortalidad en pacientes con ERC en HD existen varias herramientas que presentan sensibilidad y especificidad adecuadas; pero algunas tienen el inconveniente de reunir variables, que en ocasiones, no se disponen al momento de ingreso a HD del paciente, ni en países de pequeños y medianos ingresos. Según *Escalona-González SO* y otros⁽⁸⁾ la RNA presentó porcentajes globales de clasificación correcto de un 96,3 % en la muestra de entrenamiento y un 96,7 % en la muestra de validación. La capacidad discriminatoria fue muy buena, área COR de 0,989. Si además de estos indicadores se contemplan las métricas obtenidas en la validación de apariencia, la fiabilidad y la consistencia interna; se ratifica que el MP puede usarse de forma rutinaria para la toma de decisiones clínicas.

Además, el MP está compuesto por 12 predictores accesibles que de forma sistemática se monitorean en los pacientes con ERC en HD. Desde la perspectiva clínica, los puntajes elevados otorgados por los expertos a los predictores: enfermedad cardiovascular, sepsis y albúmina < 30 g/L, refuerza la plausibilidad biológica del MP evaluado. Estos 3 predictores se asocian de manera independiente a la mortalidad en la ERC y describen escenarios fisiopatológicos que convergen en los pacientes en HD con desgaste urémico, como el estado inflamatorio crónico, la desnutrición, sobrecarga de volumen e inmunosupresión relativa.

Los tres predictores señalados actúan como ejes estructurales del constructo a medir, teniendo en cuenta que presentan fuerte correlación ítem-total y el impacto negativo en el alfa global si son excluidos del MP; por lo que aportan coherencia conceptual y estabilidad a la herramienta predictiva.



Un elemento relevante es la inclusión en el MP de variables operativas que son propias en el proceso de la HD, como el acceso vascular mediante catéter venoso central, el inicio tardío de la HD y la ganancia de peso interdialítica excesiva. Los tres predictores mencionados otorgan sensibilidad al MP al identificar riesgos modificables. Por lo que se pueden identificar oportunidades de intervención precoces que permitan una mejor adherencia al tratamiento dialítico con mejoras reales en la calidad en la atención.

La fiabilidad y consistencia interna obtenidas resaltan que los predictores no actúan de forma aislada, mantienen una relación armónica para predecir el constructo pronóstico. Esta característica reviste especial importancia, ya que el MP desarrollado está basado en un algoritmo de la IA y en ocasiones, la propia complejidad algorítmica puede ocultar incoherencias internas.⁽²¹⁾ Además, los indicadores de validez y de discriminación destacan que el MP es interpretable y reproducible.

La validez de apariencia ha sido subestimada tradicionalmente. El desarrollo de este tipo de validez en la presente investigación, favorece la aceptabilidad, la confianza del personal clínico y la implementación práctica de la herramienta predictiva en el ámbito asistencial. La concordancia observada entre los expertos demuestra que el MP podría reducir la brecha existente entre sesgos algorítmicos y aplicación clínica.⁽²²⁾

Es importante señalar, para investigaciones futuras, la necesidad de validaciones externas prospectivas, el aumento del número de expertos para explorar si existe estabilidad del consenso en otros contextos multicéntricos e internacionales y la evaluación de otras métricas de fiabilidad, como el coeficiente omega, para desarrollar información adicional importante del MP de estructura compleja.⁽²³⁾ Los métodos y resultados obtenidos muestran evidencia sólida novedosa en el campo de predicción de mortalidad en la ERC en HD. Además, se alinean a las recomendaciones sobre la validación y uso responsable de MP basados en algoritmos de la IA.

El modelo predictivo demostró una sólida validez aparente, alta fiabilidad y adecuada consistencia interna, por lo que puede usarse como herramienta de apoyo clínico.

El colectivo de autores recomienda la validación externa del modelo predictivo de forma prospectiva y la evaluación del impacto clínico real en la optimización de la atención nefrológica.





REFERENCIAS BIBLIOGRÁFICAS

1. Bello A, Okpechi I, Osman M, Cho Y, Htay H, Jha V, et al. Epidemiology of haemodialysis outcomes [Internet]. *Nature Reviews Nephrology* 2022;47(8):835–850. DOI: <https://doi.org/10.1038/s41581-022-00542-7>
2. Chadban S, Arici M, Power A, Wu MS, Mennini FS, Arango-Álvarez JJ, et al. Projecting the economic burden of chronic kidney disease at the patient level (*Inside CKD*): a microsimulation modelling study [Internet]. *EClinicalMedicine* 2024;72(1):102615. DOI: <https://doi.org/10.1016/j.eclinm.2024.102615>
3. Escalona-González SO, González-Milán ZC. Predicción de mortalidad en pacientes con enfermedad renal crónica en hemodiálisis: inteligencia artificial frente a los modelos predictivos tradicionales [Internet]. *Revista Electrónica Dr. Zoilo E. Marinello Vidaurreta* 2025 [acceso: 06/07/2025]; 50(1):e3914. Disponible en: <https://revzoilomarinellosldcu/index.php/zmv/article/view/3914>
4. Hippisley-Cox J, Coupland C, Bafadhel M, Russell R, Sheikh A, Brindle P, et al. Development and validation of a new algorithm for improved cardiovascular risk prediction [Internet]. *Nature Medicine* 2024;30(5):1440-7. DOI: <https://doi.org/10.1038/s41591-024-02905-y>
5. Williams K, Michalska S, Cohen E, Szomszor M, Grant J. Exploring the application of machine learning to expert evaluation of research impact [Internet]. *Plos one* 2023;18(8):e0288469. DOI: <https://doi.org/10.1371/journal.pone.0288469>
6. Rahrooh A, Garlid AO, Bartlett K, Coons W, Petousis P, Hsu W, et al. Towards a framework for interoperability and reproducibility of predictive models [Internet]. *Journal of Biomedical Informatics* 2024;149(1):104551. DOI: <https://doi.org/10.1016/j.jbi.2023.104551>
7. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view [Internet]. *Journal of medical Internet research*. 2016;18(12):e323. DOI: <https://doi.org/10.2196/jmir.5870>





8. Escalona-González SO, Caballero-Mota Y, Rodríguez-Alvarez Y, León-Acebo M, González-Milán ZC, Ricardo-Páez B, et al. Red neuronal artificial para la predicción de mortalidad de pacientes con enfermedad renal crónica [Internet]. Revista Cubana de Medicina Militar. 2024 [acceso: 06/07/2025]; 53(3):e024038408. Disponible en:
<https://revmedmilitar.sld.cu/index.php/mil/article/view/38408>
9. Burguet Lago I, Rodríguez Rabelo A, Jorge Chacón D. Aplicación de tecnologías para la determinación de la competencia de los expertos [Internet]. Revista Cubana de Ciencias Informáticas 2019 [acceso: 06/07/2025];13(1):116-26. Disponible en:
http://scielo.sld.cu/scielo.php?pid=S2227-18992019000100116&script=sci_arttext&lng=en
10. Rivera-Cantos GL, Orellana-Romero JE, Barzola-Veliz VM. Validación de una estrategia educativa mediante el método de criterio de expertos o Delphi [Internet]. Revista Estudios del Desarrollo Social: Cuba y América Latina. 2023 [acceso: 06/07/2025];11(3):e40. Disponible en:
http://scielo.sld.cu/scielo.php?pid=S2308-01322023000300040&script=sci_arttext
11. Majszak M, Jebeile J. Expert judgment in climate science: How it is used and how it can be justified [Internet]. Studies in history and philosophy of science. 2023;100(1):32-8. DOI:
<https://doi.org/10.1016/j.shpsa.2023.05.005>
12. Belmar-Garrido HM. Expert validation of a Python test, reliability, difficulty and discrimination indices [Internet]. Journal of Education and Development. 2023;7(1):52. DOI:
<https://doi.org/10.20849/jed.v7i1.1320>
13. Rodríguez-Rodríguez J, Reguant-Alvarez M. Calcular la fiabilidad de un cuestionario o escala mediante el SPSS: el coeficiente alfa de *Cronbach* [Internet]. Revista de innovación e investigación. 2020;13(2):8. DOI: <https://doi.org/10.1344/reire2020.13.230048>
14. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view [Internet]. Journal of medical Internet research. 2016;18(12):e323. DOI:
<https://doi.org/10.2196/jmir.5870>
15. Díez-Sanmartín C, Sarasa Cabezuelo A, Belmonte AA. A new approach to predicting mortality in dialysis patients using sociodemographic features based on artificial intelligence





[Internet]. Artificial Intelligence in Medicine. 2023;136(1):102478. DOI:

<https://doi.org/10.1016/j.artmed.2022.102478>

16. Evangelidis N, Tong A, Manns B, Hemmelgarn B, Wheeler DC, Tugwell P, et al. Developing a set of score outcomes for trials in hemodialysis: an international Delphi survey [Internet].

American Journal of Kidney Diseases. 2017;70(4):464-475. DOI:

<https://doi.org/10.1053/j.ajkd.2016.11.029>

17. Kalkbrenner MT. Alpha, omega, and H internal consistency reliability estimates: reviewing these options and when to use them [Internet]. Counseling Outcome Research and Evaluation.

2023;14(1):77-88. DOI: <https://doi.org/10.1080/21501378.2021.1940118>

18. Peterson RA, Kim Y. On the relationship between coefficient alpha and composite reliability [Internet]. Journal of applied psychology 2013;98(1):194-8. DOI:

<https://doi.org/10.1037/a0030767>

19. Toro R, Peña-Sarmiento M, Avendaño-Prieto BL, Mejía-Vélez S, Bernal-Torres A. Análisis empírico del coeficiente Alfa de Cronbach en función de las opciones de respuesta a las preguntas, tamaño de la muestra y valores atípicos [Internet]. Revista Iberoamericana de Diagnóstico y Evaluación Psicológica. 2022 [acceso: 06/07/2025];63(1):1-18. Disponible en:

<https://www.redalyc.org/journal/4596/459671926003/html/>

20. Fitriani A, Yetti K, Sukmarini L. Contributing factors to hemodialysis adherence in Aceh, Indonesia [Internet]. Enfermería clínica. 2019;29(2):238-42. DOI:

<https://doi.org/10.1016/j.enfcli.2019.04.028>

21. Petch J, Di S, Nelson W. Opening the black box: the promise and limitations of explainable machine learning in Cardiology [Internet]. Canadian Journal of Cardiology. 2022 [acceso:

06/07/2025]; 38(2):204-213. DOI: <https://doi.org/10.1016/j.cjca.2021.09.004>

22. Nazer L, Zatarah R, Waldrip S, Ke J, Moukheiber M, Khanna A, et al. Bias in artificial intelligence algorithms and recommendations for mitigation [Internet]. PLOS digital health

2023;2(6):e0000278. DOI: <https://doi.org/10.1371/journal.pdig.0000278>





23. Walsh I, Fishman D, Garcia-Gasulla D, Titma T, Pollastri G, Harrow J, et al. DOME: recommendations for supervised machine learning validation in biology [Internet]. Nature Methods 2021;18(1):1122-7. DOI: <https://doi.org/10.1038/s41592-021-01205-4>

Conflictos de interés

Los autores declaran que no existen conflictos de interés.

Contribuciones de los autores

Conceptualización: *Sergio Orlando Escalona González, Zoraida Caridad González Milán, Yailé Caballero Mota, Yanela Rodríguez Alvarez, Diamela Henríquez Navarro, Andrés Villar Bahamonde, Raymar Molina Vega.*

Curación de datos: *Sergio Orlando Escalona González, Zoraida Caridad González Milán.*

Análisis formal: *Sergio Orlando Escalona González, Zoraida Caridad González Milán, Yailé Caballero Mota, Yanela Rodríguez Alvarez.*

Investigación: *Sergio Orlando Escalona González, Zoraida Caridad González Milán, Yailé Caballero Mota, Yanela Rodríguez Alvarez.*

Metodología: *Sergio Orlando Escalona González.*

Administración del proyecto: *Sergio Orlando Escalona González, Zoraida Caridad González Milán.*

Recursos: *Sergio Orlando Escalona González, Zoraida Caridad González Milán, Yailé Caballero Mota, Yanela Rodríguez Alvarez.*

Supervisión: *Sergio Orlando Escalona González, Zoraida Caridad González Milán, Yailé Caballero Mota, Yanela Rodríguez Alvarez.*

Validación: *Sergio Orlando Escalona González, Zoraida Caridad González Milán, Yailé Caballero Mota, Yanela Rodríguez Alvarez.*





Visualización: *Sergio Orlando Escalona González, Zoraida Caridad González Milán, Yailé Caballero Mota, Yanela Rodríguez Alvarez, Diamela Henriquez Navarro, Andrés Villar Bahamonde, Raymar Molina Vega.*

Redacción – revisión y edición: *Sergio Orlando Escalona González, Zoraida Caridad González Milán, Yailé Caballero Mota, Yanela Rodríguez Alvarez, Diamela Henriquez Navarro, Andrés Villar Bahamonde, Raymar Molina Vega.*

Declaración de disponibilidad de datos

Archivo complementario: Encuesta aplicada para la selección de expertos; tabla patrón para calcular coeficiente de argumentación de los expertos según las fuentes que han influido más en sus conocimientos; encuesta aplicada a los expertos respecto a la relación de los predictores incluidos en el modelo predictivo y su implicancia con la mortalidad; estadísticas de fiabilidad, estadísticas ítem-total, prueba Hotelling's T^2 , estadísticas de consistencia interna (PDF). Disponible en: <https://revmedmilitar.sld.cu/index.php/mil/libraryFiles/downloadPublic/167>

