The role of stress testing in the noninvasive evaluation of patients with possible or known coronary artery disease has evolved from functioning as a diagnostic tool to serving primarily as an aid in risk stratification. The results of clinical assessment and stress testing can be combined to estimate individual patient risk and guide subsequent treatment. National guidelines recommend that stress imaging preferentially be performed instead of standard exercise testing in patients with characteristics consisting of inability to exercise adequately, uninterpretable exercise ECG, or previous coronary artery disease revascularization.1,2 Aside from these patient subsets, controversy persists concerning the selection of a stress test modality attributable to uncertainty concerning the incremental prognostic value of stress imaging compared with standard exercise testing.

In this issue of Circulation: Cardiovascular Imaging, Candell-Riera et al3 report on the prognostic value of exercise MPGS in 5672 patients with known or suspected coronary artery disease. The authors apply the c-statistic, net reclassification improvement (NRI), and integrated discrimination improvement (IDI) to examine the incremental prognostic value of MPGS compared with clinical and exercise test variables for predicting the end points of all-cause death and the major cardiovascular end point of combined cardiovascular death and nonfatal myocardial infarction. The authors report 2 major findings: (1) predictive accuracy significantly increased by adding exercise variables to clinical variables for both end points; (2) adding MPGS to the model containing clinical+exercise variables significantly improved predictive accuracy for the cardiovascular end point but not for the end point of overall mortality.

In view of the abundant literature examining the prognostic value of stress MPGS, why publish another article on this topic? The major contribution of the article by Candell-Riera et al3 is the application of NRI and IDI. The standard approach of examining the association between a variable of interest (in this case MPGS) and clinical outcome consists of applying a series of tests of increasing statistical rigor. The first test determines whether a statistically significant association exists by using simple univariable analysis. The second test evaluates the independent prognostic value of the variable of interest in comparison with other known important prognostic variables in a multivariable model by selecting variables according to strength of association with the end point in a stepwise fashion. The third test examines the incremental prognostic value of the variable. The concept of incremental modeling was introduced by Ladenheim et al4 and replicates clinical decision-making. The first step includes creating a model containing only clinical variables. In this process, the next step determines whether stress testing variables improve predictive accuracy compared with accuracy provided by the clinical variables alone. The third step examines whether imaging variables add to the value of the clinical+stress test model.

Global model $\chi^2$ values are calculated at each step. If there is a statistically significant increase in the global $\chi^2$ value progressing from 1 model to the next, then the conclusion is that the addition of these variables has added incremental value. The area under the receiver-operating characteristic curve and associated c-statistic can be determined for each model. This approach is well established and has become the standard by which the majority of recent studies have been performed, examining the prognostic accuracy of MPGS compared with standard exercise testing. The results convey important information to the reader, including identification of the variables in each statistical model which contain the most important prognostic information and illustrating whether a model containing MPGS variables is significantly better for predicting outcome compared with the clinical-exercise model. Although an improvement in the global model $\chi^2$ or c-statistic can be interpreted as a positive finding, this approach fails to convey clinically important information to the reader. These results do not indicate what percentage of the study population is reclassified as higher or lower risk using MPGS and whether the reclassification results primarily from accurately identifying more high-risk or more low-risk patients. Comparing predictive accuracy between 2 models using NRI and IDI can

**Editorial**

**Net Reclassification Improvement and Integrated Discrimination Improvement**

**New Standards for Evaluating the Incremental Value of Stress Imaging for Risk Assessment**

Todd D. Miller, MD; J. Wells Askew, MD

For nuclear myocardial perfusion gated SPECT (MPGS), patient risk can be determined simply by dichotomizing images as normal/abnormal,3 but risk stratification can be refined by categorizing the images on the basis of summed perfusion scores.4,5 These scores include the summed stress score, which is a reflection of the extent and severity of combined infarction and ischemia, and the summed different score, an indicator of the extent and severity of ischemia.

The authors apply the c-statistic, net reclassification improvement (NRI), and integrated discrimination improvement (IDI) to examine the incremental prognostic value of MPGS compared with clinical and exercise test variables for predicting the end points of all-cause death and the major cardiovascular end point of combined cardiovascular death and nonfatal myocardial infarction. The authors report 2 major findings: (1) predictive accuracy significantly increased by adding exercise variables to clinical variables for both end points; (2) adding MPGS to the model containing clinical+exercise variables significantly improved predictive accuracy for the cardiovascular end point but not for the end point of overall mortality.

In view of the abundant literature examining the prognostic value of stress MPGS, why publish another article on this topic? The major contribution of the article by Candell-Riera et al3 is the application of NRI and IDI. The standard approach of examining the association between a variable of interest (in this case MPGS) and clinical outcome consists of applying a series of tests of increasing statistical rigor. The first test determines whether a statistically significant association exists by using simple univariable analysis. The second test evaluates the independent prognostic value of the variable of interest in comparison with other known important prognostic variables in a multivariable model by selecting variables according to strength of association with the end point in a stepwise fashion. The third test examines the incremental prognostic value of the variable. The concept of incremental modeling was introduced by Ladenheim et al4 and replicates clinical decision-making. The first step includes creating a model containing only clinical variables. In this process, the next step determines whether stress testing variables improve predictive accuracy compared with accuracy provided by the clinical variables alone. The third step examines whether imaging variables add to the value of the clinical+stress test model. Global model $\chi^2$ values are calculated at each step. If there is a statistically significant increase in the global $\chi^2$ value progressing from 1 model to the next, then the conclusion is that the addition of these variables has added incremental value. The area under the receiver-operating characteristic curve and associated c-statistic can be determined for each model. This approach is well established and has become the standard by which the majority of recent studies have been performed, examining the prognostic accuracy of MPGS compared with standard exercise testing. The results convey important information to the reader, including identification of the variables in each statistical model which contain the most important prognostic information and illustrating whether a model containing MPGS variables is significantly better for predicting outcome compared with the clinical-exercise model. Although an improvement in the global model $\chi^2$ or c-statistic can be interpreted as a positive finding, this approach fails to convey clinically important information to the reader. These results do not indicate what percentage of the study population is reclassified as higher or lower risk using MPGS and whether the reclassification results primarily from accurately identifying more high-risk or more low-risk patients. Comparing predictive accuracy between 2 models using NRI and IDI can
provide this information. Additionally, the results of NRI can demonstrate how many patients are shifted into different clinical risk categories (low, moderate, high) by using a different model. The application of these principles is illustrated in the article by Candell-Riera et al. These authors show the conventional results of their study in Table 1 (univariable associations between clinical, exercise, and MPGS variables and outcome end points), Table 2 (important variables in each of the respective multivariable models), and Figure 2 (receiver-operating characteristic curves and c-statistic results). The more novel results from applying NRI and IDI are provided in Figure 3 and Table 3. Figure 3 illustrates the percentage of the population that is reclassified comparing each model with the previous model. The $P$ values associated with NRI and IDI indicate whether the reclassification was statistically significant. An important result to notice is the reclassification exercise results in some patient reclassifications that are incorrect. Figure 3A demonstrates for the end point overall mortality that the addition of MPGS to the clinical+exercise model actually results in more patients being incorrectly reclassified. For patients who died, 4.3% were incorrectly reclassified as lower risk, whereas 3.3% were correctly reclassified as higher risk, resulting in a net 1% of the population being incorrectly reclassified. Conversely, for patients who survived, the addition of MPGS resulted in 7.8% being correctly reclassified and 5.1% being incorrectly reclassified, yielding a net of 2.7% correctly reclassified. Overall, 1.7% of the population was reclassified (2.7%–1%), but this small difference was not statistically significant (NRI=0.017; $P=$NS). For the end point cardiovascular death and nonfatal myocardial infarction (Figure 3B), the value of adding MPGS to clinical+exercise variables was more favorable. Overall 12.2% of the population was correctly reclassified (2.7% correct reclassifications for patients with events, 9.5% correct reclassifications for patients without events). The NRI was 0.122 ($P<0.05$) and IDI 0.033 ($P<0.001$). Table 3 illustrates shifts between risk categories (low, moderate, high) determined by these 2 models. In this table, the correct reclassifications are shown in the cells shaded gray. An important observation is that the major effect of applying NRI and IDI is correctly reclassifying low-risk patients, primarily by shifting patients who do not experience events from the moderate-risk category to the low-risk category, rather than correctly identifying more high-risk patients. For instance, for the cardiovascular mortality or myocardial infarction end point, the addition of MPGS to the clinical+exercise model resulted in 966 patients (671+295) without an event being correctly reclassified as lower risk, whereas only 24 patients (1+23) with an event were correctly reclassified as higher risk. Because the addition of MPGS resulted in some incorrect reclassifications, the number of patients with an event who were classified as high-risk increased very modestly from 86 using the exercise model to 96 using the MPGS model.

Candell-Riera et al.6 also include cost estimates (and appropriately recommend cautious interpretation) for each reclassification of $4078.00 for the cardiovascular end point and $12408.00 for the mortality end point. The reader should appreciate that the major benefit of this additional cost of applying MPGS instead of exercise testing alone in the entire population is correctly identifying low-risk patients. These cost estimates reflect only the cost of imaging and do not address the cost of invasive downstream procedures, such as coronary angiography or revascularization. Even assuming that the correct reclassification of some patients into a higher-risk category could lead to an intervention that could be life-saving, the cost per year-of-life-saved would be substantially higher than the imaging costs of reclassification provided by Candell-Riera.6

The concept of applying nuclear imaging for reclassifying patient risk is not new.8–10 More recently, this concept has been popularized through the work of Pencina et al.11 and the introduction of the terms NRI and IDI. Only a handful of recent nuclear cardiology publications have applied this approach.12–14 The present study by Candell-Riera et al.8 can be added to this small group of important studies.

The application of NRI and IDI provides a superior method for analyzing the incremental value of performing a presumably better technique in place of a less accurate technique. In contrast to the traditional but more abstract concepts of comparing global $\chi^2$ and c-statistic values, this new methodology illustrates the percentage of a population that is accurately reclassified, providing clinically meaningful results rather than only statistically significant differences between $\chi^2$ and c-statistic values. The results of reclassification can be especially useful to address questions related to cost-effectiveness. For low-risk patients, can enough patients be accurately reclassified as low-risk to avoid unnecessary catheterizations and revascularization procedures to offset the initially higher cost of performing MPGS versus exercise testing alone in the population? For high-risk patients, can enough high-risk patients be accurately reclassified to justify the higher cost of MPGS versus exercise testing according to accepted standards expressed as the cost per year-of-life-saved? NRI and IDI have been applied to other areas besides stress imaging and should represent the new standards to judge the incremental risk stratification accuracy added by more expensive technologies versus less expensive techniques.

**Disclosures**

Dr Miller has a consulting agreement with Astellas Pharma.

**References**


**Key Words:** Editorials