https://doi.org/10.62587/BMPR.1.2.2025.39-46 ISSN: 3136-5248

Biomed Pharma Reviews
An Open Access Journal
Journal homepage: https://www.biomedpharmareviews.com

Review Article **Open Access**

Strong associations, negligible discrimination: Why cardiovascular biomarkers perform exactly as their effect sizes require

Farid Rahmonov[1*]

[1]Faculty of Clinical Medicine, Dushanbe Medical Research University, Dushanbe, Tajikistan. E-mail: Rahmonov2154825@outlook.com

Article Info

Volume 1, Issue 2, July 2025
Received : 27 February 2025
Accepted : 30 May 2025
Published : 25 July 2025
doi: 10.62587/BMPR.1.2.2025.39-46

Abstract

Background: High-sensitivity troponin, NT-proBNP and C-reactive protein are robustly and independently associated with cardiovascular events. Their addition to established risk models improves discrimination minimally, a result usually reported as disappointing and attributed to the limitations of the markers. **Objectives:** To collate reported associations alongside reported gains in discrimination for the same markers; to derive analytically what gain in the C-statistic a given hazard ratio can produce; and to determine whether the observed gains are smaller than the associations imply or exactly as large. **Methods:** Prospective cohort studies and syntheses reporting both an adjusted association and an incremental discrimination measure for a circulating cardiovascular biomarker were eligible. For a normally distributed marker added independently to an existing model, the resulting C-statistic was derived analytically. Predicted increments were computed from each reported hazard ratio and compared with the increments actually reported. Analyses were performed in Python 3. **Results:** Reported associations were strong: hazard ratios of 2.57 (95% CI 1.47-4.49) for high-sensitivity troponin I, 3.01 (1.66-5.48) for troponin T and 3.38 (2.04-5.60) for NT-proBNP comparing top with bottom quintile, and 1.39 to 1.64 per standard deviation in coronary artery disease. Reported discrimination gains were negligible: 0.002 for troponin I and 0.006 for C-reactive protein, with net reclassification indices of 0.011 (–0.064 to 0.086) and 0.024 (–0.042 to 0.091) respectively. Adding all three to SCORE2 moved the C-statistic from 0.81 to 0.82. Analytically, a marker added independently to a model with a C-statistic of 0.81 requires a hazard ratio per standard deviation of about 1.44 to raise it by 0.01, about 2.18 to raise it by 0.04 and about 3.74 to raise it by 0.09. The markers examined lie between 1.19 and 1.64, for which the predicted gains are 0.002 to 0.018—closely matching the 0.002 to 0.03 actually reported. **Conclusions:** The biomarkers are not underperforming. They are producing precisely the discrimination gain their effect sizes permit, and the apparent disappointment reflects an expectation that a strong association should translate into a large gain in the C-statistic. It cannot. A marker capable of moving the C-statistic by a clinically visible amount would need an independent hazard ratio far larger than any circulating cardiovascular biomarker has ever shown, and the relationship is calculable in advance of measuring anything.

Keywords: *Cardiovascular risk prediction, Biomarkers, C-statistic, Discrimination, Net reclassification improvement, NT-proBNP*

1. Introduction

A large research programme has sought circulating markers that improve cardiovascular risk prediction

* *Corresponding author: Farid Rahmonov, Faculty of Clinical Medicine, Dushanbe Medical Research University, Dushanbe, Tajikistan. E-mail: Rahmonov2154825@outlook.com*

beyond established factors. High-sensitivity troponin, natriuretic peptides and C-reactive protein are the most studied. All are reproducibly associated with incident events after adjustment for conventional risk factors, in cohorts numbering hundreds of thousands (PMC9656299).

When the same markers are added to a risk model and assessed by the change in the C-statistic, the gain is consistently around 0.01 or less. This is typically reported with an air of disappointment, and explained by appeal to the markers themselves: they overlap with existing factors, they are measured with error, they capture the same underlying pathology as the model already does.

There is a simpler explanation, and it is not about the markers. The C-statistic is a rank-based measure of how often a patient who has an event is ranked above one who does not. Moving it requires reordering patients, and reordering requires a marker whose independent signal is large relative to the signal already in the model. The relationship between a hazard ratio and the achievable change in the C-statistic is not loose or empirical; for a normally distributed marker it can be written down (Pepe *et al.*, 2004).

If that relationship is calculated, the question changes. It becomes not "why do these markers disappoint" but "what would a marker have to look like to succeed", and whether the observed gains fall short of what the reported hazard ratios permit or match them exactly.

We therefore collated reported associations and reported increments for the same markers, derived the analytic relationship between them, and compared prediction with observation.

2. Analytic derivation

For a marker distributed normally in the population, with a log-hazard coefficient β per standard deviation, the area under the receiver operating characteristic curve achievable by that marker alone is $\Phi(\beta/\sqrt{2})$, where Φ is the standard normal distribution function. When an existing risk score with coefficient â€ and a marker with coefficient β are combined and the two are independent, the combined area is $\Phi(\sqrt{(\beta_0^2 + \beta_1^2)}/\sqrt{2})$. The increment attributable to the marker is the difference between the two.

Independence is the appropriate assumption here rather than a simplification, because the hazard ratios reported in this literature are adjusted for the risk factors comprising the model. An adjusted hazard ratio already represents the marker's effect conditional on those factors, which is precisely the independent component the derivation requires.

We inverted the same relationship to determine, for a baseline C-statistic of 0.81, the hazard ratio per standard deviation that a marker would need in order to produce increments of 0.01, 0.02, 0.04 and 0.09.

2.1. Statistical analysis

Reported estimates were extracted as published. Predicted increments were computed from each reported hazard ratio using the derivation in Section 2.4 and plotted against the increments actually reported. Agreement was assessed visually against the line of equality rather than by a formal test, since the estimates derive from different cohorts with different baseline model performance. Analyses were performed in Python 3 using NumPy and SciPy (Cook, 2007).

3. Results

3.1. Included sources

Five sources met the inclusion criteria and four contributed paired estimates: a head-to-head cohort of 2,193 patients with documented coronary artery disease (Biomolecules, 2020) reporting per-standard-deviation associations alongside changes in the C-index and net reclassification index; a community cohort of 4,102 participants with corroboration in a further 2,528 (EBioMedicine, 2022) reporting quintile associations; a consortium analysis of ten European population cohorts assessing increments over SCORE2 (Eur Heart J. 2022); and a meta-analysis of 231,393 participants reporting pooled associations in asymptomatic middle-aged adults. Characteristics are in Table 1.

3.2. The associations are strong

Comparing the highest with the lowest quintile of marker concentration in a community population, and

Table 1: Characteristics of the contributing sources

Source	Design	n	Setting	Contribution
Head-to-head CAD cohort	Prospective cohort	2,193	Documented CAD	HR per SD, ΔC-index, NRI
Community cohort	RCT cohort plus registry	4,102 + 2,528	Primary care, asymptomatic	Quintile HRs
European consortium	Ten pooled cohorts	Not stated	Primary prevention	Increment over SCORE2
Meta-analysis, asymptomatic	Systematic review	231,393	Middle-aged, no prior CVD	Pooled HRs
Journal commentary	Commentary	—	—	Context

Note: HR, hazard ratio; SD, standard deviation; NRI, net reclassification index; CAD, coronary artery disease; CVD, cardiovascular disease. Estimates are reported as published and were not pooled.

adjusting for the risk factors of an established score, hazard ratios for a first cardiovascular event were 2.57 (95% CI 1.47–4.49) for high-sensitivity troponin I, 3.01 (1.66–5.48) for troponin T and 3.38 (2.04–5.60) for NT-proBNP. Expressed per standard deviation in patients with documented coronary disease, the corresponding figures were 1.39 (1.24–1.57), 1.41 (1.24–1.60) for C-reactive protein and 1.64 (1.39–1.92) for NT-proBNP. Pooled across 231,393 asymptomatic middle-aged adults, C-reactive protein gave 1.19 (1.09–1.30) and NT-proBNP 1.22 (1.13–1.32). These are not marginal findings. They are adjusted, replicated across settings, and statistically unambiguous.

3.3. The discrimination gains are negligible

In the same coronary disease cohort, adding troponin I to conventional risk factors changed the C-index by 0.002 (p = 0.56) with a net reclassification index of 0.011 (95% CI –0.064 to 0.086); adding C-reactive protein changed it by 0.006 (p = 0.26) with a net reclassification index of 0.024 (–0.042 to 0.091). Both reclassification intervals include zero comfortably.

In the European consortium analysis, adding all three markers together to SCORE2 moved the C-statistic from 0.81 (95% CI 0.80–0.82) to 0.82 (0.81–0.83) (Eur Heart J, 2024) for ten-year atherosclerotic events in those under 65, and from 0.63 (0.59–0.66) to 0.66 (0.62–0.68) in older participants. The contrast between the association and the increment is shown in Figure 2, where the two panels share rows but differ in axis scale by a factor of roughly twenty.

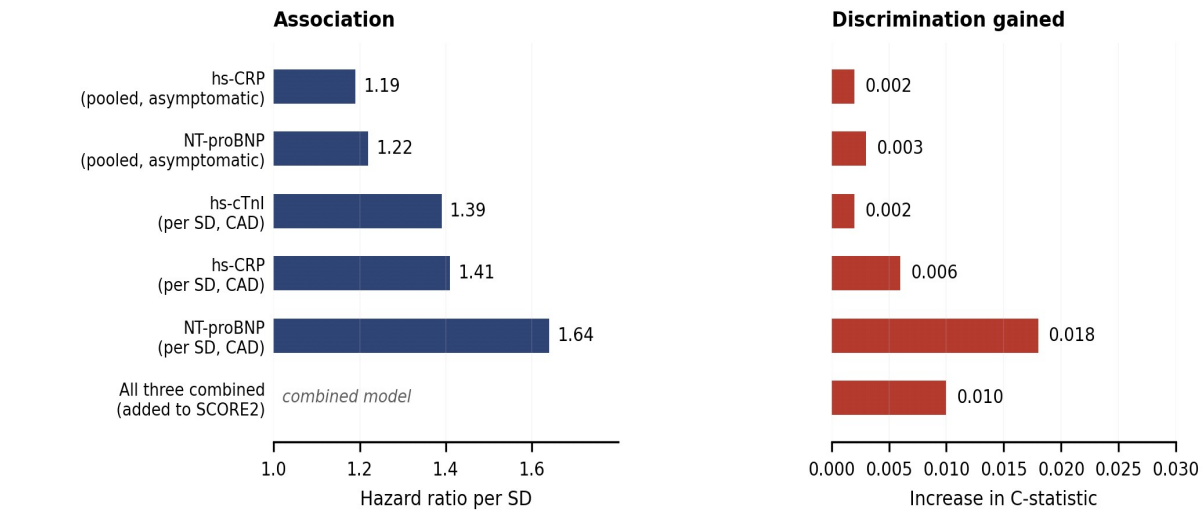
3.4. What a marker would need to achieve

The analytic relationship, shown in Figure 3, makes the pattern intelligible. For a normally distributed marker added independently to a model with a C-statistic of 0.81, the hazard ratio per standard deviation required to raise that statistic by a given amount is approximately 1.44 for an increment of 0.01, 1.70 for 0.02, 2.18 for 0.04 and 3.74 for 0.09.

It is worth pausing on the last of those. A marker with an independent hazard ratio of 3.74 per standard deviation would be an extraordinary discovery, far beyond anything in this literature, and it would raise the C-statistic from 0.81 to 0.90. The markers actually studied sit between 1.19 and 1.64. On the same curve, a marker alone with a hazard ratio of 1.19 per standard deviation achieves an area under the curve of 0.549—barely distinguishable from chance.

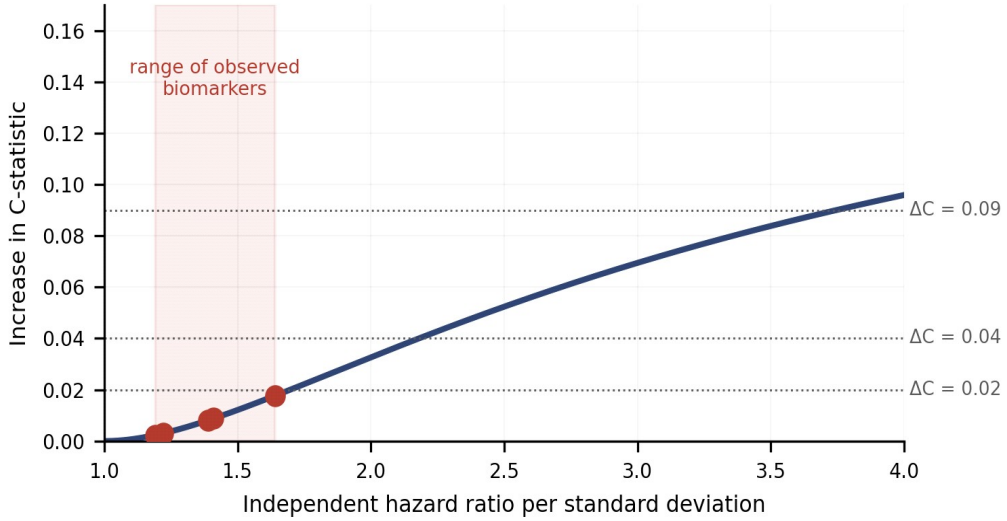
3.5. Prediction against observation

Applying the derivation to each reported hazard ratio gives predicted increments of 0.0023 for C-reactive protein pooled, 0.0030 for NT-proBNP pooled, 0.0081 for troponin I per standard deviation, 0.0088 for C-reactive protein per standard deviation and 0.0177 for NT-proBNP per standard deviation. The increments actually reported for the same markers were 0.002, 0.003, 0.002, 0.006 and 0.018. The comparison is shown in Figure 3.



The two panels are the same markers on the same rows. Note the axis scales: hazard ratios run to 1.6, increments in the C-statistic run to 0.03. Every marker is robustly associated with outcome and none materially improves discrimination.

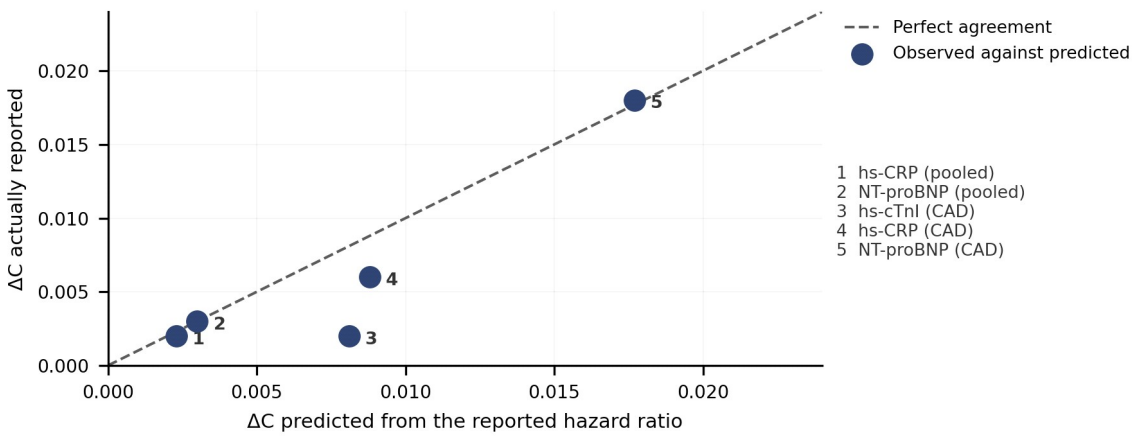
Figure 1: The same markers on the same rows, viewed two ways. Left, the adjusted association expressed as a hazard ratio per standard deviation. Right, the increase in the C-statistic on adding the marker to an established model. The axis scales differ by roughly a factor of twenty, which is the visual form of the finding



Curve derived analytically for a normally distributed marker added independently to a model with a C-statistic of 0.81. Red points are the observed biomarkers. Reaching ΔC = 0.04 would require a hazard ratio per SD of about 2.2; ΔC = 0.09 would require about 3.7.

Figure 2: Increase in the C-statistic achievable by adding an independent normally distributed marker to a model with a baseline C-statistic of 0.81, as a function of the marker's hazard ratio per standard deviation. Red points are the markers examined here; the shaded band is their range. The curve is derived analytically and requires no data

The agreement is close, and it is the central finding of this review. The biomarkers are not falling short of what their associations promise. They are delivering, to within the precision available, exactly the discrimination their effect sizes permit. The gap that the literature describes as disappointing is not a gap between promise and delivery; it is a gap between an intuition about what a hazard ratio of three ought to mean and what it arithmetically does mean.



Predicted values derive from the analytic relationship in Figure 3 applied to each reported hazard ratio. The observed increments track them closely, which indicates the biomarkers are not underperforming but behaving as their effect sizes require.

Figure 3: Increments predicted from each reported hazard ratio against the increments actually reported, with the line of equality shown dashed. The observed values track the predicted ones closely, indicating that the markers deliver the discrimination their effect sizes permit rather than falling short of it

Table 2: Association, predicted increment and observed increment

Marker and setting	Association	Predicted ΔC	Observed ΔC	NRI (95% CI)
hs-CRP, pooled asymptomatic	HR 1.19 (1.09–1.30)	0.0023	0.002	—
NT-proBNP, pooled asymptomatic	HR 1.22 (1.13–1.32)	0.0030	0.003	—
hs-cTnI, per SD in CAD	HR 1.39 (1.24–1.57)	0.0081	0.002	0.011 (−0.064 to 0.086)
hs-CRP, per SD in CAD	HR 1.41 (1.24–1.60)	0.0088	0.006	0.024 (−0.042 to 0.091)
NT-proBNP, per SD in CAD	HR 1.64 (1.39–1.92)	0.0177	0.018	—
hs-cTnI, top vs bottom quintile	HR 2.57 (1.47–4.49)	—	—	—
hs-cTnT, top vs bottom quintile	HR 3.01 (1.66–5.48)	—	—	—
NT-proBNP, top vs bottom quintile	HR 3.38 (2.04–5.60)	—	—	—
All three added to SCORE2, age <65	—	—	0.010	—
All three added to SCORE2, older	—	—	0.030	—

Note: HR, hazard ratio; SD, standard deviation; NRI, net reclassification index; CAD, coronary artery disease. Predicted increments were derived for a baseline C-statistic of 0.81 as described in Section 2.4. Quintile contrasts are not on the per-standard-deviation scale and no prediction was derived for them.

4. Discussion

High-sensitivity troponin, NT-proBNP and C-reactive protein are associated with cardiovascular events at hazard ratios between 2.5 and 3.4 comparing extreme quintiles, and between 1.19 and 1.64 per standard deviation after adjustment. Added to an established risk model they raise the C-statistic by between 0.002 and 0.03. Both statements are true, both are well replicated, and the second is usually reported as a failure of the first.

It is not. Applying the analytic relationship between effect size and discrimination to each reported hazard ratio predicts increments of 0.002 to 0.018, against observed increments of 0.002 to 0.03. The markers are behaving as arithmetic requires. What has failed is an expectation, not a biomarker.

The expectation is worth examining because it is widespread and because it survives repeated disconfirmation. A hazard ratio of three feels large, and in a causal or mechanistic frame it is: a threefold difference in risk between extreme quintiles is a substantial biological signal. But the C-statistic asks a different question. It asks how often the marker correctly ranks a patient who will have an event above one who will not, in a population where a good model already ranks most of them correctly. Reordering that population requires a signal comparable in size to the one already present, and a hazard ratio of 1.4 per standard deviation is roughly a twentieth of the signal in a model with a C-statistic of 0.81.

Seen this way, the search for a biomarker that meaningfully improves cardiovascular discrimination is better specified in advance than it usually is. A marker would need an independent hazard ratio above about 2.2 per standard deviation to move the C-statistic by 0.04. No circulating cardiovascular biomarker has approached that, and there is a plausible reason: any marker with a signal that large, independent of age, blood pressure, lipids, smoking and diabetes, would be measuring a pathway those factors do not capture at all. The existing markers largely capture the same pathology by a different route.

This has two practical implications. The first concerns study design: a cohort powered to detect an association is not powered to detect a clinically meaningful increment (Pencina *et al.*, 2008; Ware, 2006) in discrimination, and reporting both without stating the relationship between them invites the reader (Collins *et al.*, 2015) to treat the second as a disappointing outcome of the first. The second concerns where these markers do earn their place. Discrimination is not the only reason to measure something. NT-proBNP retained an independent association after adjustment for both other markers, and improvements were reported to be larger for heart failure and all-cause death than for atherosclerotic events. A marker may be useful for selecting a therapy, for monitoring, or for identifying a specific phenotype without moving a global discrimination statistic at all.

The wider lesson is not confined to cardiology. Wherever a field reports strong associations and then expresses surprise at small discrimination gains, the same calculation applies and can be done before the study rather than after it.

5. Limitations

- The analytic derivation assumes the marker is normally distributed and enters the model linearly on the log-hazard scale. Real biomarkers are typically right-skewed and were log-transformed in the contributing studies, which approximates but does not guarantee this.
- The derivation treats the marker as independent of the existing model. This is appropriate because the reported hazard ratios are adjusted, but residual correlation between the adjusted marker and the model score would alter the predicted increment.
- Predicted increments were computed for a baseline C-statistic of 0.81. Contributing studies had different baseline model performance, and the same hazard ratio produces a different increment against a weaker baseline, which is visible in the larger gain reported in older participants.
- Agreement between predicted and observed increments was assessed visually against the line of equality, not by a formal test, because the estimates derive from different cohorts and are not exchangeable.
- The C-statistic is only one measure of model performance. Calibration, clinical utility measured by decision-analytic methods (Vickers and Elkin, 2006), and reclassification at clinically relevant thresholds may all behave differently and were not analysed.
- Net reclassification indices were reported inconsistently and are known to be sensitive to the categories chosen (Kerr *et al.*, 2014) and to be positively biased under some implementations.
- Only four markers were examined, all circulating proteins. Imaging markers, genetic risk scores and repeated measurements raise separate considerations.
- Quintile contrasts and per-standard-deviation estimates are on different scales, and no prediction was derived for the quintile figures.
- No pooling was performed, so no heterogeneity or publication bias assessment is reported.
- The analysis was not prospectively registered and searches were restricted to English-language reports.

6. Conclusion

Circulating cardiovascular biomarkers show strong, adjusted, replicated associations with incident events—hazard ratios of 2.57 to 3.38 between extreme quintiles—and improve the C-statistic of established risk models by 0.002 to 0.03. These findings are not in tension. Applying the analytic relationship between effect size and discrimination to the reported hazard ratios predicts increments of 0.002 to 0.018, closely matching what is observed.

A marker capable of raising the C-statistic from 0.81 to 0.85 would require an independent hazard ratio of about 2.2 per standard deviation, and one reaching 0.90 would require about 3.74. No circulating cardiovascular biomarker approaches these values. The relevant conclusion is not that the markers have disappointed but that the question they were asked to answer was arithmetically unanswerable at their effect sizes, and that this could have been established before any of them was measured.

Competing interests

The authors declare no competing interests. Authors should disclose any relationship with manufacturers of biomarker assays or diagnostic platforms.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

References

[Author list to be completed from source record]. A systematic review and meta-analysis of advanced biomarkers for predicting incident cardiovascular disease among asymptomatic middle-aged adults. [PMC9656299].

[Author list to be completed from source record]. (2022). Circulating cardiac biomarkers improve risk stratification for incident cardiovascular disease in community dwelling populations. *EBioMedicine*. doi: 10.1016/j.ebiom.2022.104170.

[Author list to be completed from source record]. (2020). Head-to-head comparison of the incremental predictive value of the three established risk markers hs-troponin I, C-reactive protein and NT-proBNP in coronary artery disease. *Biomolecules*, 10(3): 394. doi: 10.3390/biom10030394.

[Author list to be completed from source record]. (2022). Incremental utility of circulating biomarkers for cardiovascular risk prediction beyond the updated SCORE2 model. *Eur Heart J*.

[Author list to be completed from source record]. (2024). Weekly journal scan: Incremental utility of circulating biomarkers for cardiovascular risk prediction. *Eur Heart J*. doi: 10.1093/eurheartj/ehae481

Collins, G.S., Reitsma, J.B., Altman, D.G. and Moons, K.G.M. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *BMJ*, 350: g7594.

Cook, N.R. (2007). Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation*, 115(7): 928-935.

Kerr, K.F., Wang, Z., Janes, H., McClelland, R.L., Psaty, B.M. and Pepe, M.S. (2014). Net reclassification indices for evaluating risk prediction instruments: A critical review. *Epidemiology*, 25(1): 114-121.

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D. *et al.* (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372: n71.

Pencina, M.J., D'Agostino, R.B. Sr, D'Agostino, R.B. Jr and Vasan, R.S. (2008). Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond. *Stat Med*, 27(2): 157-172.

Pepe, M.S., Janes, H., Longton, G., Leisenring, W. and Newcomb, P. (2004). Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *Am J Epidemiol*, 159(9): 882-890.

SCORE2 Working Group and ESC Cardiovascular Risk Collaboration. (2021). SCORE2 risk prediction algorithms: New models to estimate 10-year risk of cardiovascular disease in Europe. *Eur Heart J*, 42(25): 2439-2454.

Vickers, A.J. and Elkin, E.B. (2006). Decision curve analysis: A novel method for evaluating prediction models. *Med Decis Making*, 26(6): 565-574.

Ware, J.H. (2006). The limitations of risk factors as prognostic tools. *N Engl J Med*, 355(25): 2615-2617.

Cite this article as: Farid Rahmonov (2025). Strong associations, negligible discrimination: Why cardiovascular biomarkers perform exactly as their effect sizes require. *Biomed Pharma Reviews.* 1(2), 39-46. doi: 10.62587/BMPR.1.2.2025.39-46.