

Del pretest a la decisión clínica: el valor de la reclasificación del riesgo

Santiago Decotto  · Ana Miceli  · Rodolfo Pizarro 

Servicio de Cardiología, Hospital Italiano de Buenos Aires.
Ciudad Autónoma de Buenos Aires. Argentina.

Acta Gastroenterol Latinoam 2026;56(2):127-130

Recibido: 22/06/2026 / Aceptado: 26/06/2026 / Publicado online: 30/06/2026 / <https://doi.org/10.52787/agl.v56i2.664>

Introducción

En medicina nos encontramos constantemente en búsqueda de nuevas herramientas diagnósticas y pronósticas. Es así que continuamente se incorporan variables clínicas, biomarcadores séricos, técnicas de imagen y nuevos modelos predictivos con el objetivo de mejorar la estimación del riesgo y guiar la toma de decisiones en nuestros pacientes. Sin embargo, la mera asociación estadística entre una variable y un desenlace clínico no garantiza su utilidad en la práctica asistencial.

Frente a la aparición de un nuevo marcador, su utilidad no radica únicamente en la capacidad de predecir eventos, sino en si aporta información capaz de modificar la estimación del riesgo de un paciente de manera clínicamente significativa.¹ En otras palabras, ¿permite tomar mejores decisiones?

Toda evaluación clínica comienza con la estimación

de la probabilidad pretest, a partir de antecedentes, factores de riesgo, hallazgos clínicos relevantes y estudios complementarios. Sobre esta base, los pacientes suelen clasificarse en categorías de riesgo bajo, intermedio o alto, que orientan conductas diagnósticas y terapéuticas específicas.² La incorporación de una nueva variable adquiere verdadero valor cuando logra desplazar a un individuo desde una categoría de riesgo hacia otra más apropiada, modificando potencialmente la conducta terapéutica.

Este concepto, conocido como **reclasificación** del riesgo, ha cobrado creciente relevancia en la evaluación de nuevos biomarcadores y modelos predictivos.³ En los últimos años, numerosos estudios han demostrado que mejoras aparentemente modestas en las medidas tradicionales de discriminación no siempre se traducen en cambios clínicamente relevantes. Como consecuencia, han surgido herramientas metodológicas destinadas a evaluar no solo la capacidad predictiva de un modelo, sino también su impacto sobre la clasificación de los pacientes y, en última instancia, sobre la toma de decisiones.

El objetivo de esta revisión es describir los fundamentos de la reclasificación del riesgo, analizar sus principales métricas y discutir su relevancia para la evaluación crítica de nuevas herramientas diagnósticas y pronósticas.

¿Cómo evaluamos un modelo predictivo?

Tradicionalmente, el rendimiento de un modelo predictivo se ha evaluado mediante su **capacidad de discriminación**, es decir, su habilidad para diferenciar correctamente a los pacientes que presentarán un evento de aquellos que no lo harán.

Correspondencia: Santiago Decotto
Correo electrónico: santiago.decotto@hospitalitaliano.org.ar

La herramienta más utilizada para este propósito es la curva **ROC** (*Receiver Operating Characteristic*), que representa la relación entre sensibilidad y especificidad para distintos puntos de corte de una determinada prueba diagnóstica. El **área bajo la curva** (AUC) resume esta capacidad discriminativa en un único valor, donde 0,5 representa una discriminación equivalente al azar y 1,0 una discriminación perfecta. En términos generales, cuanto mayor es el AUC, mejor es la capacidad del modelo para distinguir entre pacientes con y sin eventos.^{4,5}

La curva ROC permite visualizar cómo varía el rendimiento diagnóstico al modificar el punto de corte de una prueba. Cada punto de la curva representa un valor umbral con un equilibrio diferente entre sensibilidad y especificidad. Uno de los métodos más utilizados para seleccionar un punto de corte es el *índice de Youden*, que identifica el valor que maximiza simultáneamente ambas medidas (sensibilidad y especificidad). Sin embargo, el punto de corte estadísticamente óptimo no siempre coincide con el más adecuado desde una perspectiva clínica. En la práctica, la elección del umbral dependerá del contexto clínico y de las consecuencias que puedan tener los errores diagnósticos. Así, en algunas situaciones se prioriza la sensibilidad para minimizar falsos negativos, como ocurre en las pruebas utilizadas como estrategia de tamizaje de enfermedades potencialmente graves, mientras que en otras es preferible maximizar la especificidad para evitar falsos positivos e intervenciones innecesarias, particularmente cuando el objetivo es confirmar un diagnóstico antes de indicar un tratamiento específico.

Por este motivo, la comparación de áreas bajo la curva se ha convertido en una práctica habitual para evaluar nuevos biomarcadores o modelos pronósticos. Sin embargo, aunque la discriminación es una propiedad importante, no necesariamente refleja el impacto clínico de una nueva variable. En muchos casos, la incorporación de un biomarcador puede generar mejoras mínimas en el AUC y, aun así, modificar de manera significativa la clasificación del riesgo de determinados pacientes, o viceversa. Esta limitación impulsó el desarrollo de nuevas herramientas destinadas a evaluar el valor incremental de una prueba más allá de la discriminación tradicional.

Concepto de reclasificación

Como se mencionó, aunque las medidas de discriminación son útiles para evaluar la capacidad predictiva de un modelo, no siempre reflejan su impacto sobre la toma de decisiones. En la práctica clínica, las decisiones rara vez se basan en probabilidades exactas; por el contrario, los pacientes suelen agruparse en categorías de riesgo que orientan conductas diagnósticas y terapéuticas específicas.

Desde esta perspectiva, una mejora en la capacidad discriminativa de un modelo no necesariamente implica un beneficio clínico. Un biomarcador puede aumentar marginalmente el AUC sin modificar la clasificación de ningún paciente. Por el contrario, una variación aparentemente modesta en los parámetros tradicionales puede traducirse en una reclasificación significativa de individuos ubicados cerca de los puntos de corte clínicamente relevantes.

El concepto de **reclasificación** surge precisamente para responder a esta necesidad. Su objetivo es evaluar si la incorporación de una nueva variable permite asignar a los pacientes a categorías de riesgo más apropiadas, acercando la predicción estadística a la toma de decisiones clínicas.⁶

¿Cómo evaluamos la reclasificación?

En 2008 **Pencina y col.** introdujeron una nueva medida denominada **índice de reclasificación neta** (NRI).⁷ Este fue desarrollado para evaluar si un nuevo marcador proporciona una mejora clínicamente relevante en la predicción de eventos. Su fundamento es sencillo: un nuevo marcador útil desde el punto de vista clínico debería incrementar el riesgo estimado de los pacientes que efectivamente presentarán el evento y reducirlo en aquellos que permanecerán libres de él, favoreciendo así una reclasificación más adecuada del riesgo individual. El objetivo es aumentar tanto la proporción de individuos con evento con reclasificación ascendente (reestratificación a categoría de mayor riesgo al aplicar el modelo nuevo), como la proporción de individuos sin evento con reclasificación descendente (reestratificación a categoría de menor riesgo al aplicar el modelo nuevo).

Para su cálculo, la población se divide en individuos que presentaron el evento de interés y aquellos que permanecieron libres de él durante el seguimiento. En los sujetos con eventos se considera favorable una reclasificación hacia categorías de mayor riesgo, mientras que en aquellos sin eventos se considera favorable una reclasificación hacia categorías de menor riesgo.

El NRI cuantifica el balance neto entre estas reclasificaciones correctas e incorrectas. De esta manera, valores positivos indican que el nuevo modelo mejora la clasificación del riesgo respecto del modelo original, mientras que valores cercanos a cero sugieren una escasa contribución adicional. Cuanto mayor sea el NRI, mayor será la capacidad del nuevo marcador para reclasificar adecuadamente a los individuos. La **figura central** resume de manera esquemática los principios que sustentan el cálculo del NRI y los distintos escenarios de reclasificación observados tras la incorporación de una nueva variable a un modelo predictivo.

Aplicabilidad y generalización de los modelos predictivos

Un aspecto fundamental a considerar es que si bien la sensibilidad, la especificidad y el AUC son propiedades intrínsecas de una prueba o modelo, su utilidad clínica varía significativamente según la **probabilidad pretest** de la población en la que se apliquen. Así, una misma prueba puede presentar un elevado valor predictivo positivo en poblaciones de alto riesgo y un rendimiento mucho menor cuando se utiliza en poblaciones de bajo riesgo. De manera inversa, el valor predictivo negativo suele incrementarse a medida que disminuye la prevalencia del evento.

Esta dependencia de la probabilidad pretest tiene implicancias relevantes para la validación de modelos predictivos. En poblaciones con una prevalencia extremadamente baja o elevada de un evento los valores predictivos pueden resultar artificialmente favorables. Por este motivo, la validación de un marcador suele ser particularmente informativa en **poblaciones de riesgo intermedio**, donde existe una mayor incertidumbre diagnóstica y en las que una herramienta predictiva tiene mayores posibilidades de reclasificar pacientes de manera clínicamente significativa.⁸

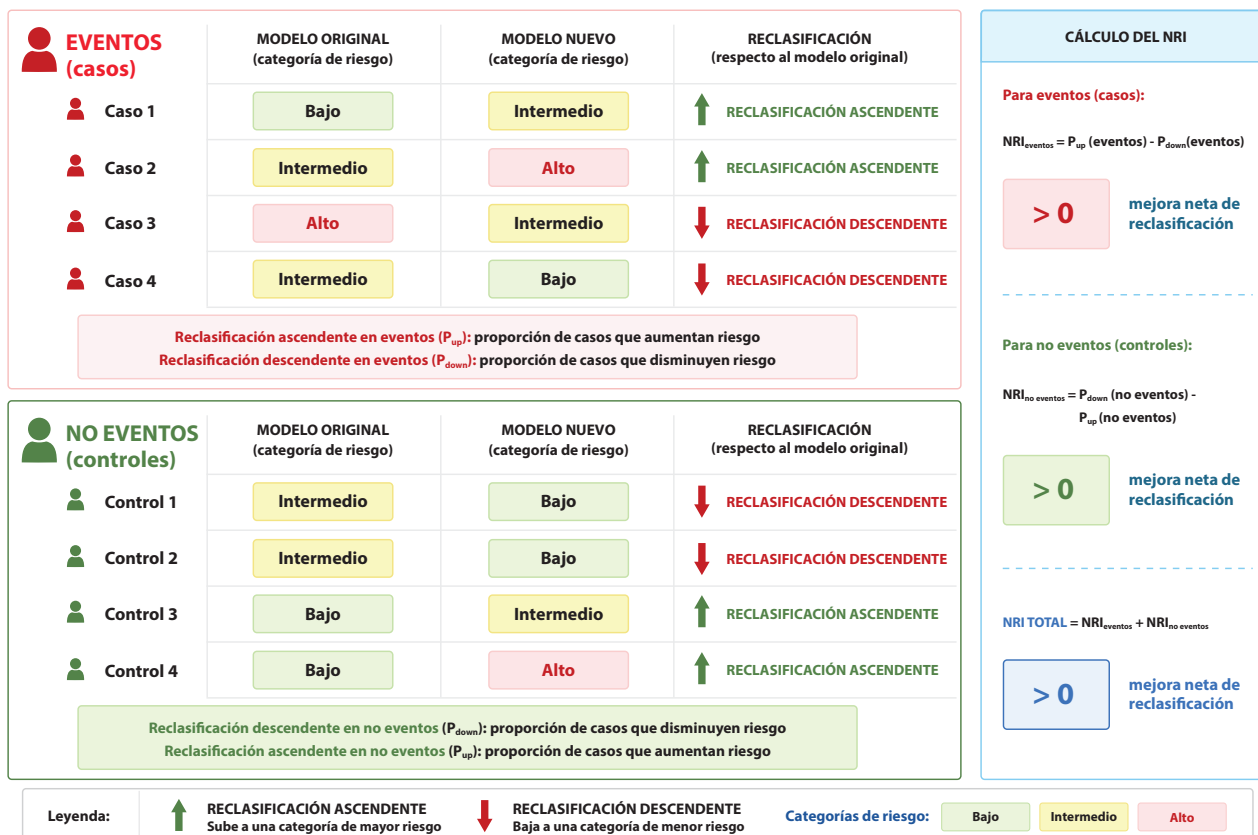
Conclusiones

La evaluación de nuevas herramientas diagnósticas y pronósticas no debería limitarse exclusivamente a las medidas tradicionales de discriminación. Si bien la curva ROC y el AUC continúan siendo herramientas fundamentales para valorar el rendimiento de un modelo, su capacidad para reflejar el impacto clínico de una nueva variable es limitada. La reclasificación del riesgo aporta una perspectiva complementaria al evaluar si la información adicional modifica de manera apropiada la estimación de riesgo individual y, potencialmente, la conducta médica.

Este concepto adquiere especial relevancia en poblaciones de riesgo intermedio, donde la incertidumbre diagnóstica es mayor y donde una nueva herramienta tiene mayores posibilidades de modificar la estimación de riesgo y cambiar una conducta clínica. En este contexto, la capacidad de reclasificar correctamente a los pacientes puede resultar más valiosa que pequeñas mejoras en las medidas tradicionales de discriminación.

Finalmente, el objetivo no es solamente predecir mejor, sino clasificar mejor a los pacientes para tomar decisiones más adecuadas.

Figura central. Se resumen de manera esquemática los principios que sustentan el cálculo del índice de reclasificación neta (NRI) y los distintos escenarios de reclasificación observados tras la incorporación de una nueva variable a un modelo predictivo



Propiedad intelectual. Los autores declaran que los datos y la figura presentes en el manuscrito son originales y se realizaron en sus instituciones pertenecientes.

Financiamiento. Los autores declaran que no hubo fuentes de financiación externas.

Conflictos de interés. Los autores declaran no tener conflictos de interés en relación con este artículo.

Aviso de derechos de autor



© 2026 Acta Gastroenterológica Latinoamericana. Este es un artículo de acceso abierto publicado bajo los términos de la Licencia Creative Commons Attribution (CC BY-NC-SA 4.0), la cual permite el uso, la distribución y la reproducción de forma no comercial, siempre que se cite al autor y la fuente original.

Cite este artículo como: Decotto S, Miceli A y Pizarro R. Del pretest a la decisión clínica: el valor de la reclasificación del riesgo. *Acta Gastroenterol Latinoam.* 2026;56(2):127-130. <https://doi.org/10.52787/agl.v56i2.664>

Referencias

1. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* 2007;115(7):928-35.
2. Pauker SG, Kassirer JP. The threshold approach to clinical decision making. *N Engl J Med* 1980;302(20):1109-17.
3. Hlatky MA, Greenland P, Arnett DK, *et al.* Criteria for evaluation of novel markers of cardiovascular risk: a scientific statement from the American Heart Association. *Circulation* 2009;119(17):2408-16.
4. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143(1):29-36.
5. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem* 1993;39(4):561-77.
6. Kerr KF, Wang Z, Janes H, McClelland RL, Psaty BM, Pepe MS. Net reclassification indices for evaluating risk prediction instruments: a critical review. *Epidemiology* 2014;25(1):114-21.
7. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 2008;27(2):157-72; discussion 207-12.
8. Greenland P, Alpert JS, Beller GA, *et al.* 2010 ACCF/AHA guideline for assessment of cardiovascular risk in asymptomatic adults: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol* 2010;56(25):e50-103.

From Pretest to Clinical Decision: The Value of Risk Reclassification

Santiago Decotto  · Ana Miceli  · Rodolfo Pizarro 

Cardiology Department, Hospital Italiano de Buenos Aires.
Ciudad Autónoma de Buenos Aires. Argentina.

Acta Gastroenterol Latinoam 2026;56(2):131-134

Received: 22/06/2026 / Accepted: 26/06/2026 / Published online: 30/06/2026 / <https://doi.org/10.52787/agl.v56i2.664>

Introduction

In medicine, we are constantly seeking new diagnostic and prognostic tools. As a result, clinical variables, serum biomarkers, imaging techniques, and novel predictive models are continually being incorporated with the aim of improving risk assessment and guiding clinical decision-making for our patients. However, a mere statistical association between a variable and a clinical outcome does not guarantee its usefulness in clinical practice.

When a new marker emerges, its value lies not only in its ability to predict events, but also in whether it provides information capable of modifying a patient's risk estimation in a clinically meaningful way.¹ In other words, does it enable better decision-making?

Every clinical evaluation begins with an estimate of the pretest probability, based on medical history, risk factors, relevant clinical findings, and complementary diagnostic tests. On this basis, patients are typically classified into categories of low-, intermediate-, or high-risk, which guide specific diagnostic and therapeutic strategies.² The incorporation of a new variable acquires true value when it shifts an individual from one risk category to a more appropriate one, potentially altering the therapeutic approach.

This concept, known as **risk reclassification**, has become increasingly important in the evaluation of new biomarkers and predictive models.³ In recent years, numerous studies have shown that seemingly modest improvements in traditional measures of discrimination do not always translate into clinically relevant changes. As a result, methodological tools have been developed to assess not only a model's predictive performance, but also its impact on patient classification and, ultimately, on clinical decision-making.

The aim of this review is to describe the fundamentals of risk reclassification, analyze its principal metrics, and discuss its relevance for the critical evaluation of new diagnostic and prognostic tools.

How Do We Evaluate a Predictive Model?

Traditionally, the performance of a predictive model has been evaluated based on **discriminatory ability**, that is, its capacity to correctly distinguish between

Correspondence: Santiago Decotto
Email: santiago.decotto@hospitalitaliano.org.ar

patients who will experience an event and those who will not.

The most widely used tool for this purpose is the **Receiver Operating Characteristic (ROC) curve**, which represents the relationship between sensitivity and specificity for different cutoff points of a given diagnostic test. The **area under the curve (AUC)** summarizes this discriminatory ability in a single value, where 0.5 represents discrimination equivalent to chance and 1.0 indicates perfect discrimination. In general, the higher the AUC, the better the model's ability to distinguish between patients with and without events.^{4,5}

The ROC curve allows us to visualize how diagnostic performance varies when the cutoff point of the test is modified. Each point on the curve represents a threshold value with a different balance between sensitivity and specificity. One of the most commonly used methods for selecting a cutoff point is the **Youden index**, which identifies the value that simultaneously maximizes both measures (sensitivity and specificity). However, the statistically optimal cutoff point does not always coincide with the most appropriate one from a clinical perspective.

In practice, the choice of threshold will depend on the clinical context and the potential consequences of diagnostic errors. Thus, in some situations, sensitivity is prioritized to minimize false negatives -as is the case with tests used as a screening strategy for potentially serious diseases- while in others, it is preferable to maximize specificity to avoid false positives and unnecessary interventions, particularly when the goal is to confirm a diagnosis before prescribing a specific treatment.

For this reason, comparing the **area under the curve (AUC)** has become standard practice for evaluating new biomarkers or prognostic models. However, although discriminatory power is an important property, it does not necessarily reflect the clinical impact of a new variable. In many cases, the inclusion of a biomarker may result in minimal improvements in the AUC and yet significantly alter the risk classification of certain patients, or vice versa. This limitation has driven the development of new tools designed to evaluate the **incremental value** of a test beyond traditional measures of discrimination.

The Concept of Reclassification

As mentioned, although measures of discrimination are useful for evaluating a model's predictive ability, they do not always reflect its impact on decision-making. In clinical practice, decisions are rarely based on exact probabilities; rather, patients are typically grouped into risk

categories that guide specific diagnostic and therapeutic strategies.

From this perspective, an improvement in a model's discriminatory ability does not necessarily imply a clinical benefit. A biomarker may marginally increase the AUC without changing the classification of any patient. Conversely, an apparently modest change in traditional parameters can result in a significant reclassification of individuals located near clinically relevant decision cutoff points.

The concept of **reclassification** emerged precisely to address this need. Its purpose is to evaluate whether the addition of a new variable allows patients to be assigned to more appropriate risk categories, bringing statistical prediction closer to clinical decision-making.⁶

How Do We Assess Reclassification?

In 2008, **Pencina *et al.*** introduced a new metric called the **Net Reclassification Index (NRI)**.⁷ It was developed to evaluate whether a new marker provides a clinically relevant improvement in prediction of events. Its rationale is straightforward: a clinically useful new marker should increase the estimated risk for patients who will actually experience the event and reduce it for those who will remain event-free, thereby promoting a more appropriate classification of individual risk. The goal is to increase both the proportion of individuals who experienced an event and were reclassified upward (restratified into a higher-risk category when applying the new model) and the proportion of individuals who did not experience an event and were reclassified downward (restratified into a lower-risk category when applying the new model).

To calculate this, the population is divided into individuals who experienced the event of interest and those who remained event-free during follow-up. For event-experiencing subjects, reclassification into higher-risk categories is considered favorable, whereas for event-free subjects, reclassification into lower-risk categories is considered favorable.

The **NRI** quantifies the net balance between these correct and incorrect reclassifications. Thus, positive values indicate that the new model improves risk classification compared with the original model, whereas values close to zero suggest little additional benefit. The higher the NRI, the greater the new marker's ability to appropriately reclassify individuals.

The **Central Figure** schematically summarizes the principles underlying the calculation of the NRI and the different reclassification scenarios observed after incorporating a new variable into a predictive model.

Applicability and Generalizability of Predictive Models

An important consideration is that although sensitivity, specificity, and AUC are intrinsic properties of a test or model, their clinical utility varies substantially according to the **pretest probability** of the population in which they are applied. Thus, the same test may have a high positive predictive value in high-risk populations and perform considerably less well when used in low-risk populations. Conversely, the negative predictive value generally increases as the prevalence of the event decreases.

This dependence on pretest probability has important implications for the validation of predictive models. In populations with an extremely low or high prevalence of an event, predictive values may be artificially favorable. For this reason, the validation of a marker is often particularly informative in **intermediate-risk populations**, where there is greater diagnostic uncertainty and where a predictive tool has a higher likelihood of reclassifying patients in a clinically meaningful way.⁸

Conclusions

The evaluation of new diagnostic and prognostic tools should not be limited exclusively to traditional measures of discrimination. Although the ROC curve and the AUC remain fundamental tools for assessing a model’s performance, their ability to reflect the clinical impact of a new variable is limited. Risk reclassification provides a complementary perspective by evaluating whether additional information appropriately modifies the individual risk estimation and, potentially, clinical decision-making.

This concept is particularly relevant in intermediate-risk populations, where diagnostic uncertainty is greater and where a new tool has a higher likelihood of altering risk estimation and changing clinical management. In this context, the ability to correctly reclassify patients may be more valuable than small improvements in traditional measures of discrimination.

Ultimately, the goal is not only to predict better, but also to classify patients more effectively in order to support more appropriate clinical decisions.

Central Figure. Schematic overview of the principles underlying the calculation of the NRI and the different reclassification scenarios observed after adding a new variable to a predictive model

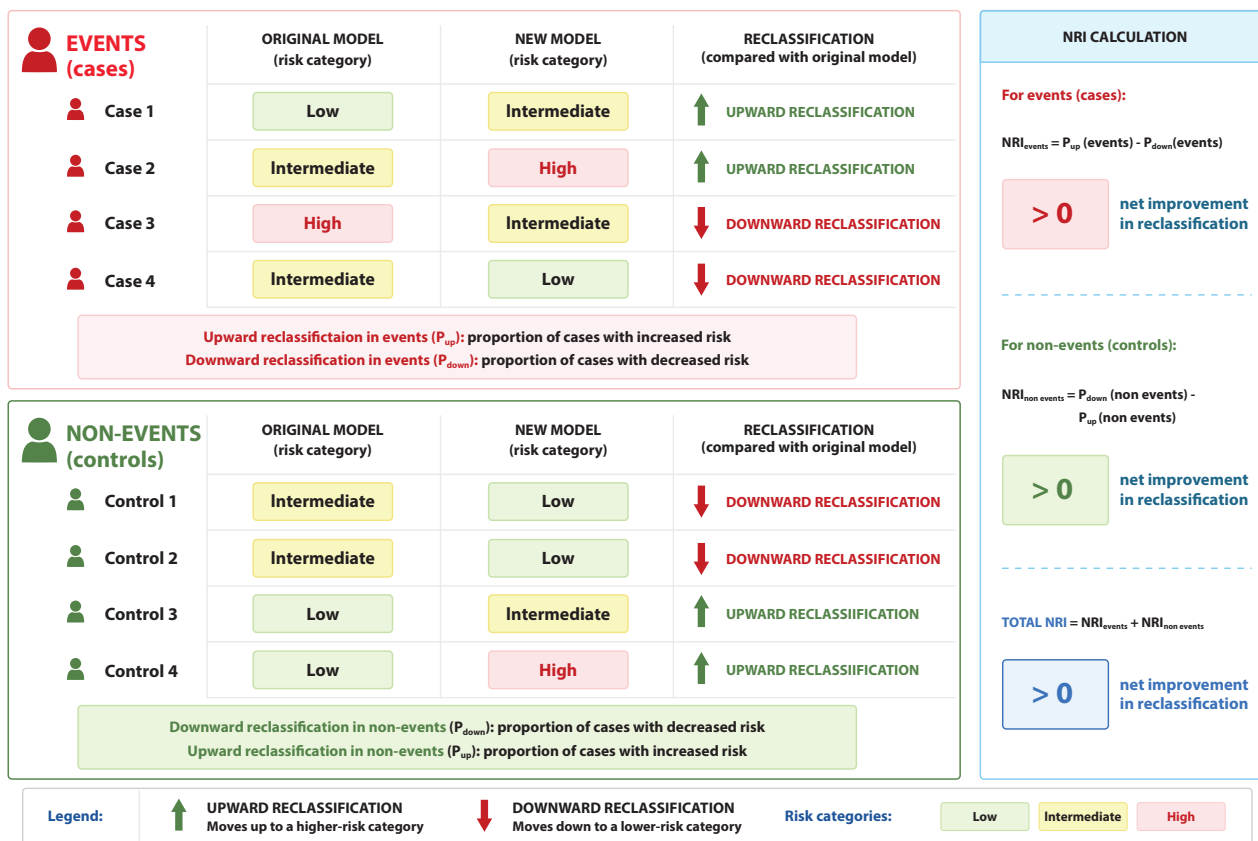


Figure prepared with the assistance of artificial intelligence.

Intellectual property. *The authors declare that the data and figure in this article are original and were carried out at their institutions.*

Funding. *The authors declare that there were no external sources of funding.*

Conflict of interest. *The authors declare that they have no conflicts of interest in relation to this article.*

Copyright



© 2026 *Acta Gastroenterológica latinoamericana*. This is an open-access article released under the terms of the Creative Commons Attribution (CC BY-NC-SA 4.0) license, which allows non-commercial use, distribution, and reproduction, provided the original author and source are acknowledged.

Cite this article as: Decotto S, Miceli A y Pizarro R. From Pretest to Clinical Decision: The Value of Risk Reclassification. *Acta Gastroenterol Latinoam*. 2026;56(2):131-134. <https://doi.org/10.52787/agl.v56i2.664>

References

1. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* 2007;115(7):928-35.
2. Pauker SG, Kassirer JP. The threshold approach to clinical decision making. *N Engl J Med* 1980;302(20):1109-17.
3. Hlatky MA, Greenland P, Arnett DK, *et al.* Criteria for evaluation of novel markers of cardiovascular risk: a scientific statement from the American Heart Association. *Circulation* 2009;119(17):2408-16.
4. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143(1):29-36.
5. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem* 1993;39(4):561-77.
6. Kerr KF, Wang Z, Janes H, McClelland RL, Psaty BM, Pepe MS. Net reclassification indices for evaluating risk prediction instruments: a critical review. *Epidemiology* 2014;25(1):114-21.
7. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 2008;27(2):157-72; discussion 207-12.
8. Greenland P, Alpert JS, Beller GA, *et al.* 2010 ACCF/AHA guideline for assessment of cardiovascular risk in asymptomatic adults: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol* 2010;56(25):e50-103.