Appendix: mathematical background and calculations for CCTA

Ewout Steyerberg, Ben van Calster, Andrew Vickers; March 21, 2016

We give some mathematical background on the calculation of alternative summary measures for incremental value of a test. All are based on binary classification of patients as high versus low risk. For the clinical application of CCTA, we assume a threshold of 5%, but the full range of 1% to 15% might be considered [1].

Net absolute effect is referred to as Net Absolute Reclassification Improvement (NARI). Net Benefit is abbreviated as NB. Another summary measure is the Net Reclassification Improvement (NRI), which can be used for binary classification, but also for more than 2 categories [3, 6]. All these three measures (NARI, NB, NRI) are weighted sums of the change in sensitivity (delta(sensitivity)) and change in specificity (delta(specificity)), where sensitivity is the true positive (TP) rate (TP/Nevent) and specificity is the true negative (TN) rate (TN/Nnonevent). The total number of patients is indicated by N.

**NRI calculation**

\[
NRI = \text{delta(sensitivity)} + \text{delta(specificity)},
\]

or

\[
NRI = \text{delta}(TP/Nevent) + \text{delta}(TN/Nnonevent),
\]

or

\[
NRI = \text{delta}(TP/Nevent) - \text{delta}(FP/Nnonevent).
\]

NRI is the simple sum of two conditional probabilities: the changes in sensitivity, or true positive (TP) rate and specificity, or true negative (TN) rate. With more false positives (FP), as in the case study of CCTA, NRI would simply subtract the FP fraction from the TP fraction. This sum implicitly weighs by the event rate (Nevent/N). If the event rate is 8% as in this case, we weight specificity changes, i.e. more TN (or less FP), as 8:92 as important compared to TP changes.

**Net Benefit calculation**

For Net Benefit, we define the difference between 2 models as delta(NB). It is common to define NB in terms of TP and a weighted number of FP classifications:

\[
\text{delta(NB)} = (\text{delta}(TP) - w^{*}\text{delta}(FP)) / N.
\]

But NB can also be defined in changes in sensitivity and specificity:

\[
\text{delta(NB)} = \text{Nevent/N}^{*}\text{delta(sensitivity)} - \text{Nnonevent/N}^{*}w^{*}\text{delta(specificity)}.
\]

The crucial aspect of NB is the weight w to relate FP to TP classifications [5].

If \(w = \text{Nevent} / \text{Nnonevent}\) (the odds of the event rate), then

\[
\text{delta(NB)} = \text{NRI}^{*}\text{Nevent/N},
\]

or

\[
\text{delta(NB)} = \text{Nevent/N}^{*}\text{delta(sensitivity)} - \text{Nnonevent/N}^{*}\text{Nevent/N}^{*}\text{delta(specificity)} = \text{Nevent/N}^{*}\text{delta(sensitivity)} - \text{Nevent/N}^{*}\text{Nevent}^{*}\text{delta(specificity)}.
\]
Using the event rate of 8% for the weight of FP vs TP classifications is consistent with using 8% as a decision threshold; at that point NRI and NB are equivalent except by a constant (the event rate). If we use a 50% decision threshold to define high risk patients, $w = 1$. With that weight, NB = NARI. This weighting would imply that FP classifications are as important as TP classifications.

**Net Absolute Reclassification Improvement (NARI) calculation**

NARI = $\text{Nevent/N} \times \delta(\text{sensitivity}) + \text{Nnonevent/N} \times \delta(\text{specificity})$, or $\text{NARI} = (\delta(\text{TP}) - \delta(\text{FP})) / \text{N}$.

So, NARI is a weighted sum of NRI, which would just add $\delta(\text{sensitivity})$ and $\delta(\text{specificity})$ without weighting for event rate and non-event rate as in the NARI. NARI is equal to NB if we use $w=1$. The second formulation clarifies that NARI is simply the change in accuracy of a classification rule, where accuracy = (TP+TN)/N. Accuracy is a notoriously poor summary measure for performance evaluation. In the case study, the 8% event rate would imply an accuracy of at least 92% if all patients would be classified as negative.

**Numerical illustration**

We illustrate the calculation of NRI, NB, NARI and accuracy for a 5% decision threshold, which is plausible based on the European Society of Cardiology guidelines (see Table below, based on original Table 3) [1].

<table>
<thead>
<tr>
<th>RCRI scores only</th>
<th>Model that included CCTA findings</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Event</td>
<td>No Event</td>
<td></td>
</tr>
<tr>
<td>&lt;5%</td>
<td>&lt;5%</td>
<td>&gt;=5%</td>
<td>&gt;=5%</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>191</td>
<td>114</td>
</tr>
<tr>
<td>&gt;=5%</td>
<td>0</td>
<td>59</td>
<td>47</td>
</tr>
</tbody>
</table>

There are net 10 TP classifications among 74 patients with an event, and 114 – 47 = 67 more FP classifications among 881 patients without an event, for a total N of 955 patients. The increase in sensitivity was $10/74 = 13.5%$; the decrease in specificity was $67/881 = 7.6%$.

NRI = $\delta(\text{sensitivity}) + \delta(\text{specificity}) = 13.5% - 7.6% = 0.059$.

$\delta(\text{NB}) = (\delta(\text{TP}) - w\delta(\text{FP})) / \text{N} = (10 - 5/95\times 67) / 955 = 0.0068$.

NARI = $\text{Nevent/N} \times \delta(\text{sensitivity}) + \text{Nnonevent/N} \times \delta(\text{specificity}) = 74/955 \times 13.5% + 881/955 \times -7.6% = -0.060$.

Equivalently, the accuracy decreases from 38.1% to 32.1%, or $-6.0%$. Note that NRI is the sum of 2 conditional percentages, and is hence unitless.
In contrast, delta(NB), NARI, and delta(accuracy) are scaled by N and hence interpretable as a percentage of the total sample.

In the case study, both the c statistic, the NRI, and NB give a positive result on the incremental value of CCTA, while NARI suggests a net negative effect, identical to what a simplistic evaluation of accuracy would suggest.

**Role of decision thresholds**
The c statistic considers the range of all possible decision thresholds. The NRI and NARI can be used for multiple thresholds, such as 5%, 5-15%, and 15%, and also for binary decision making with e.g. a risk threshold of 5%. NB is well defined for a single risk threshold, where we can consider different weighting schemes for the improvement in sensitivity and specificity. Both NRI and NARI have no solid basis, in contrast to NB, which is a proper scoring rule with a basis in decision sciences [3,4,5]. NARI and net effect evaluations hence have no place in summarizing the incremental performance of a marker or test.

References:


