Reviewer #2: The authors have been largely responsive to my comments but there are still a couple of issues that need to be clarified.

1. You seem overly positive on categorical NRI and overly negative on NRI at event rate.
   a. I would prefer that it is stressed more forcefully that categorical NRI should be presented separately for events and non-events.

   I have omitted the sentence that stated this on page 7 and emphasized the point by replacing it with, “Note that the components of the NRI should always be reported separately for those with and without events, so that these can be appropriately interpreted and reweighted if desired.”

   b. While it is true that the clinical interpretation of NRI(p) depends on the adequacy of the threshold, its interpretation as a measure of distance between distributions of risk between events and non-events does not. In that sense reclassification from the null (AARD, Max Youden) is a global discrimination measure that corresponds to the AUC and it is proper. This should be noted.

   On page 7 I added the sentence, “It is a proper measure of global discrimination measure that serves as a measure of distance between the distributions of risk between events and non-events.”

   c. The fact that 28% have CVD risk above 3% does not necessarily mean that 3% is the wrong threshold. The AHA/ACC guidelines offer an optional threshold of 5% and some argue that statins could benefit individuals at an even lower risk.

   On page 8 I rephrased this to, “This threshold is lower than the 7.5% used in current recommendations for statin therapy. Ideally, cost-benefit considerations should be used to obtain the optimal thresholds to translate these measures to clinical importance.”

2. You should state more clearly that discrimination slope and rescaled Brier are only asymptotically equivalent (and not equivalent). I still think it would be useful to give the authors a choice whether to base IDI on discrimination slope versus rescaled Brier noting the risks - slope being only...
asymptotically proper and rescaled Brier taking negative values.

I have added that these two measures are asymptotically equivalent on page 9 bottom.

3. Please add a citation by Hajime Uno to the section on AUC in survival - it provides the most elegant way to handle censoring.

I added a reference to the work of Uno on page 5, “Uno et al14 also introduced a modified c-statistic that does not depend on the study-specific censoring distribution.”

4. Split sample validation based on one split is not appropriate. If you want to recommend this approach, ask for average of multiple splits.

This may not always be feasible, but I changed it to, “ideally taking an average over multiple splits,” on page 12 and to “(ideally multiple) split samples” in the box.

Associate Editor: please add that net benefit is proper.

I added a sentence on page 11 describing the net benefit, “The net benefit, which is a proper measure, compares the benefits and risks of decisions, weighting by their relative harms or tradeoff.”