

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

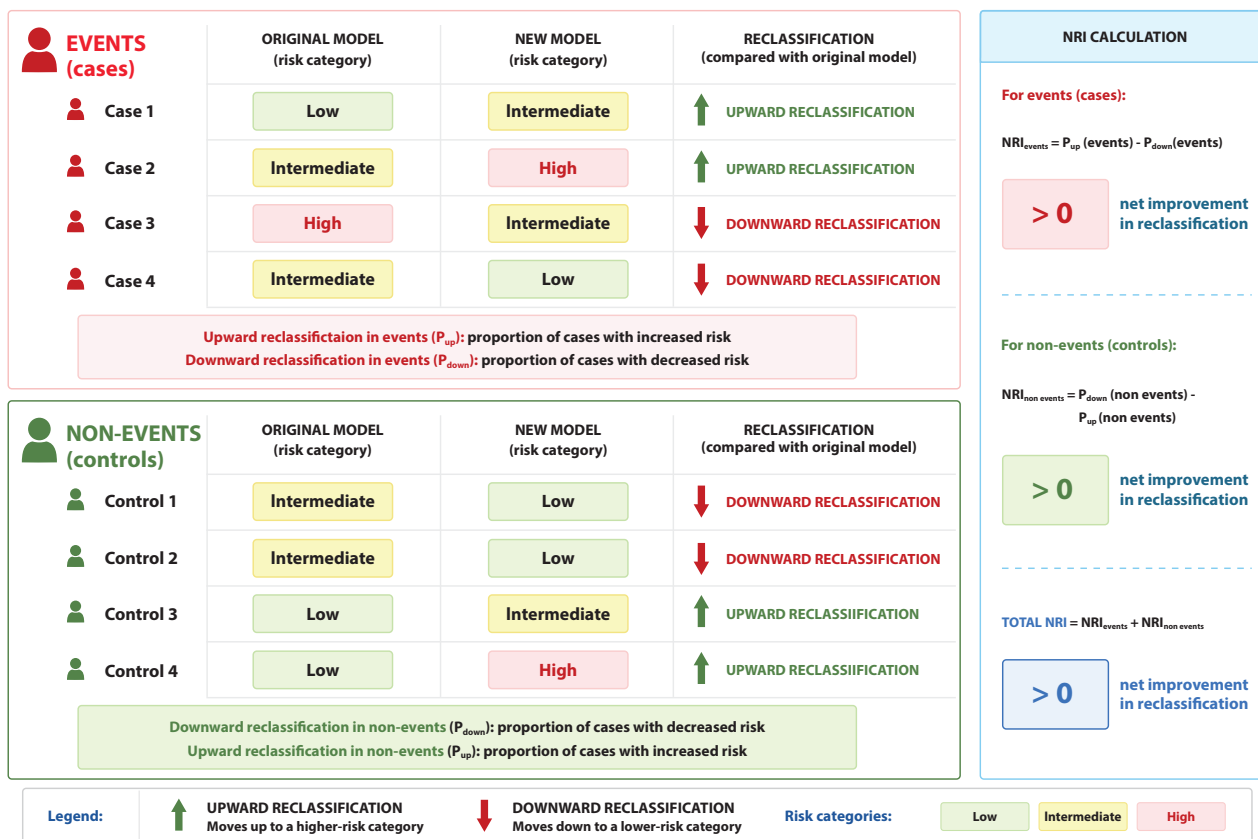


Figure prepared with the assistance of artificial intelligence.

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.