Net Reclassification Improvement: Computation, Interpretation, and Controversies

A Literature Review and Clinician's Guide

Maarten J.G. Leening, MD, MSc; Moniek M. Vedder, MSc; Jacqueline C.M. Witteman, PhD; Michael J. Pencina, PhD; and Ewout W. Steyerberg, PhD

The net reclassification improvement (NRI) is an increasingly popular measure for evaluating improvements in risk predictions. This article details a review of 67 publications in high-impact general clinical journals that considered the NRI. Incomplete reporting of NRI methods, incorrect calculation, and common misinterpretations were found. To aid improved applications of the NRI, the article elaborates on several aspects of the computation and interpretation in various settings. Limitations and controversies are discussed, including the effect of miscalibration of prediction models, the use of the continuous NRI and "clinical NRI," and the relation with decision analytic measures. A systematic approach toward presenting NRI analysis is proposed: Detail and motivate the methods used for computation of the NRI, use clinically meaningful risk cutoffs for the category-based NRI, report both NRI components, address issues of calibration, and do not interpret the overall NRI as a percentage of the study population reclassified. Promising NRI findings need to be followed with decision analytic or formal cost-effectiveness evaluations.

Ann Intern Med. 2014;160:122-131. www.annals.org
For author affiliations, see end of text.

Since the introduction of the term *risk factor* more than 50 years ago in this journal (1), many such factors have been identified. Risk factors have been incorporated into statistical models to predict occurrence of disease, to more adequately diagnose patients, and to predict outcomes after disease has been diagnosed. A substantial number of clinical guidelines have incorporated risk prediction models to aid clinicians in everyday decision making in various fields of medicine, including cardiology, oncology, and respiratory medicine (2–8).

Many markers, such as biomarkers, genetic factors, and imaging results, have been proposed to improve these prediction models. In the past 3 decades, the most commonly used measure to quantify these improvements has been the change in the c-statistic, also known as the area under the receiver-operating characteristic curve (AUC). Studies have emphasized the limitations of the AUC, including the difficulty in interpreting the usually small changes in this statistic and the relation of the magnitude of improvement to the performance of the baseline model (9–12). A more relevant criterion may be to assess whether the addition of the marker to an existing model will influence clinical practice (13), which is the case if the newly predicted risk crosses a clinically meaningful threshold for an individual. This has led to the introduction of the concept of risk reclassification (14), which involves cross-tabulating categories of predicted risk for 2 models—usually one with the new marker under study and the other without it—to see how persons are classified differently when these models are used. The subsequent changes in risk classification can be quantified by the net reclassification improvement (NRI) (15). Risk reclassification analysis with the NRI has become popular: More than 1000 publications have cited the 2008 article that introduced the NRI (15). However, reporting of the methods used is of heterogeneous quality (16), and misconceptions are common in interpreting the NRI (17).

In this article, we aim to provide a systematic assessment of the reporting practices in analyses involving the NRI and address some controversies relating to its use and interpretation. We also make recommendations on how to report and interpret the NRI (18).

See also:

Print
Editorial comment. 136

Overview of Current Reporting

Literature Search and Data Extraction

We systematically collected studies that computed the NRI or discussed results from NRI analysis. We used the Thomson Reuters Web of Knowledge (version 5.9) to identify all publications that cited 1 of 4 methodological articles by Pencina and colleagues (15, 19–21) or a methodological review on reclassification measures by Cook and Ridker (22). The search was last updated on 23 April 2013 and yielded 1250 unique citations (**Appendix Figure 1**, available at www.annals.org). We selected all 67 citations in the 4 general clinical journals with the highest impact factors (*New England Journal of Medicine*, *The Lancet*, *Journal of the American Medical Association*, and *Annals of Internal Medicine*) (22–88) for data extraction (**Appendix Tables 1** and **2**, available at www.annals.org). Our rationale was that these articles may be expected to have broad impact and be used as examples for others.

Two evaluators independently extracted data from the publications. Cases on which the evaluators disagreed were discussed with a third evaluator to reach consensus. All

publications were searched for NRI calculations or results. If these were found, we checked which version of the NRI was used: the category-based NRI (15) or the continuous (category-free) NRI (20) (**Table 1**). Next, we reviewed all articles to determine whether risk categories corresponding to diagnostic or treatment thresholds from clinical guidelines were used to evaluate the category-based NRI or whether other categorization was justified. We determined which NRI components were reported: solely the overall NRI, or the event NRI and the nonevent NRI (**Table 1**). Moreover, we categorized studies that reported estimates of the overall NRI on the basis of whether they reported it as a unitless statistic or a percentage.

Results

The predominant reason for citing one of the methodological articles was the computation of NRI estimates ($n = 39$) (**Table 2**). In 2 (5%) articles, only the continuous NRI was computed. In 5 articles, the NRI was used to compare 2 different models instead of the nested addition of 1 or more new risk markers to a simpler model.

Of the 37 articles that computed category-based NRI results, 34 (92%) detailed the cutoffs for the risk categories chosen. The number of risk categories defined in the computation of the NRI varied between 2 and 6, with 3 being the most common number (**Appendix Table 1**). These risk categories were justified in the text, by references, or both ways in 15 (41%) instances and fully matched clinically meaningful categories with clear implications from guidelines in 4 (11%) instances (**Table 2**). For outcomes other than atherosclerotic cardiovascular disease, the rationale for the risk categorization could not be traced in 10 of 12 instances. Another 8 studies on the prediction of various manifestations of cardiovascular disease used cutoffs for the NRI that are the subject of ongoing debate (28, 60, 70, 89, 90)—for example, a 10-year risk cutoff of 6% (rather than 10%) for low risk for coronary heart disease. Fourteen publications applied cutoffs for coronary risk stratification to broader definitions of cardiovascular disease (**Appendix Table 1**).

Among 38 prospective studies that calculated the NRI, 30 (79%) clearly reported the time horizon at which the risk predictions were evaluated. In 7 of 30 (23%) instances where both predicted horizon and observed follow-up were detailed, we could infer that the authors studied a predicted horizon beyond the observed follow-up time (**Table 2**). We identified another 7 studies that used events occurring beyond the predicted horizon in the reclassification analysis.

Nearly all studies reported the overall NRI. Only 11 (28%) articles presented its components—the event NRI and the nonevent NRI—in the results section. However, 25 (68%) presented reclassification tables stratified for events and nonevents (**Table 2**), which allowed for computation of both NRI components by a knowledgeable reader. By combining the components presented in the text and the reclassification tables, we identified 29 (74%) stud-

Table 1. **Formulas and Interpretation of the NRI**

Type	Formula and Interpretation
Category-based NRI*	
Event NRI	Pr(uplevent) − Pr(downlevent) = (number of events classified up − number of events classified down)/number of events The net percentage of persons with the event of interest correctly classified upward Can be interpreted as a percentage with a range of −100% to 100%†
Nonevent NRI	Pr(downlnonevent) − Pr(uplnonevent) = (number of nonevents classified down − number of nonevents classified up)/number of nonevents The net percentage of persons without the event of interest correctly classified downward Can be interpreted as a percentage with a range of −100% to 100%†
Overall NRI	[Pr(uplevent) − Pr(downlevent)] + [Pr(downlnonevent) − Pr(uplnonevent)] = event NRI + nonevent NRI The sum of the net percentages of correctly reclassified persons with and without the event of interest; this statistic is implicitly weighted for the event rate and cannot be interpreted as a percentage Theoretical range is –2 to 2
Continuous NRI‡	
Event NRI	Pr(higherlevent) − Pr(lowerlevent) = (number of events with increased predicted risk − number of events with decreased predicted risk)/number of events The net percentage of persons with the event of interest correctly assigned a higher predicted risk Can be interpreted as a percentage with a range of −100% to 100%†
Nonevent NRI	Pr(lowerlnonevent) − Pr(higherlnonevent) = (number of nonevents with decreased predicted risk − number of nonevents with increased predicted risk)/number of nonevents The net percentage of persons without the event of interest correctly assigned a lower predicted risk Can be interpreted as a percentage with a range of −100% to 100%†
Overall NRI	[Pr(higherlevent) − Pr(lowerlevent)] + [Pr(lowerlnonevent) − Pr(higherlnonevent)] = event NRI + nonevent NRI The sum of the net percentages of persons with and without the event of interest correctly assigned a different predicted risk; this statistic is implicitly weighted for the event rate and cannot be interpreted as a percentage Theoretical range is −2 to 2

NRI = net reclassification improvement; Pr = probability.
* Assumes that clinically meaningful categories of predicted risk can be defined.
† Negative percentages are interpreted as a worsening in risk classification (i.e., the number of incorrectly reclassified events [or nonevents] exceeds the number of correctly reclassified events [or nonevents]).
‡ Does not consider any categorization.

Table 2. **Results From the Literature Review on Reporting of the NRI**

Reporting of NRI Feature	Studies, *n* (%)
Reason for citing methodological article on NRI	
Claimed to have calculated NRI	39 (58)*
Discussed NRI results from previous analysis	4 (6)*
Suggested alternative methods for quantifying predictive abilities	16 (24)*
Computed other (non-NRI) measures elaborated on in this article	8 (12)*
Risk categorization	
Only continuous (category-free) NRI computed	2 (5)†
Categorization for computing NRI detailed	34 (92)‡
Categorization for computing NRI justified in text	10 (27)‡
Reference given for NRI categorization	14 (38)‡
Categorization for computing NRI corresponded to diagnostic or therapeutic implications in clinical guidelines	4 (11)‡
Time horizon and follow-up	
Predicted horizon detailed	30 (79)§
Observed follow-up detailed (mean, median, or maximum)	37 (97)§
Predicted time horizon longer than observed follow-up	7 (23)‖
Components¶	
Overall NRI	36 (92)†
Event NRI and nonevent NRI in text or tables	11 (28)†
Reclassification table for main findings	25 (68)‡
Unit¶	
Reported as a percentage	24 (67)**
Interpreted as a percentage or proportion	8 (22)**

NRI = net reclassification improvement.
* Of all 67 publications included in the literature review.
† Of 39 studies that calculated the NRI.
‡ Of 37 studies that calculated the category-based NRI.
§ Of 38 prospective studies that calculated the NRI.
‖ Of 30 prospective studies that calculated the NRI and detailed the predicted horizon and follow-up.
¶ **Table 1** provides more details.
** Of 36 studies that reported the overall NRI.

ies with information on the event NRI and nonevent NRI presented for at least 1 reclassification analysis. Of note, 1 study claimed to have calculated the NRI, but no such results could be traced. Another study presented *P* values but no point estimates of the NRI.

Of the 36 studies presenting estimates of the overall NRI, 24 (67%) expressed it as a percentage (**Table 2**). Eight (22%) articles in our review interpreted the overall NRI as a percentage or proportion of the entire study population that was correctly reclassified or used similar wording, such as interpreting an overall NRI of 0.29 as " . . . 29% of patients were correctly reclassified . . . " (17, 39).

NRI Computation, Components, and Interpretation

Predicted Time Horizons and Follow-up

When prospective data are involved, such as cardiovascular events occurring during follow-up, the time horizon used to calculate the predicted risks should be clear. Because virtually every prospective study has some loss to follow-up, it is important to adequately handle observations with incomplete follow-up in the analysis. In our review, we found that studies published shortly after the introduction of the NRI often did not report how incomplete follow-up was handled. Some studies classified censored observations as nonevents ("naive extrapolation") or excluded persons with incomplete follow-up. Better methods have been proposed to limit loss of useful information, including Kaplan–Meier estimates of the expected number of events and nonevents ("prospective NRI") (20, 78) and inverse-probability weighting (91). Similarly, not every study has sufficient follow-up available for the predicted time horizons used in clinical guidelines (for example, 10-year risk for coronary heart disease [89]). In the articles we reviewed, authors made various attempts to overcome this problem, such as using Weibull extrapolation (48, 53), adjusting the predicted risk cutoffs by the ratio of actual to desired follow-up (24), or extrapolating the observed rates on the Kaplan–Meier survival estimates to the predicted time horizon for presentation purposes (22).

Risk Categories

The NRI was introduced with the example of the added value of high-density lipoprotein cholesterol level to coronary risk prediction in the Framingham Heart Study (15). Current clinical guidelines on primary prevention of cardiovascular disease recommend clear cutoffs for initiation of statin treatment (2, 3, 89, 90). These recommendations are supported by cost-effectiveness analyses. The NRI captures the change in a person's predicted risk that crosses one of such cutoffs and thus translates into a clinically meaningful change in treatment recommendations.

Our review of the literature confirms the findings of Tzoulaki and colleagues: Selected risk cutoffs are generally poorly motivated and rarely correspond to therapeutic implications. Both shortcomings have been shown to yield significantly higher NRI estimates (16, 81). In some cases, the existing clinical cutoffs may result in limited reclassification. For example, in a study of a population at very low risk for cardiovascular disease, only a small number of participants would be considered to be at high risk; therefore, few will cross the recommended risk thresholds after the addition of a new marker (92). Using the existing cutoffs illustrates the limited utility of a new marker in real-life application to such a low-risk population. Choosing a priori clinically meaningful cutoffs has been frequently emphasized (15, 16, 19, 20, 60, 63, 81, 92–98). In addition, the estimates of the NRI and its components increase with the number of categories (95, 99). Limiting analysis to clinically meaningful categories will forestall authors from presenting results from the cutoffs with the highest magnitude of NRI in their data. Moreover, consistent use of cutoffs enhances comparability of results on the same

markers between studies provided that the same outcome definition and time horizons are used.

Although many risk prediction algorithms are described in the medical literature, a limited number of clinical guidelines outside the field of cardiology explicitly recommend risk thresholds for use in clinical practice. In the fields where meaningful cutoffs are lacking or evolving, various options have been suggested to overcome this problem. Each has its own caveats. First, in some cases, classification thresholds exist for related outcomes. For example, a 20% 10-year risk for "hard coronary heart disease" corresponds to a 25% 10-year risk for "total coronary heart disease" (100). In these situations, a conversion factor based on the ratio of event rates—in this example, a ratio of 1.25—can be used to translate cutoffs from one application to another. Such conversion assumes that the associated clinical implications are similar for the different outcome definitions, which may not always be true. For example, the protective effect of statins on the occurrence of cardiovascular manifestations other than coronary heart disease, such as heart failure, may be less (101). Similarly, conversion factors can be used to define risk cutoffs for different predicted time horizons (for example, 30- vs. 10-year risks [102]). In the absence of published conversion factors, the data under study can be examined to define the relative occurrence of the outcomes. Second, some researchers have suggested defining risk categories based on the event rate. A cutoff equal to the event rate would be used for binary classification, and cutoffs equal to half the event rate, the event rate, and twice the event rate would be used when more than 2 categories are desired (99, 103). Such cutoffs, however, have no direct clinical interpretation. The appropriateness of risk cutoffs should be related to the anticipated use of the prediction model. As an example, myocardial infarction risk thresholds for a model used to select patients with chest pain for early discharge from an emergency department will be much lower than those for a model used to identify patients with chest pain who will benefit from early invasive coronary angiography. Third, the continuous NRI was introduced as an alternative in the absence of any categorization (**Table 1**) (20). However, it does not quantify the clinical impact of risk reclassification (see Limitations and Controversies). The relation between cutoffs and the risk distribution in the data can be elegantly visualized in reclassification graphs with superimposed cutoffs (**Appendix Figure 2**, available at www.annals.org).

Case–Control Studies

Because of cost and feasibility, the predictive value of new biomarkers is often studied in subsets of persons with events and nonevents from larger prospective studies, especially when the event rates are low. The NRI can be used in both cohort studies and (nested) case–control studies (20). In the latter, the researcher determines the ratio of events (cases) to nonevents (controls) by selective oversampling of cases, which implies artificial weighting by the investigators (43). This should not lead to different estimates in magnitude of the NRI compared with results derived from a full cohort provided that the cases and controls are randomly selected (20, 104). However, difficulties arise when selected controls are not representative of the entire underlying subset they were drawn from, as in the case when matching on certain risk factors (even as simple as age and sex) is done (104–106). This can be overcome by weighting for the inverse of the sampling probability for cases and controls (101, 104).

Components and Interpretation

Although the article that introduced the NRI recommended reporting the components of the overall NRI (15), we noticed in our review that a limited number of studies did so. The components are easier to interpret than the combined number: When only 1 cutoff is being evaluated, the event NRI equals the improvement in sensitivity and the nonevent NRI equals the improvement in specificity (15). The NRI components then express the net percentages of persons with or without events correctly reclassified (**Table 1**). Negative percentages for the components are interpreted as a net worsening in risk classification. The overall NRI is the sum of these 2 underlying components; as a result, an identical point estimate of this statistic may have different interpretations depending on its components (62, 93). Large positive values of the event NRI indicate that the investigated marker aids in the detection of persons with the outcome of interest. This enables clinicians to initiate targeted treatment and thereby prevent events. On the other hand, an overall NRI driven by the nonevent NRI indicates the marker's property of correctly decreasing risk estimates for nonevents and is thus useful for reducing overtreatment. However, such markers will have limited contribution to decreasing the burden of disease. This illustrates the difficulty of interpreting the overall NRI without knowledge of its components (107). Although it is tempting to do so, the overall NRI cannot be interpreted as the "net percentage of persons correctly reclassified" in a straightforward manner (48) because of the implicit weighting by the event rate: The overall NRI is the sum of 2 fractions with different denominators (the number of events and nonevents) (17). Such misinterpretations may have contributed to the popularity of the overall NRI, which therefore should not be presented as a percentage but as a unitless statistic (17). Moreover, the components of the overall NRI may be reasonably well-interpretable, whereas their sum is less so because of the implicit weighting related to the event rate (the costs of misclassification are assumed to be proportional to the odds of nonevents) (**Table 1**) (108).

As with most summary statistics, the NRI should not be interpreted on its own but in the context of complementary statistical measures. If a marker is not associated with the outcome or does not yield an increase in the

AUC, a positive NRI should not be expected (94). In rare instances where this does occur, random chance or differences in calibration between the models are the most likely causes. Also, presenting reclassification tables (in tabular or graphical form) will aid in the broader interpretation of summarized reclassification statistics (**Appendix Figure 2** and **Appendix Table 3**, available at www.annals.org).

Limitations and Controversies

Miscalibration

Unlike such rank-based statistics as the AUC, the NRI is affected by miscalibration of a model (that is, the average predicted risk is not close to the event rate) (108–110). Systematic miscalibration does not occur when the performance of models is assessed on the same data set that was used to develop them but is often present when prediction models are validated in other populations. A well-recognized example of this phenomenon is the application of the Framingham cardiovascular risk models to European populations (111–113). When performing a head-to-head comparison between a Framingham function (using the published coefficients and baseline hazard) and a new risk function developed from the data under study, one might find an NRI that favors the new model and no difference in the AUCs (114, 115). This discrepancy can be avoided by deriving both the reference model and the model including the marker under investigation from the same data set that is used to compute the NRIs or by recalibrating both models in case of independent validation (116).

The traditional Hosmer–Lemeshow goodness-of-fit test is strongly dependent on the sample size of the study (117). Therefore, calibration might better be assessed graphically in a plot with predicted risks on the horizontal axis and observed event rates on the vertical axis, as in Koller and colleagues' example (54). For perfectly calibrated models, the plot forms a diagonal line where the observed event rates equal the predicted risks. Such graphs can show systematic underestimation or overestimation as well as issues of overfitting (which can be quantified using the calibration slope [118]).

Classification or Reclassification?

Some researchers have argued that before addressing the issue of reclassification, one should first focus on risk classification and examine the margins of a reclassification table (43). Accordingly, examining reclassification is useful only to the extent to which it quantifies change in the size of these margins. This might be of particular relevance in head-to-head comparisons of nonnested models with substantial reclassification (that is, if the 2 models have low correlation). In this case, as in the example shown by Koller and colleagues (54), knowing how many persons are classified in the clinically relevant subgroups is of greater interest than the exact reclassification within the inner cells of the table (93, 96). Therefore, when choosing between competing models for clinical practice, the main question is which one leads to better classification (which relates to both discrimination and calibration of the models). On the other hand, when the focus is primarily on the potential of a new marker, the improvements in discrimination and subsequent risk reclassification that it can induce are of primary interest.

Continuous NRI

The continuous NRI was originally proposed to overcome the problem of selecting categories in applications where they do not naturally exist (20). It does not require any risk categorization and considers all changes in predicted risk for all events and nonevents. This has several consequences. First, most changes in predicted risk do not translate into changes in clinical management; for example, a middle-aged woman whose 10-year predicted coronary risk doubles from 1% to 2% will probably not be treated differently (92, 119). Therefore, the interpretation of the continuous NRI is different from that of the category-based NRI (**Table 1**) (11). Second, when the addition of a normally distributed marker is considered, the continuous NRI is less affected by the performance of the baseline model and can therefore be seen as a rescaling of the measures of association (for example, an odds ratio of 1.65 per SD corresponds to a continuous NRI of 0.395) (11, 21). Consequently, the continuous NRI is often positive for relatively weak markers (11). Moreover, it is strongly affected by miscalibration, especially in the setting of external validation (110).

As such, the continuous NRI is less suitable for head-to-head comparisons of competing models unless these models have been developed from the same data or are correctly calibrated. The most appealing application of the continuous NRI comes in quantifying the effect of an added predictor in settings where the distributions of other risk factors may not be representative of the population (120). For example, when the same marker for coronary risk prediction is evaluated in 2 populations, one with wide and the other with narrow age ranges, the conclusions about its usefulness might be different if based on the increment in AUC (12). The continuous NRI, however, would give a consistent message and is therefore marker-descriptive rather than model-descriptive. Furthermore, its magnitude should be assessed on its own scale (11) and should not be compared with that of the category-based version.

Clinical NRI

Reclassification measures, including the NRI, can be used to evaluate markers in specific subgroups of the study population defined by the reference model. Specifically, the added value of new risk markers may be of greater importance in persons with a risk categorization that has more uncertainty about the clinical implications (for example, persons at intermediate risk for coronary heart disease [33, 48, 62, 72, 73, 86]). This "clinical NRI" (121), however, has been found to be biased because it does not take into

account incorrect reclassification from other risk categories into the intermediate-risk category (62). Adding randomly generated noninformative markers to existing prediction models leads to positive clinical NRIs more frequently than expected on the basis of chance (99, 122). A method for correcting this systematic overestimation has been published (122).

Decision Analytic Measures

The overall NRI implicitly weights for the event rate, p, with $1/p$ and $1/(1 - p)$ serving as costs for false-negative results (events classified downward) and false-positive results (nonevents classified upward), respectively (108, 123). However, a different weighting of false-positive and false-negative results is often more clinically appropriate (98). This can readily be incorporated in a weighted version of the NRI if the event NRI and nonevent NRI are presented separately or when a reclassification table is provided (20, 124). In its broadest form, the weighted NRI can be interpreted as the average savings (for example, in dollars or quality-adjusted life-years) per person resulting from using the new model instead of the old one (20).

The weighted NRI is a decision analytic measure and is mathematically a transformation of changes in net benefit and relative utility (124). These measures use the harm–benefit ratio to define an optimum decision threshold for binary classification as high risk versus low risk (125). The harm–benefit ratio also defines the weights of true-positive and false-positive classifications to calculate a single summary measure (124–126). However, the use of such decision analytic measures is limited by the fact that weights for harms and benefits are not firmly established in most fields of medicine (126), although a range of decision thresholds can be considered in a sensitivity analysis with visualization in a "decision curve" (127).

The nonweighted category-based NRI analysis is regarded as an early-stage analysis in the evaluation of new markers or prediction models. For assessment of the potential clinical utility of promising markers, decision analytic approaches are needed in the next step, after the NRI analyses but before a full formal cost-effectiveness analysis that incorporates changes in costs and clinical outcomes in more detail (13).

Table 3. **Recommendations for Reporting the NRI**

Article Section	Recommendation
Methods	
Type of NRI	Specify the type of NRI computed in the methods section of the manuscript (category-based and/or continuous NRI).
Follow-up	Specify the horizon of risk prediction if the NRI was computed for prognostic evaluations (e.g., 10-y risk). Describe how censored observations (e.g., persons lost to follow-up before the specified horizon) were handled. Use the event status at the predicted time horizon and ignore events occurring beyond the predicted time horizon (e.g., when predicting 10-y risk for CHD, consider participants with a myocardial infarction occurring after 10 y of follow-up as nonevents).
Cutoffs	For category-based NRI, the categorization should ideally have clear consequences in clinical practice. When possible, give references to formal clinical guidelines used to define the risk categories for the computation of the NRI. If alternative cutoffs were used, clearly motivate them.
Results	
Components	Report the NRIs for events and nonevents separately. Reclassification tables stratified for persons with and without the event of interest are informative beyond the NRI (e.g., **Appendix Table 3**).
Unit	The event and nonevent NRIs can be presented as percentages. However, the overall NRI has no units and should therefore not be presented as a percentage (**Table 1**).
Calibration	Provide information on the calibration of the models being compared.
Discussion	
Interpretation	The components of the overall NRI can be interpreted as a net percentage of the number of persons with or without events. However, the overall NRI should not be interpreted as a net percentage of the study population correctly reclassified.
Comparisons	Do not draw strong comparative conclusions based on direct comparisons of NRIs obtained in different populations or using different outcomes or cutoffs.

CHD = coronary heart disease; NRI = net reclassification improvement.

Recommendations

In our literature review, we encountered several common flaws in the presentation and interpretation of the NRI and insufficient documentation of the computational methods. On the basis of our observations, we make the following recommendations for clinical research (18) (**Table 3**).

Clearly defining which type of NRI is used is essential because their applicability and relevance vary substantially. The most appropriate NRI type and cut points depend on several factors, as discussed in this review. We recommend separate reporting of the NRI for events and nonevents in all circumstances. Also, the sum of the NRI components should not be interpreted as a percentage. If authors choose to present the category-based NRI, they should discuss the implied costs of misclassification by the event rate. The cutoffs selected for the NRI analyses should preferably match risk thresholds that have clear clinical implications or can be motivated on clinical grounds. In general, the category-based NRI is directly applicable in settings where meaningful risk categories exist and models are well-calibrated. If either of these conditions is not satisfied, one must carefully determine what information the NRI offers and whether it can be interpreted meaningfully. Using cut-

offs that have no direct clinical meaning impedes the interpretation of the category-based NRI. Several methods have been proposed to define cut points in situations where meaningful thresholds do not exist, but each has its own caveats. Presenting graphical displays similar to a decision curve (127) for a range of cutoffs could be considered as an alternative. The continuous NRI can be recommended in only a few settings, including those where the primary focus is on the strength of the marker rather than model performance. Authors must be careful not to overinterpret the magnitude of the continuous NRI, which is usually much larger than that of the category-based NRI, and must ascertain that the models are well-calibrated. Finally, for mathematical reasons, we recommend against calculating *P* values for any of the forms of the NRI when the contribution of a new marker is being evaluated (128, 129). Instead, after a marker has been shown to be statistically significantly associated with the outcome, only CIs for the NRI should be presented.

Our recommendations are meant to improve completeness, transparency, and clinical relevance of research involving risk reclassification. However, because the scientific debate on the NRI and related performance measures is ongoing, our recommendations may be subject to advances or additions in the future.

From Erasmus MC - University Medical Center Rotterdam, Rotterdam, the Netherlands, and Duke University, Durham, North Carolina.

Grant Support: By the Netherlands Organisation for Health Research and Development (ZonMw) and the Netherlands Organisation for Scientific Research (NWO) (ZonMw HTA grant 80-82500-98-10208; Vici grant 918-76-619) and the Center for Translational Molecular Medicine (PCMM project grant). The funding sources had no role in the design or conduct of the study; the collection, management, analysis, or interpretation of the data; or the preparation, review, or approval of the manuscript.

Potential Conflicts of Interest: Disclosures can be viewed at www.acponline.org/authors/icmje/ConflictOfInterestForms.do?msNum=M13-1522.

Requests for Single Reprints: Maarten J.G. Leening, MD, MSc, Department of Epidemiology, Erasmus MC - University Medical Center Rotterdam, PO Box 2040, 3000 CA Rotterdam, the Netherlands; e-mail, m.leening@erasmusmc.nl.

Current author addresses and author contributions are available at www.annals.org.

References

1. **Kannel WB, Dawber TR, Kagan A, Revotskie N, Stokes J 3rd.** Factors of risk in the development of coronary heart disease—six year follow-up experience. The Framingham Study. Ann Intern Med. 1961;55:33-50. [PMID: 13751193]

2. **Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults.** Executive Summary of The Third Report of The National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, And Treatment of High Blood Cholesterol In Adults (Adult Treatment Panel III). JAMA. 2001;285:2486-97. [PMID: 11368702]

3. **Perk J, De Backer G, Gohlke H, Graham I, Reiner Z, Verschuren M, et al; European Association for Cardiovascular Prevention & Rehabilitation (EACPR).** European Guidelines on cardiovascular disease prevention in clinical practice (version 2012). The Fifth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of nine societies and by invited experts). Eur Heart J. 2012;33:1635-701. [PMID: 22555213]

4. **Hamm CW, Bassand JP, Agewall S, Bax J, Boersma E, Bueno H, et al; ESC Committee for Practice Guidelines.** ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation: The Task Force for the management of acute coronary syndromes (ACS) in patients presenting without persistent ST-segment elevation of the European Society of Cardiology (ESC). Eur Heart J. 2011;32:2999-3054. [PMID: 21873419]

5. **Visvanathan K, Chlebowski RT, Hurley P, Col NF, Ropka M, Collyar D, et al; American Society of Clinical Oncology.** American Society of Clinical Oncology clinical practice guideline update on the use of pharmacologic interventions including tamoxifen, raloxifene, and aromatase inhibition for breast cancer risk reduction. J Clin Oncol. 2009;27:3235-58. [PMID: 19470930]

6. **Munshi NC, Anderson KC, Bergsagel PL, Shaughnessy J, Palumbo A, Durie B, et al; International Myeloma Workshop Consensus Panel 2.** Consensus recommendations for risk stratification in multiple myeloma: report of the International Myeloma Workshop Consensus Panel 2. Blood. 2011;117:4696-700. [PMID: 21292777]

7. **Worth LJ, Lingaratnam S, Taylor A, Hayward AM, Morrissey S, Cooney J, et al; Australian Consensus Guidelines 2011 Steering Committee.** Use of risk stratification to guide ambulatory management of neutropenic fever. Australian Consensus Guidelines 2011 Steering Committee. Intern Med J. 2011;41:82-9. [PMID: 21272172]

8. **Bates SM, Jaeschke R, Stevens SM, Goodacre S, Wells PS, Stevenson MD, et al; American College of Chest Physicians.** Diagnosis of DVT: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. Chest. 2012;141:e351S-418S. [PMID: 22315267]

9. **Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P.** Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. Am J Epidemiol. 2004;159:882-90. [PMID: 15105181]

10. **Cook NR.** Use and misuse of the receiver operating characteristic curve in risk prediction. Circulation. 2007;115:928-35. [PMID: 17309939]

11. **Pencina MJ, D'Agostino RB, Pencina KM, Janssens ACJW, Greenland P.** Interpreting incremental value of markers added to risk prediction models. Am J Epidemiol. 2012;176:473-81. [PMID: 22875755]

12. **Austin PC, Steyerberg EW.** Predictive accuracy of risk factors and markers: a simulation study of the effect of novel markers on different performance measures for logistic regression models. Stat Med. 2013;32:661-72. [PMID: 22961910]

13. **Hlatky MA, Greenland P, Arnett DK, Ballantyne CM, Criqui MH, Elkind MSV, et al; American Heart Association Expert Panel on Subclinical Atherosclerotic Diseases and Emerging Risk Factors and the Stroke Council.** Criteria for evaluation of novel markers of cardiovascular risk: a scientific statement from the American Heart Association. Circulation. 2009;119:2408-16. [PMID: 19364974]

14. **Cook NR, Buring JE, Ridker PM.** The effect of including C-reactive protein in cardiovascular risk prediction models for women. Ann Intern Med. 2006;145:21-9. [PMID: 16818925]

15. **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;27:157-72. [PMID: 17569110]

16. **Tzoulaki I, Liberopoulos G, Ioannidis JPA.** Use of reclassification for assessment of improved prediction: an empirical evaluation. Int J Epidemiol. 2011;40:1094-105. [PMID: 21325392]

17. **Leening MJG, Steyerberg EW.** Fibrosis and mortality in patients with dilated cardiomyopathy [Letter]. JAMA. 2013;309:2547-8. [PMID: 23800926]

18. **Lilford RJ, Richardson A, Stevens A, Fitzpatrick R, Edwards S, Rock F, et al.** Issues in methodological research: perspectives from researchers and commissioners. Health Technol Assess. 2001;5:1-57. [PMID: 11368832]

19. **Pencina MJ, D'Agostino RB, Vasan RS.** Statistical methods for assessment of added usefulness of new biomarkers. Clin Chem Lab Med. 2010;48:1703-11. [PMID: 20716010]

20. **Pencina MJ, D'Agostino RB Sr, Steyerberg EW.** Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. Stat Med. 2011;30:11-21. [PMID: 21204120]
21. **Pencina MJ, D'Agostino RB Sr, Demler OV.** Novel metrics for evaluating improvement in discrimination: net reclassification and integrated discrimination improvement for normal variables and nested models. Stat Med. 2012;31:101-13. [PMID: 22147389]
22. **Cook NR, Ridker PM.** Advances in measuring the effect of individual predictors of cardiovascular risk: the role of reclassification measures. Ann Intern Med. 2009;150:795-802. [PMID: 19487714]
23. **Adabag AS, Therneau TM, Gersh BJ, Weston SA, Roger VL.** Sudden death after myocardial infarction. JAMA. 2008;300:2022-9. [PMID: 18984889]
24. **Auer R, Bauer DC, Marques-Vidal P, Butler J, Min LJ, Cornuz J, et al; Health ABC Study.** Association of major and minor ECG abnormalities with coronary heart disease events. JAMA. 2012;307:1497-505. [PMID: 22496264]
25. **Breteler MMB.** Mapping out biomarkers for Alzheimer disease [Editorial]. JAMA. 2011;305:304-5. [PMID: 21245188]
26. **Buckley DI, Fu R, Freeman M, Rogers K, Helfand M.** C-reactive protein as a risk factor for coronary heart disease: a systematic review and meta-analyses for the U.S. Preventive Services Task Force. Ann Intern Med. 2009;151:483-95. [PMID: 19805771]
27. **Chou R, Arora B, Dana T, Fu R, Walker M, Humphrey L.** Screening asymptomatic adults with resting or exercise electrocardiography: a review of the evidence for the U.S. Preventive Services Task Force. Ann Intern Med. 2011; 155:375-85. [PMID: 21930855]
28. **Cook NR.** Biomarkers for prediction of cardiovascular events [Letter]. JAMA. 2009;302:2089. [PMID: 19920231]
29. **Cornelis MC, Qi L, Zhang C, Kraft P, Manson J, Cai T, et al.** Joint effects of common genetic variants on the risk for type 2 diabetes in U.S. men and women of European ancestry. Ann Intern Med. 2009;150:541-50. [PMID: 19380854]
30. **de Boer IH, Levin G, Robinson-Cohen C, Biggs ML, Hoofnagle AN, Siscovick DS, et al.** Serum 25-hydroxyvitamin D concentration and risk for major clinical disease events in a community-based population of older adults: a cohort study. Ann Intern Med. 2012;156:627-34. [PMID: 22547472]
31. **de Lemos JA, Drazner MH, Omland T, Ayers CR, Khera A, Rohatgi A, et al.** Association of troponin T detected with a highly sensitive assay and cardiac structure and mortality risk in the general population. JAMA. 2010;304:2503-12. [PMID: 21139111]
32. **deFilippi CR, de Lemos JA, Christenson RH, Gottdiener JS, Kop WJ, Zhan M, et al.** Association of serial measures of cardiac troponin T using a sensitive assay with incident heart failure and cardiovascular mortality in older adults. JAMA. 2010;304:2494-502. [PMID: 21078811]
33. **den Ruijter HM, Peters SAE, Anderson TJ, Britton AR, Dekker JM, Eijkemans MJ, et al.** Common carotid intima-media thickness measurements in cardiovascular risk prediction: a meta-analysis. JAMA. 2012;308:796-803. [PMID: 22910757]
34. **Devereaux PJ, Chan MTV, Alonso-Coello P, Walsh M, Berwanger O, Villar JC, et al; Vascular Events In Noncardiac Surgery Patients Cohort Evaluation (VISION) Study Investigators.** Association between postoperative troponin levels and 30-day mortality among patients undergoing noncardiac surgery. JAMA. 2012;307:2295-304. [PMID: 22706835]
35. **Di Angelantonio E, Gao P, Pennells L, Kaptoge S, Caslake M, Thompson A, et al; Emerging Risk Factors Collaboration.** Lipid-related markers and cardiovascular disease prediction. JAMA. 2012;307:2499-506. [PMID: 22797450]
36. **Eddy DM, Adler J, Patterson B, Lucas D, Smith KA, Morris M.** Individualized guidelines: the potential for increasing quality and reducing costs. Ann Intern Med. 2011;154:627-34. [PMID: 21536939]
37. **Farooq V, van Klaveren D, Steyerberg EW, Meliga E, Vergouwe Y, Chieffo A, et al.** Anatomical and clinical characteristics to guide decision making between coronary artery bypass surgery and percutaneous coronary intervention for individual patients: development and validation of SYNTAX Score II. Lancet. 2013; 381:639-50. [PMID: 23439103]
38. **Fonarow GC, Pan W, Saver JL, Smith EE, Reeves MJ, Broderick JP, et al.** Comparison of 30-day mortality models for profiling hospital performance in acute ischemic stroke with vs without adjustment for stroke severity. JAMA. 2012;308:257-64. [PMID: 22797643]
39. **Gulati A, Jabbour A, Ismail TF, Guha K, Khwaja J, Raza S, et al.** Association of fibrosis with mortality and sudden cardiac death in patients with nonischemic dilated cardiomyopathy. JAMA. 2013;309:896-908. [PMID: 23462786]
40. **Helfand M, Buckley DI, Freeman M, Fu R, Rogers K, Fleming C, et al.** Emerging risk factors for coronary heart disease: a summary of systematic reviews conducted for the U.S. Preventive Services Task Force. Ann Intern Med. 2009; 151:496-507. [PMID: 19805772]
41. **Hingorani AD, Psaty BM.** Primary prevention of cardiovascular disease: time to get more or less personal? JAMA. 2009;302:2144-5. [PMID: 19920239]
42. **Hlatky MA.** Framework for evaluating novel risk markers [Editorial]. Ann Intern Med. 2012;156:468-9. [PMID: 22431679]
43. **Janes H, Pepe MS, Gu W.** Assessing the value of risk predictions by using risk stratification tables. Ann Intern Med. 2008;149:751-60. [PMID: 19017593]
44. **Janssens ACJW, Ioannidis JPA, van Duijn CM, Little J, Khoury MJ; GRIPS Group.** Strengthening the reporting of genetic risk prediction studies: The GRIPS Statement. Ann Intern Med. 2011;154:421-5. [PMID: 21403077]
45. **Kaptoge S, Di Angelantonio E, Lowe G, Pepys MB, Thompson SG, Collins R, et al; Emerging Risk Factors Collaboration.** C-reactive protein concentration and risk of coronary heart disease, stroke, and mortality: an individual participant meta-analysis. Lancet. 2010;375:132-40. [PMID: 20031199]
46. **Kaptoge S, Di Angelantonio E, Pennells L, Wood AM, White IR, Gao P, et al; Emerging Risk Factors Collaboration.** C-reactive protein, fibrinogen, and cardiovascular disease prediction. N Engl J Med. 2012;367:1310-20. [PMID: 23034020]
47. **Kathiresan S, Melander O, Anevski D, Guiducci C, Burtt NP, Roos C, et al.** Polymorphisms associated with cholesterol and risk of cardiovascular events. N Engl J Med. 2008;358:1240-9. [PMID: 18354102]
48. **Kavousi M, Elias-Smale S, Rutten JHW, Leening MJG, Vliegenthart R, Verwoert GC, et al.** Evaluation of newer risk markers for coronary heart disease risk classification: a cohort study. Ann Intern Med. 2012;156:438-44. [PMID: 22431676]
49. **Keller T, Zeller T, Ojeda F, Tzikas S, Lillpopp L, Sinning C, et al.** Serial changes in highly sensitive troponin I assay and early diagnosis of myocardial infarction. JAMA. 2011;306:2684-93. [PMID: 22203537]
50. **Kengne AP, Echouffo-Tcheugui JB, Sobngwi E.** Coronary artery calcium for guiding statin treatment [Letter]. Lancet. 2012;379:312. [PMID: 22284653]
51. **Khera AV, Cuchel M, de la Llera-Moya M, Rodrigues A, Burke MF, Jafri K, et al.** Cholesterol efflux capacity, high-density lipoprotein function, and atherosclerosis. N Engl J Med. 2011;364:127-35. [PMID: 21226578]
52. **Kim WR, Biggins SW, Kremers WK, Wiesner RH, Kamath PS, Benson JT, et al.** Hyponatremia and mortality among patients on the liver-transplant waiting list. N Engl J Med. 2008;359:1018-26. [PMID: 18768945]
53. **Kivimäki M, Batty GD, Hamer M, Ferrie JE, Vahtera J, Virtanen M, et al.** Using additional information on working hours to predict coronary heart disease: a cohort study. Ann Intern Med. 2011;154:457-63. [PMID: 21464347]
54. **Koller MT, Leening MJG, Wolbers M, Steyerberg EW, Hunink MGM, Schoop R, et al.** Development and validation of a coronary risk prediction model for older U.S. and European persons in the Cardiovascular Health Study and the Rotterdam Study. Ann Intern Med. 2012;157:389-97. [PMID: 22986376]
55. **Lubitz SA, Yin X, Fontes JD, Magnani JW, Rienstra M, Pai M, et al.** Association between familial atrial fibrillation and risk of new-onset atrial fibrillation. JAMA. 2010;304:2263-9. [PMID: 21076174]
56. **Lyssenko V, Jonsson A, Almgren P, Pulizzi N, Isomaa B, Tuomi T, et al.** Clinical risk factors, DNA variants, and the development of type 2 diabetes. N Engl J Med. 2008;359:2220-32. [PMID: 19020324]
57. **Manolio TA.** Genomewide association studies and assessment of the risk of disease. N Engl J Med. 2010;363:166-76. [PMID: 20647212]
58. **Martínez ME, Thompson P, Messer K, Ashbeck EL, Lieberman DA, Baron JA, et al.** One-year risk for advanced colorectal neoplasia: U.S. versus U.K. risk-stratification guidelines. Ann Intern Med. 2012;157:856-64. [PMID: 23247939]
59. **Matsushita K, Mahmoodi BK, Woodward M, Emberson JR, Jafar TH, Jee SH, et al; Chronic Kidney Disease Prognosis Consortium.** Comparison of risk prediction using the CKD-EPI equation and the MDRD study equation for estimated glomerular filtration rate. JAMA. 2012;307:1941-51. [PMID: 22570462]
60. **McEvoy JW.** Coronary artery calcium score and cardiovascular event prediction [Letter]. JAMA. 2010;304:741-2. [PMID: 20716732]
61. **Meigs JB, Shrader P, Sullivan LM, McAteer JB, Fox CS, Dupuis J, et al.** Genotype score in addition to common risk factors for prediction of type 2 diabetes. N Engl J Med. 2008;359:2208-19. [PMID: 19020323]

62. **Melander O, Newton-Cheh C, Almgren P, Hedblad B, Berglund G, Engström G, et al.** Novel and conventional biomarkers for prediction of incident cardiovascular events in the community. JAMA. 2009;302:49-57. [PMID: 19567439]
63. **Melander O, Newton-Cheh C, Wang TJ.** Biomarkers for prediction of cardiovascular events—reply. JAMA. 2009;302:2090.
64. **Omland T, de Lemos JA, Sabatine MS, Christophi CA, Rice MM, Jablonski KA, et al; Prevention of Events with Angiotensin Converting Enzyme Inhibition (PEACE) Trial Investigators.** A sensitive cardiac troponin T assay in stable coronary artery disease. N Engl J Med. 2009;361:2538-47. [PMID: 19940289]
65. **Palomaki GE, Melillo S, Bradley LA.** Association between 9p21 genomic markers and heart disease: a meta-analysis. JAMA. 2010;303:648-56. [PMID: 20159873]
66. **Paynter NP, Chasman DI, Buring JE, Shiffman D, Cook NR, Ridker PM.** Cardiovascular disease risk prediction with and without knowledge of genetic variation at chromosome 9p21.3. Ann Intern Med. 2009;150:65-72. [PMID: 19153409]
67. **Paynter NP, Chasman DI, Paré G, Buring JE, Cook NR, Miletich JP, et al.** Association between a literature-based genetic risk score and cardiovascular events in women. JAMA. 2010;303:631-7. [PMID: 20159871]
68. **Peralta CA, Shlipak MG, Judd S, Cushman M, McClellan W, Zakai NA, et al.** Detection of chronic kidney disease with creatinine, cystatin C, and urine albumin-to-creatinine ratio and association with progression to end-stage renal disease and mortality. JAMA. 2011;305:1545-52. [PMID: 21482744]
69. **Pischon T, Boeing H, Hoffmann K, Bergmann M, Schulze MB, Overvad K, et al.** General and abdominal adiposity and risk of death in Europe. N Engl J Med. 2008;359:2105-20. [PMID: 19005195]
70. **Pletcher MJ, Tice JA, Pignone M.** Modeling cardiovascular disease prevention [Letter]. JAMA. 2010;303:835. [PMID: 20197528]
71. **Polak JF, Pencina MJ, Pencina KM, O'Donnell CJ, Wolf PA, D'Agostino RB Sr.** Carotid-wall intima-media thickness and cardiovascular events. N Engl J Med. 2011;365:213-21. [PMID: 21774709]
72. **Polonsky TS, McClelland RL, Jorgensen NW, Bild DE, Burke GL, Guerci AD, et al.** Coronary artery calcium score and risk classification for coronary heart disease prediction. JAMA. 2010;303:1610-6. [PMID: 20424251]
73. **Ripatti S, Tikkanen E, Orho-Melander M, Havulinna AS, Silander K, Sharma A, et al.** A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses. Lancet. 2010;376:1393-400. [PMID: 20971364]
74. **Rosenberg S, Elashoff MR, Beineke P, Daniels SE, Wingrove JA, Tingley WG, et al; PREDICT (Personalized Risk Evaluation and Diagnosis in the Coronary Tree) Investigators.** Multicenter validation of the diagnostic accuracy of a blood-based gene expression test for assessing obstructive coronary artery disease in nondiabetic patients. Ann Intern Med. 2010;153:425-34. [PMID: 20921541]
75. **Schelbert EB, Cao JJ, Sigurdsson S, Aspelund T, Kellman P, Aletras AH, et al.** Prevalence and prognosis of unrecognized myocardial infarction determined by cardiac magnetic resonance in older adults. JAMA. 2012;308:890-6. [PMID: 22948699]
76. **Schnabel RB, Sullivan LM, Levy D, Pencina MJ, Massaro JM, D'Agostino RB Sr, et al.** Development of a risk score for atrial fibrillation (Framingham Heart Study): a community-based cohort study. Lancet. 2009;373:739-45. [PMID: 19249635]
77. **Selvin E, Steffes MW, Zhu H, Matsushita K, Wagenknecht L, Pankow J, et al.** Glycated hemoglobin, diabetes, and cardiovascular risk in nondiabetic adults. N Engl J Med. 2010;362:800-11. [PMID: 20200384]
78. **Steyerberg EW, Pencina MJ.** Reclassification calculations for persons with incomplete follow-up [Letter]. Ann Intern Med. 2010;152:195-6. [PMID: 20124243]
79. **Tammemägi MC, Katki HA, Hocking WG, Church TR, Caporaso N, Kvale PA, et al.** Selection criteria for lung-cancer screening. N Engl J Med. 2013;368:728-36. [PMID: 23425165]
80. **Tangri N, Stevens LA, Griffith J, Tighiouart H, Djurdjev O, Naimark D, et al.** A predictive model for progression of chronic kidney disease to kidney failure. JAMA. 2011;305:1553-9. [PMID: 21482743]
81. **Tzoulaki I, Liberopoulos G, Ioannidis JPA.** Assessment of claims of improved prediction beyond the Framingham risk score. JAMA. 2009;302:2345-52. [PMID: 19952321]
82. **Wacholder S, Hartge P, Prentice R, Garcia-Closas M, Feigelson HS, Diver WR, et al.** Performance of common genetic variants in breast-cancer risk models. N Engl J Med. 2010;362:986-93. [PMID: 20237344]
83. **Wilson PWF.** Challenges to improve coronary heart disease risk assessment [Editorial]. JAMA. 2009;302:2369-70. [PMID: 19952326]
84. **Wormser D, Kaptoge S, Di Angelantonio E, Wood AM, Pennells L, Thompson A, et al; Emerging Risk Factors Collaboration.** Separate and combined associations of body-mass index and abdominal adiposity with cardiovascular disease: collaborative analysis of 58 prospective studies. Lancet. 2011;377:1085-95. [PMID: 21397319]
85. **Wormser D, Di Angelantonio E, Sattar N, Collins R, Thompson S, Danesh J, et al.** Body-mass index, abdominal adiposity, and cardiovascular risk – reply. Lancet. 2011;378:228.
86. **Yeboah J, McClelland RL, Polonsky TS, Burke GL, Sibley CT, O'Leary D, et al.** Comparison of novel risk markers for improvement in cardiovascular risk assessment in intermediate-risk individuals. JAMA. 2012;308:788-95. [PMID: 22910756]
87. **Zethelius B, Berglund L, Sundström J, Ingelsson E, Basu S, Larsson A, et al.** Use of multiple biomarkers to improve the prediction of death from cardiovascular causes. N Engl J Med. 2008;358:2107-16. [PMID: 18480203]
88. **Zoungas S, Patel A, Chalmers J, de Galan BE, Li Q, Billot L, et al; ADVANCE Collaborative Group.** Severe hypoglycemia and risks of vascular events and death. N Engl J Med. 2010;363:1410-8. [PMID: 20925543]
89. **Greenland P, Alpert JS, Beller GA, Benjamin EJ, Budoff MJ, Fayad ZA, et al; American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines.** 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. Circulation. 2010;122:e584-636. [PMID: 21098428]
90. **Mosca L, Benjamin EJ, Berra K, Bezanson JL, Dolor RJ, Lloyd-Jones DM, et al.** Effectiveness-based guidelines for the prevention of cardiovascular disease in women—2011 update: a guideline from the American Heart Association. Circulation. 2011;123:1243-62. [PMID: 21325087]
91. **Cai T, Tian L, Lloyd-Jones DM.** Comparing costs associated with risk stratification rules for t-year survival. Biostatistics. 2011;12:597-609. [PMID: 21415016]
92. **Leening MJG, Cook NR.** Net reclassification improvement: a link between statistics and clinical practice. Eur J Epidemiol. 2013;28:21-3. [PMID: 23291877]
93. **Pepe MS.** Problems with risk reclassification methods for evaluating prediction models. Am J Epidemiol. 2011;173:1327-35. [PMID: 21555714]
94. **Mihaescu R, van Zitteren M, van Hoek M, Sijbrands EJ, Uitterlinden AG, Witteman JCM, et al.** Improvement of risk prediction by genomic profiling: reclassification measures versus the area under the receiver operating characteristic curve. Am J Epidemiol. 2010;172:353-61. [PMID: 20562194]
95. **Mühlenbruch K, Heraclides A, Steyerberg EW, Joost HG, Boeing H, Schulze MB.** Assessing improvement in disease prediction using net reclassification improvement: impact of risk cut-offs and number of risk categories. Eur J Epidemiol. 2013;28:25-33. [PMID: 23179629]
96. **Pepe MS, Janes H.** Commentary: Reporting standards are needed for evaluations of risk reclassification. Int J Epidemiol. 2011;40:1106-8. [PMID: 21571811]
97. **McGeechan K, Macaskill P, Irwig L, Liew G, Wong TY.** Assessing new biomarkers and predictive models for use in clinical practice: a clinician's guide. Arch Intern Med. 2008;168:2304-10. [PMID: 19029492]
98. **Greenland S.** The need for reorientation toward cost-effective prediction: comments on 'Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond' by M. J. Pencina et al., Statistics in Medicine (DOI: 10.1002/sim.2929). Stat Med. 2008;27:199-206. [PMID: 17729377]
99. **Cook NR, Paynter NP.** Performance of reclassification statistics in comparing risk prediction models. Biom J. 2011;53:237-58. [PMID: 21294152]
100. **National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III).** Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. Circulation. 2002;106:3143-421. [PMID: 12485966]
101. **Cook NR, Paynter NP, Eaton CB, Manson JE, Martin LW, Robinson JG, et al.** Comparison of the Framingham and Reynolds Risk scores for global

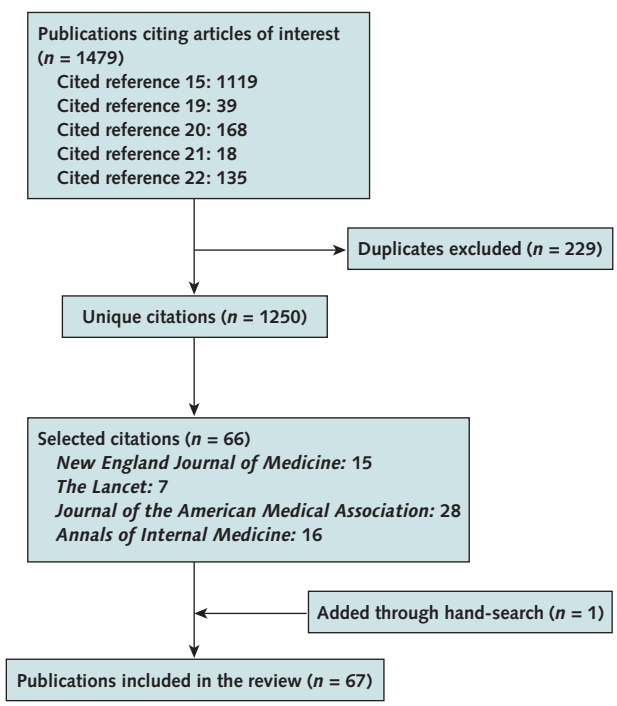
cardiovascular risk prediction in the multiethnic Women's Health Initiative. Circulation. 2012;125:1748-56, S1-11. [PMID: 22399535]
102. **Pencina MJ, D'Agostino RB Sr, Larson MG, Massaro JM, Vasan RS.** Predicting the 30-year risk of cardiovascular disease: the Framingham Heart Study. Circulation. 2009;119:3078-84. [PMID: 19506114]
103. **Takahara M, Katakami N, Kaneto H, Shimomura I.** Risk categorization for calculating net reclassification improvement [Letter]. Eur J Epidemiol. 2013; 28:607-9. [PMID: 23839540]
104. **Ganna A, Reilly M, de Faire U, Pedersen N, Magnusson P, Ingelsson E.** Risk prediction measures for case-cohort and nested case-control designs: an application to cardiovascular disease. Am J Epidemiol. 2012;175:715-24. [PMID: 22396388]
105. **Pepe MS, Fan J, Seymour CW, Li C, Huang Y, Feng Z.** Biases introduced by choosing controls to match risk factors of cases in biomarker research. Clin Chem. 2012;58:1242-51. [PMID: 22730452]
106. **Rothman KJ, Greenland S, Lash TL.** Modern Epidemiology. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008.
107. **Kavousi M, Leening MJG, Witteman JCM.** Markers for prediction of cardiovascular disease risk [Letter]. JAMA. 2012;308:2561. [PMID: 23268505]
108. **Pencina MJ, D'Agostino RB, Sr., D'Agostino RB, Jr., Vasan RS.** Comments on 'Integrated discrimination and net reclassification improvements—Practical advice'. Stat Med. 2008;27:207-12. [PMID: 17569110]
109. **Pepe MS, Feng Z, Gu JW.** Comments on 'Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond' by M. J. Pencina et al., Statistics in Medicine (DOI: 10.1002/sim.2929). Stat Med. 2008;27:173-81. [PMID: 17671958]
110. **Hilden J, Gerds TA.** A note on the evaluation of novel biomarkers: do not rely on integrated discrimination improvement and net reclassification index. Stat Med. 2013. [PMID: 23553436]
111. **Brindle P, Emberson J, Lampe F, Walker M, Whincup P, Fahey T, et al.** Predictive accuracy of the Framingham coronary risk score in British men: prospective cohort study. BMJ. 2003;327:1267. [PMID: 14644971]
112. **Hense HW, Schulte H, Löwel H, Assmann G, Keil U.** Framingham risk function overestimates risk of coronary heart disease in men and women from Germany—results from the MONICA Augsburg and the PROCAM cohorts. Eur Heart J. 2003;24:937-45. [PMID: 12714025]
113. **Koller MT, Steyerberg EW, Wolbers M, Stijnen T, Bucher HC, Hunink MGM, et al.** Validity of the Framingham point scores in the elderly: results from the Rotterdam study. Am Heart J. 2007;154:87-93. [PMID: 17584559]
114. **Merry AH, Boer JM, Schouten LJ, Ambergen T, Steyerberg EW, Feskens EJM, et al.** Risk prediction of incident coronary heart disease in The Netherlands: re-estimation and improvement of the SCORE risk function. Eur J Prev Cardiol. 2012;19:840-8. [PMID: 21551214]
115. **Siontis GCM, Tzoulaki I, Siontis KC, Ioannidis JPA.** Comparisons of established risk prediction models for cardiovascular disease: systematic review. BMJ. 2012;344:e3318. [PMID: 22628003]
116. **Steyerberg EW.** Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. New York: Springer; 2009.
117. **Kramer AA, Zimmerman JE.** Assessing the calibration of mortality benchmarks in critical care: The Hosmer-Lemeshow test revisited. Crit Care Med. 2007;35:2052-6. [PMID: 17568333]
118. **Cox DR.** Two further applications of a model for binary regression. Biometrika. 1958;45:562-5.
119. **Cook NR, Paynter NP.** Comments on 'Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers' by M. J. Pencina, R. B. D'Agostino, Sr. and E. W. Steyerberg [Letter]. Stat Med. 2012; 31:93-5. [PMID: 21344474]
120. **Vergouwe Y, Moons KGM, Steyerberg EW.** External validity of risk models: Use of benchmark values to disentangle a case-mix effect from incorrect coefficients. Am J Epidemiol. 2010;172:971-80. [PMID: 20807737]
121. **Cook NR.** Comments on 'Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond' by M. J. Pencina et al., Statistics in Medicine (DOI: 10.1002/sim.2929). Stat Med. 2008; 27:191-5. [PMID: 17671959]
122. **Paynter NP, Cook NR.** A bias-corrected net reclassification improvement for clinical subgroups. Med Decis Making. 2013;33:154-62. [PMID: 23042826]
123. **Vickers AJ, Elkin EB, Steyerberg E.** Net reclassification improvement and decision theory [Letter]. Stat Med. 2009;28:525-6. [PMID: 17907248]
124. **Van Calster B, Vickers AJ, Pencina MJ, Baker SG, Timmerman D, Steyerberg EW.** Evaluation of markers and risk prediction models: overview of relationships between NRI and decision-analytic measures. Med Decis Making. 2013;33:490-501. [PMID: 23313931]
125. **Localio AR, Goodman S.** Beyond the usual prediction accuracy metrics: reporting results for clinical decision making [Editorial]. Ann Intern Med. 2012; 157:294-5. [PMID: 22910942]
126. **Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al.** Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology. 2010;21:128-38. [PMID: 20010215]
127. **Vickers AJ, Elkin EB.** Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making. 2006;26:565-74. [PMID: 17099194]
128. **Vickers AJ, Cronin AM, Begg CB.** One statistical test is sufficient for assessing new predictive markers. BMC Med Res Methodol. 2011;11:13. [PMID: 21276237]
129. **Pepe MS, Kerr KF, Longton G, Wang Z.** Testing for improvement in prediction model performance. Stat Med. 2013;32:1467-82. [PMID: 23296397]

Current Author Addresses: Drs. Leening and Witteman: Department of Epidemiology, Erasmus MC - University Medical Center Rotterdam, Dr. Molenwaterplein 50, 3015 GE Rotterdam, the Netherlands.
Ms. Vedder and Dr. Steyerberg: Department of Public Health, Erasmus MC - University Medical Center Rotterdam, Dr. Molenwaterplein 50, 3015 GE Rotterdam, the Netherlands.
Dr. Pencina: Department of Biostatistics and Bioinformatics, Duke Clinical Research Institute, Duke University, 2400 Pratt Street, Durham, NC 27715.

Author Contributions: Conception and design: M.J.G. Leening, E.W. Steyerberg.
Analysis and interpretation of the data: M.J.G. Leening, M.M. Vedder, E.W. Steyerberg.
Drafting of the article: M.J.G. Leening.
Critical revision of the article for important intellectual content: M.J.G. Leening, M.M. Vedder, J.C.M. Witteman, M.J. Pencina, E.W. Steyerberg.
Final approval of the article: M.J.G. Leening, M.M. Vedder, J.C.M. Witteman, M.J. Pencina, E.W. Steyerberg.
Statistical expertise: M.J.G. Leening, M.J. Pencina, E.W. Steyerberg.
Obtaining of funding: E.W. Steyerberg.
Administrative, technical, or logistic support: M.J.G. Leening.
Collection and assembly of data: M.J.G. Leening, M.M. Vedder.

130. **McGeechan K, Liew G, Macaskill P, Irwig L, Klein R, Sharrett AR, et al.** Risk prediction of coronary heart disease based on retinal vascular caliber (from the Atherosclerosis Risk In Communities [ARIC] Study). Am J Cardiol. 2008; 102:58-63. [PMID: 18572036]

Appendix Figure 1. **Summary of evidence search and selection.**



The search was last updated on 23 April 2013.

Appendix Table 1. List and Main Characteristics of the 67 Articles

Study, Year (Reference)	Article Type	Marker/Comparison	Outcome of Interest/Topic	Cutoffs Used for NRI
Adabag et al, 2008 (23)	Original article	–	Sudden death after MI	NA*
Auer et al, 2012 (24)	Original article	ECG abnormalities	CHD	7.5% and 15% at 7.5 y
Breteler, 2011 (25)	Editorial	–	Dementia	NA*
Buckley et al, 2009 (26)	Meta-analysis	CRP level	CHD	NA*
Chou et al, 2011 (27)	Review	Resting or exercise ECG	CVD	NA*
Cook and Ridker, 2009 (22)	Methods	CRP level	CVD	5%, 10%, and 20% at 10 y†
Cook, 2009 (28)	Letter	CRP level	CVD	NA*
Cornelis et al, 2009 (29)	Original article	Genetic risk score	Type 2 DM	NA*
de Boer et al, 2012 (30)	Original article	25-Hydroxyvitamin D level	Composite of hip fracture, MI, cancer, and death	50 nmol/L vs. season-specific at 10 y
de Lemos et al, 2010 (31)	Original article	Troponin T level	Cardiac structure and death	NA*
deFilippi et al, 2010 (32)	Original article	Troponin T level	Heart failure and CVD death	10% and 20% at 10 y
den Ruijter et al, 2012 (33)	Original article	cIMT	MI and stroke	5%, 10%, and 20% at 10 y
Devereaux et al, 2012 (34)	Original article	Troponin T level	Death after noncardiac surgery	1%, 5%, and 10% at 30 d
Di Angelantonio et al, 2012 (35)	Original article	Cholesterol, apolipoprotein, and Lp(a) levels	CVD	10% and 20% at 10 y
Eddy et al, 2011 (36)	Original article	Hypertension guidelines	MI and stroke	Not specified
Farooq et al, 2013 (37)	Original article	Coronary revascularization strategies	Death	NA*
Fonarow et al, 2012 (38)	Original article	NIH Stroke Scale	Stroke fatality	Not specified at 30 d
Gulati et al, 2013 (39)	Original article	Myocardial fibrosis	Death and major arrhythmia	5%, 10%, and 20% at 5 y (death); 15% at 5 y (major arrhythmia)
Helfand et al, 2009 (40)	Review	CAC score; leukocyte count; periodontal disease; ABI; cIMT; and CRP, Lp(a), homocysteine, and fasting glucose levels	CHD	NA*
Hingorani and Psaty, 2009 (41)	Commentary	–	CVD	NA*
Hlatky, 2012 (42)	Editorial	–	CHD	NA*
Janes et al, 2008 (43)	Methods	–	Risk stratification tables	NA*
Janssens et al, 2011 (44)	Methods	–	GRIPS statement	NA*
Kaptoge et al, 2010 (45)	Original article	CRP level	CHD, stroke, and death	NA*
Kaptoge et al, 2012 (46)	Original article	CRP and fibrinogen levels	CVD	10% and 20% at 10 y
Kathiresan et al, 2008 (47)	Original article	Genetic risk score	CVD	10% and 20% at 10 y
Kavousi et al, 2012 (48)	Original article	CKD; leukocyte count; CAC score; cIMT; PAD; PWV; and vWF antigen, NT-proBNP, fibrinogen, CRP, homocysteine, and uric acid levels	CHD	10% and 20% at 10 y†
Keller et al, 2011 (49)	Original article	Serial changes in troponin I level	MI	NA*
Kengne et al, 2012 (50)	Letter	CRP level and CAC score	CVD	NA*
Khera et al, 2011 (51)	Original article	Cholesterol efflux capacity	Obstructive CAD	NA*
Kim et al, 2008 (52)	Original article	Hyponatremia	Death in ESLD	NA*
Kivimäki et al, 2011 (53)	Original article	Working hours	CHD	5% and 10% at 10 y
Koller et al, 2012 (54)	Original article	BMI, CRP level, cIMT, ABI, and ECG-LVH	CHD	10% and 20% at 10 y
Lubitz et al, 2010 (55)	Original article	Familial atrial fibrillation	Atrial fibrillation	5% and 10% at 8 y
Lyssenko et al, 2008 (56)	Original article	Genetic polymorphisms	Type 2 DM	10% and 20% at an unspecified horizon
Manolio, 2010 (57)	Review	–	Genetic risk prediction	NA*
Martinez et al, 2012 (58)	Original article	U.K. and U.S. guidelines	Advanced colorectal dysplasia	Number, type, and size of adenomas at 1 y
Matsushita et al, 2012 (59)	Original article	CKD-EPI and MDRD equations	Death and ESRD	eGFR of 90, 60, 45, 30, and 15 mL/min per 1.73 m^2 at an unspecified horizon
McEvoy, 2010 (60)	Letter	CAC score	CHD	NA*
Meigs et al, 2008 (61)	Original article	Genetic risk score	Type 2 DM	2% and 8% at 8 to 10 y
Melander et al, 2009 (62)	Original article	CRP, cystatin C, Lp-PLA2, MR-proADM, MR-proANP, and NT-proBNP levels	CHD and CVD	6%, 10%, and 20% at 10 y
Melander et al, 2009 (63)	Letter reply	CRP level	CVD	NA*
Omland et al, 2009 (64)	Original article	Troponin T level	CVD death, heart failure, and MI	NA*
Palomaki et al, 2010 (65)	Meta-analysis	Chromosome 9p21 polymorphisms	CHD	5%, 10%, and 20% at 10 y†
Paynter et al, 2009 (66)‡	Original article	Chromosome 9p21.3 polymorphisms	CVD	5%, 10%, and 20% at 10 y†
Paynter et al, 2010 (67)	Original article	Genetic risk score	CVD	5%, 10%, and 20% at 10 y
Peralta et al, 2011 (68)	Original article	Creatinine level, cystatin C level, and urine albumin–creatinine ratio	Death and ESRD	Continuous NRI at an unspecified horizon

Continued on following page

Appendix Table 1—**Continued**

Study, Year (Reference)	Article Type	Marker/Comparison	Outcome of Interest/Topic	Cutoffs Used for NRI
Pischon et al, 2008 (69)	Original article	BMI and abdominal adiposity	Death	2.5%, 5%, and 7.5% at 5 y
Pletcher et al, 2010 (70)	Letter	–	CVD	NA*
Polak et al, 2011 (71)	Original article	cIMT	CVD	6% and 20% at 10 y†
Polonsky et al, 2010 (72)	Original article	CAC score	CHD	3% and 10% at 5 y
Ripatti et al, 2010 (73)	Original article	Genetic risk score	CHD	5%, 10%, and 20% at 10 y
Rosenberg et al, 2010 (74)	Original article	Gene expression test	Presence of obstructive CAD	20% and 50%
Schelbert et al, 2012 (75)	Original article	Unrecognized MI	Death	Continuous NRI at an unspecified horizon
Schnabel et al, 2009 (76)	Original article	Echocardiographic measurements	Atrial fibrillation	5% and 15% at 10 y
Selvin et al, 2010 (77)	Original article	Glycated hemoglobin level	Type 2 DM, CHD, and death	5%, 10%, and 20% at 10 y
Steyerberg and Pencina, 2010 (78)	Letter	CRP level	CVD	5%, 10%, and 20% at 10 y†
Tammemägi et al, 2013 (79)	Original article	Smoking intensity and history of cancer	Lung cancer	1% and 2% at 6 y
Tangri et al, 2011 (80)	Original article	Calcium, phosphate, bicarbonate, and albumin levels	CKD	Not specified
Tzoulaki et al, 2009 (81)	Review	86 predictors	CHD	NA*
Wacholder et al, 2010 (82)	Original article	Genetic polymorphisms	Breast cancer	NA*
Wilson, 2009 (83)	Editorial	–	CHD	NA*
Wormser et al, 2011 (84)	Original article	BMI and abdominal adiposity	CHD and stroke	5%, 10%, and 20% at 10 y
Wormser et al, 2011 (85)	Letter reply	BMI and abdominal adiposity	CHD	NA*
Yeboah et al, 2012 (86)	Original article	cIMT, CAC score, brachial FMD, ABI, CRP level, and family history	CHD and CVD	5% and 20% at 10 y†
Zethelius et al, 2008 (87)	Original article	Troponin I, NT-proBNP, cystatin C, and CRP levels	CVD death	6% and 20% at an unspecified horizon
Zoungas et al, 2010 (88)	Original article	Severe hypoglycemia	CVD	NA*

ABI = ankle–brachial index; BMI = body mass index; CAC = coronary artery calcium; CAD = coronary artery disease; CHD = coronary heart disease; cIMT = carotid intima–media thickness; CKD = chronic kidney disease; CKD-EPI = Chronic Kidney Disease Epidemiology Collaboration; CRP = C-reactive protein; CVD = cardiovascular disease; DM = diabetes mellitus; ECG = electrocardiography; ECG-LVH = electrocardiographic left ventricular hypertrophy; eGFR = estimated glomerular filtration rate; ESLD = end-stage liver disease; ESRD = end-stage renal disease; FMD = flow-mediated dilation; GRIPS = Genetic Risk Prediction Studies; Lp(a) = lipoprotein(a); Lp-PLA2 = lipoprotein-associated phospholipase A2; MDRD = Modification of Diet in Renal Disease; MI = myocardial infarction; MR-proADM = midregional proadrenomedullin; MR-proANP = midregional proatrial natriuretic peptide; NA = not applicable; NIH = National Institutes of Health; NRI = net reclassification improvement; NT-proBNP = N-terminal fragment of prohormone B-type natriuretic peptide; PAD = peripheral arterial disease; PWV = pulse wave velocity; vWF = von Willebrand factor.

* NRI was not calculated.

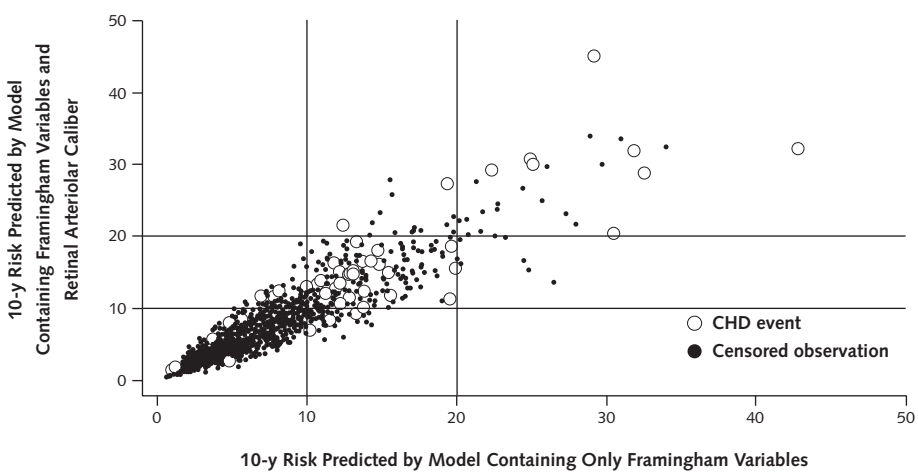
† Observations from a follow-up period shorter than the predicted time horizon were used.

‡ Identified through hand-search with erroneous citation linkage to a methodological article on NRI.

Appendix Table 2. **Summary Characteristics of the 67 Articles**

Characteristic	Studies, *n* (%)
Journal	
New England Journal of Medicine	15 (22)
The Lancet	7 (10)
Journal of the American Medical Association	28 (42)
Annals of Internal Medicine	17 (25)
Year of print publication	
2008	8 (12)
2009	13 (19)
2010	16 (24)
2011	12 (18)
2012	15 (22)
2013	3 (4)
Cited methodological article	
Pencina et al, 2008 (15)	56 (84)
Pencina et al, 2010 (19)	3 (4)
Pencina et al, 2011 (20)	9 (13)
Pencina et al, 2012 (21)	0 (0)
Cook and Ridker, 2009 (22)	11 (16)
Country of address for correspondence	
Australia	1 (1)
Canada	2 (3)
Finland	1 (1)
Germany	2 (3)
Greece	1 (1)
The Netherlands	6 (9)
Norway	1 (1)
South Africa	1 (1)
Sweden	4 (6)
Switzerland	1 (1)
United States	39 (58)
United Kingdom	8 (12)

Appendix Figure 2. **Example of a reclassification graph with superimposed cut points of predicted risk.**



The graph shows 10-y risk for incident CHD in women from the ARIC (Atherosclerosis Risk in Communities) Study predicted by a model containing only the Framingham risk score variables (horizontal axis) against risk predicted by a model containing Framingham risk score variables and retinal arteriolar caliber (vertical axis). Lines at predicted risks of 10% and 20% are superimposed to show reclassification over clinically relevant cut points (2, 89) and thereby create a visual representation of a reclassification table (**Appendix Table 3**). Of note, most women in this study have a low (<10%) predicted risk for CHD, both with the Framingham variables and with the model that includes retinal arteriolar caliber. The graph also shows that a limited number of women are reclassified over the cut points (i.e., only a small proportion of dots lies in the off-diagonal cells of the graph). CHD = coronary heart disease. (Reproduced from McGeechan and colleagues [130] with permission of the *American Journal of Cardiology*.)

Appendix Table 3. **Example of a Risk Reclassification Table Stratified by Event Status***

Model Containing Only Framingham Risk Score Variables	Model Containing Framingham Risk Score Variables and Coronary Artery Calcium Score			
	<10% Risk	10%–20% Risk	>20% Risk	Total
<10% risk		50	4	125
Persons with event, *n*	71			
Persons without event, *n*	2015	315	16	2346
Total persons, *n*	2086	365	20	2471
Observed risk (95% CI), %	3 (2–5)	14 (10–19)	21 (6–60)	–
10%–20% risk				
Persons with event, *n*	19	75	55	149
Persons without event, *n*	262	364	144	770
Total persons, *n*	281	439	199	919
Observed risk (95% CI), %	7 (4–12)	17 (13–22)	28 (20–37)	–
>20% risk				
Persons with event, *n*	0	9	62	71
Persons without event, *n*	17	60	140	217
Total persons, *n*	17	69	202	288
Observed risk (95% CI), %	0 (0–0)	13 (6–27)	31 (23–40)	–
Total persons, *n*				
With event	90	134	121	345
Without event	2294	739	300	3333
Total	2384	873	421	3678

* The Table shows reclassification for 10-y risk for incident coronary heart disease in participants from the Rotterdam Study predicted by a model containing only the Framingham risk score variables against risk predicted by a model containing Framingham risk score variables and coronary artery calcium score. The numbers are rounded due to the use of Kaplan–Meier estimates for persons with incomplete follow-up. (Reproduced from reference 48.)