Screening for dementia:

Q* index as a global measure of test accuracy revisited

Andrew J Larner

1 Cognitive Function Clinic, Walton Centre for Neurology and Neurosurgery, Liverpool, United Kingdom

Correspondence: AJ Larner, Cognitive Function Clinic, Walton Centre for Neurology and Neurosurgery, Lower Lane, Fazakerley, Liverpool, L9 7LJ, United Kingdom
e-mail: a.larner@thewaltoncentre.nhs.uk

Abstract
Receiver operating characteristic (ROC) curves intersect the downward diagonal through ROC space at a point, the Q* index, where by definition sensitivity and specificity are equal. Aside from its use in meta-analysis, Q* index has also been suggested as a possible global parameter summarising test accuracy of cognitive screening instruments and as a definition for optimal test cut-off. Area under the ROC curve (AUC ROC) is a recognised measure of test accuracy. This study compared different methods for determining Q* index (both graphical and calculation from diagnostic odds ratio) and AUC ROC (integration and calculation from diagnostic odds ratio) using the dataset of a prospective screening test accuracy study of the Mini-Addenbrooke’s Cognitive Examination. The different methods did not agree. DOR-based calculations gave a very sensitive cut-off but with poorer global metrics than the graphical method. DOR-based calculations are not recommended for defining optimal test cut-off or test accuracy.

Keywords: cut-off; dementia; diagnosis; MACE; Q* index; ROC curve; screening accuracy
Introduction
Screening for dementia in clinical practice usually involves the administration of at least one cognitive screening instrument. The definition of test cut-off or dichotomisation point for such instruments has profound implications for screening accuracy.

Optimal test cut-off or decision threshold may be defined in various ways based on a receiver operating characteristic (ROC) plot of the “hit rate” or Sensitivity or true positive rate (TPR) (ordinate) against the false positive rate (FPR), or [1 – Specificity], (abscissa) across the range of possible test scores. Of these methods, the Euclidean index and the Youden index are probably the most commonly used to define optimal cut-off.

The Euclidean index (d) is the point minimally distant from the top left hand (“north west”) corner of the ROC plot, with coordinates (0,1), with the ROC intersect of a line joining these points designated c,\(^1,2\) such that:

\[
d = \min \sqrt{[(1 – \text{Sens}(c))^2 + (1 – \text{Spec}(c))^2]}
\]

Youden index (Y) is a global or unitary measure of test performance given by [Sensitivity + Specificity – 1]. Maximal Y corresponds to the point on the ROC curve which is the maximal vertical distance from the diagonal line (y = x, or Sens = 1 – Spec, or TPR = FPR), the line representing the performance of a random classifier. Also known as c*, this point minimizes misclassification:\(^3\)

\[
Y = \max [\text{Sens}(c^*) + \text{Spec}(c^*) – 1]
\]

When c and c* do not agree, c* is preferable since it maximizes the overall rate of correct classification.\(^4\)

Using a summary ROC (SROC) curve based on data from meta-analyses, another index was proposed to determine optimal test cut-off: the Q* index. This is the point where the downward, negative, or anti-diagonal line through ROC space (y = 1 – x, or Sens = Spec, or Sens – Spec = 0, or TPR = 1 – FPR) intersects the ROC curve.\(^5\) Q* index may also be defined as the “point of indifference on the ROC curve”, where the sensitivity and specificity are equal, or, in other words, where the probabilities of incorrect test results are equal for disease cases and non-cases (i.e. indifference between false positive and false negative diagnostic errors, with both assumed to be of equal value). For the particular case of a symmetrical ROC curve, Q* will be equal to the Euclidean index.\(^6\) Q* index has also been suggested as useful global measure of test accuracy,\(^7\) although its use for the comparison of diagnostic tests is controversial and has been generally discouraged.\(^8\)

As well as being assessed graphically, as the point of intersection of the ROC curve and the anti-diagonal through ROC space, Q* may also be calculated based on its relationship to the diagnostic odds ratio (DOR). DOR, or the cross-product ratio, is the ratio of the product of true positives and true negatives and of false negatives and false positives.\(^9,10\) Q* may be calculated as:\(^6\)

\[
Q^* = \sqrt{\text{DOR}/[1 + \sqrt{\text{DOR}]}}
\]
In addition to defining appropriate cut-offs, clinicians are often interested in the overall accuracy of a test. This may also be measured from the ROC curve as the area under the curve (AUC), typically calculated by integration. AUC may also be calculated, based on its relationship to DOR:

\[
AUC = \frac{DOR}{(DOR - 1)^2 [(DOR - 1) - \ln(DOR)]}
\]

Comparison of Q* index and AUC values determined by these different methods has not, to the author’s knowledge, been previously examined.

Therefore, the aims of the current study were:

- To define and compare optimal cut-offs using the ROC-based indices d, Y, and Q*, the latter assessed both graphically and by calculation using DOR.
- To use the optimal cut-offs thus defined to examine the implications for test sensitivity and specificity and predictive values.
- To compare AUC values for test accuracy calculated both by integration and by calculation using DOR.

These aims were realised by reanalysing the dataset of a large pragmatic prospective screening test accuracy study of the Mini-Addenbrooke’s Cognitive Examination (MACE), a cognitive screening instrument widely used in the assessment of patients suspected to have dementia and lesser degrees of cognitive impairment. MACE comprises tests of attention, memory (7-item name and address), verbal fluency, clock drawing, and memory recall, takes around 5-10 minutes to administer, and has a score range 0-30 (impaired to normal). In the index MACE study, two cut-off points were identified: ≤25/30 had both high sensitivity and specificity (both ≥ 0.85); and ≤21/30 had high specificity (1.00). The current study builds on previously reported data from the pragmatic prospective test accuracy study of MACE.

Methods
Optimal MACE cut-off as defined by minimal Euclidean index (d), by maximal Youden index (Y), and by Q* index was determined. Plots of d and Y against MACE cut-offs were used to determine their optima. Q* index was measured graphically, by reading off the point of intersection of the ROC plot with the anti-diagonal.

Q* index was also calculated at each MACE cut-off, encompassing the range of MACE cut-offs defined in the index study, from the corresponding value of DOR, using Walter’s formula:

\[
Q^* = \sqrt{DOR}/[1 + \sqrt{DOR}]
\]

At each of the “optimal” cut-offs thus defined by d, Y, and Q* index, test sensitivity, specificity, and predictive values were then calculated and the results compared. Clinical utility indexes (CUI+, CUI-), parameters indicative of the value of a
diagnostic method for ruling in or ruling out a diagnosis, were also calculated and classified,\textsuperscript{14} where:

\[
\begin{align*}
CUI^+ &= Sens \times PPV \\
CUI^- &= Spec \times NPV
\end{align*}
\]

From these, the global summary screening utility index (SSUI) was calculated and classified:\textsuperscript{15}

\[
SSUI = CUI^+ + CUI^-
\]

Test accuracy was assessed by means of the area under the ROC curve (AUC). AUC was calculated by the standard method of integration with 95% confidence intervals, and also by calculation, using Walter’s formula:\textsuperscript{6}

\[
AUC = \frac{DOR}{(DOR - 1)^2[(DOR - 1) - \ln(DOR)]}
\]

**Results**

Optimal MACE cut-off differed according to method. Plots of d and Y versus MACE test cut-offs showed minimum d at \( \leq 19/30 \), and maximum Y at \( \leq 20/30 \). Graphically, Q* index was 0.8 at cut-off \( \leq 18/30 \) (Figure 1).

By calculation, Q* index varied with diagnostic OR at each test cut-off (Table 1, 3\textsuperscript{rd} column), with the optimal value at \( \leq 23/30 \). Of note, the calculated Q* index at cut-off \( \leq 18/30 \) corresponded with the value (0.8) determined graphically.

Comparison of test sensitivity, specificity, and predictive values at each of the defined “optimal” cut-offs showed variations in all parameters, with the exception of NPV which remained fairly constant (Table 2). Sensitivity and specificity were not equal for the calculated Q* index, unlike the graphically determined Q* index.

Clinical utility indexes and summary screening utility index values for each cut-off method showed the calculated Q* index to have the poorest performance of the methods examined (Table 3).

AUC by integration was 0.886 (95% confidence interval 0.856-0.916).\textsuperscript{11} Calculated AUC showed a maximum of 0.941 at the optimal cut-off of \( \leq 23/30 \) (Table 1, 4\textsuperscript{th} column). The calculated AUC most closely corresponding to the AUC by integration was between the cut-offs of \( \leq 19/30 \) and \( \leq 20/30 \), the optima defined by d and Y respectively.

**Discussion**

Optimal cut-off values defined by different ROC-based methods differ, and this has implications for test performance in terms of the sensitivity, specificity, and predictive values calculated from these cut-offs.

Symmetrical ROC curves have constant DOR. Hence for the ROC diagonal line (\( y = x \), or Sens = 1 – Spec, or TPR = FPR), DOR = 1, indicative of a useless test. In an
asymmetric ROC plot, the more usual situation in clinical practice (Figure 1), DOR varies with test cut-off or threshold, with corresponding variation in the value of $Q^*$ index calculated using Walter’s equation. For the anti-diagonal line ($y = 1 - x$, or Sens = Spec, or Sens – Spec = 0, or TPR = 1 – FPR), DOR will vary from $\infty$ at coordinates (0,1) through 1 at the intersection with the ROC diagonal (Sens = Spec = 0.5) to $-\infty$ at the bottom right hand (“south east”) corner at coordinates (1,0).

In the current study, the optimal MACE cut-off defined by $Q^*$ index by calculation from DOR appeared to be an outlier compared to the other methods examined for defining optimal cut-off. This confirmed the findings of a previous study looking at the use of various global measures to define MACE cut-offs where optimal cut-off defined by DOR was an outlier.\textsuperscript{13} This higher cut-off gave better sensitivity but poorer specificity and positive predictive value than the other defined “optimal” cut-offs (Table 2). Furthermore, the AUC by calculation from DOR gave a higher value of AUC than by integration.

Clinicians desirous of using highly sensitive tests to avoid missing cases of dementia, even at the cost of more false positives, might therefore be inclined to use the DOR-based method to define test cut-off. However, it is recognised that DOR tends to give the most optimistic results by choosing the best quality of a test and ignoring its weaknesses, producing extremely high cut-points and treating false negatives and false positives as equally undesirable. Moreover, DOR may be less intuitive as a test measure than sensitivity and specificity.\textsuperscript{8} It may be noted that, of the methods examined here, $Q^*$ index by calculation from DOR gave the lowest value for the Youden index (Table 2), and for clinical utility indexes and the summary screening utility index (Table 3).

Hence this study provides additional evidence of the shortcomings of DOR and DOR-based calculations\textsuperscript{16,17} and its use cannot therefore be recommended for defining optimal test cut-off or test accuracy.
References

Figures and Tables

Figure 1: MACE ROC plot for diagnosis of dementia, with chance diagonal ($y = x$) and anti-diagonal ($y = 1 - x$) lines, the latter permitting the graphical estimate of $Q^*$ index.
Table 1: Q* index and AUC values calculated from diagnostic odds ratio at each MACE cut-off value

<table>
<thead>
<tr>
<th>MACE cut-off</th>
<th>DOR</th>
<th>Calculated Q* index</th>
<th>Calculated AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤25/30</td>
<td>52.0</td>
<td>0.878</td>
<td>0.941</td>
</tr>
<tr>
<td>≤24/30</td>
<td>38.7</td>
<td>0.862</td>
<td>0.927</td>
</tr>
<tr>
<td>≤23/30</td>
<td>52.8</td>
<td>0.879</td>
<td>0.941</td>
</tr>
<tr>
<td>≤22/30</td>
<td>47.7</td>
<td>0.873</td>
<td>0.937</td>
</tr>
<tr>
<td>≤21/30</td>
<td>32.2</td>
<td>0.850</td>
<td>0.917</td>
</tr>
<tr>
<td>≤20/30</td>
<td>25.1</td>
<td>0.834</td>
<td>0.902</td>
</tr>
<tr>
<td>≤19/30</td>
<td>18.9</td>
<td>0.813</td>
<td>0.882</td>
</tr>
<tr>
<td>≤18/30</td>
<td>16.0</td>
<td>0.800</td>
<td>0.869</td>
</tr>
<tr>
<td>≤17/30</td>
<td>12.8</td>
<td>0.782</td>
<td>0.850</td>
</tr>
</tbody>
</table>

Abbreviations: DOR = diagnostic odds ratio; AUC = area under the ROC curve
Table 2: MACE sensitivity, specificity, Youden index, and predictive values at optimal cut-offs defined by various global metrics

<table>
<thead>
<tr>
<th>Metric</th>
<th>Optimal cut-off</th>
<th>Sens</th>
<th>Spec</th>
<th>Y</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q* index (graphically)</td>
<td>≤18/30</td>
<td>0.80</td>
<td>0.80</td>
<td>0.60</td>
<td>0.42</td>
<td>0.96</td>
</tr>
<tr>
<td>d</td>
<td>≤19/30</td>
<td>0.86</td>
<td>0.76</td>
<td>0.62</td>
<td>0.38</td>
<td>0.97</td>
</tr>
<tr>
<td>Y</td>
<td>≤20/30</td>
<td>0.91</td>
<td>0.71</td>
<td>0.62</td>
<td>0.36</td>
<td>0.98</td>
</tr>
<tr>
<td>Q* index (by calculation)</td>
<td>≤23/30</td>
<td>0.98</td>
<td>0.49</td>
<td>0.47</td>
<td>0.25</td>
<td>0.99</td>
</tr>
</tbody>
</table>

Abbreviations: d = Euclidean index; NPV = negative predictive value; PPV = positive predictive value; Y = Youden index
Table 3: MACE clinical and screening summary utility indexes and their classification at the optimal cut-offs defined by various global metrics

<table>
<thead>
<tr>
<th>Metric</th>
<th>Optimal cut-off</th>
<th>CUI+</th>
<th>CUI-</th>
<th>SSUI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q* index (graphically)</td>
<td>≤18/30</td>
<td>0.336 (very poor)</td>
<td>0.768 (good)</td>
<td>1.104 (adequate)</td>
</tr>
<tr>
<td>d</td>
<td>≤19/30</td>
<td>0.327 (very poor)</td>
<td>0.737 (good)</td>
<td>1.064 (adequate)</td>
</tr>
<tr>
<td>Y</td>
<td>≤20/30</td>
<td>0.378 (poor)</td>
<td>0.696 (good)</td>
<td>1.023 (adequate)</td>
</tr>
<tr>
<td>Q* index (by calculation)</td>
<td>≤23/30</td>
<td>0.245 (very poor)</td>
<td>0.485 (poor)</td>
<td>0.730 (poor)</td>
</tr>
</tbody>
</table>

Abbreviations: d = Euclidean index; CUI+, CUI- = positive and negative clinical utility index; SSUI = summary screening utility index; Y = Youden index