



This is a repository copy of *What is an ROC curve?*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/113754/>

Version: Accepted Version

Article:

Hoo, Z.H. orcid.org/0000-0002-7067-3783, Candlish, J. and Teare, M.D. (2017) What is an ROC curve? *Emergency Medicine Journal*, 34 (6). pp. 357-359. ISSN 1472-0205

<https://doi.org/10.1136/emmermed-2017-206735>

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Emergency Medicine Journal**What is a ROC curve?**

Journal:	<i>Emergency Medicine Journal</i>
Manuscript ID	Draft
Article Type:	Concepts
Date Submitted by the Author:	n/a
Complete List of Authors:	Hoo, Zhe Hui; University of Sheffield, School of Health and Related Research (SchARR) Candlish, Jane; University of Sheffield, SchARR Teare, Dawn; University of Sheffield, School of Health and Related Research
Keywords:	statistics

SCHOLARONE™
Manuscripts

Title:

What is a ROC curve?

Author list:

Zhe Hui Hoo^{1,2}

Jane Candlish¹

M Dawn Teare¹

¹ School of Health and Related Research (ScHARR), University of Sheffield, Sheffield, UK

² Sheffield Adult CF Centre, Northern General Hospital, Sheffield, UK

Corresponding author information:

Zhe Hui Hoo

Room 1.03, Innovation Centre, 217 Portobello, Sheffield S1 4DP, UK

Email: z.hoo@sheffield.ac.uk

Telephone: 0114 226 6281

Word count: 1268 words

Key words: ROC curve, diagnostic accuracy

ABSTRACT

The paper by Body et al is concerned with the evaluation of decision aids which can be used to identify potential acute coronary syndromes (ACS) in the ED. The authors previously developed the Manchester Acute Coronary Syndromes model (MACS) decision aid which uses several clinical variables and two biomarkers to 'rule in' and 'rule out' ACS. However, one of the two biomarkers (H-FABP) is not widely used so a revised decision aid has been developed (Troponin-only Manchester Acute Coronary Syndromes, T-MACS) which include a single biomarker hs-cTnT. In this issue, the authors show how they derive a revised decision aid and describe its performance in a number of independent diagnostic cohort studies. Decision aids (as well as other types of "diagnostic tests") are often evaluated in terms of diagnostic testing parameters such as the area under the ROC curve, sensitivity and specificity. In this article, we explain how the ROC analysis is conducted and why it is an essential step towards developing a test with the desirable levels of sensitivity and specificity.

INTRODUCTION

Making a diagnosis is an important role for clinicians, and often a range of tests are used to supplement history-taking and examination. However, not all diagnostic tests are equal, hence it is important to evaluate the diagnostic abilities of those tests.

What physicians want to know is whether a specific condition is present, or not present. Diagnostic tests can produce either yes / no answers (binary results) or a number on a continuous scale. In the diagnosis of deep vein thrombosis (DVT), for example, Doppler ultrasound would produce a binary result (DVT present or absent). However, the serum D-dimer level used to detect DVT gives a number in a range of possible values. Diagnostic tests that produce continuous results are often dichotomised because the outcome of interest is generally binary. An example would be the threshold (or cut-off point) of 500 micrograms/L to distinguish between a 'positive' and a 'negative' D-dimer result.[1]

One of the challenges in interpreting the results of diagnostic tests that produce continuous measures is the selection of the threshold to distinguish a 'positive' test from a 'negative' test. We used the paper by Body et al[2] as an example to illustrate the role of receiver operating characteristic (ROC) curve in choosing the threshold cut-offs for the newly derived T-MACS score. The original MACs score used two thresholds to aid clinical decision making. A low threshold was set to 'rule-out' a diagnosis of ACS and a higher threshold was used to 'rule-in' a diagnosis of ACS. Patients with test results falling between these two thresholds would require further clinical observation. The T-MACS (and the original MACS) score estimate the probability of ACS from 0 (lowest probability) to 1 (highest probability).

GENERATING A ROC CURVE

To produce a ROC curve, the sensitivities and specificities for different values of a continuous test measure are first tabulated. This results, essentially, in a list of a various test values and the corresponding sensitivity and specificity of the test at that value. Then, the graphical ROC curve is produced by plotting sensitivity (true positive rate) on the y-axis against 1– specificity (false positive rate) on the x-axis for the various values tabulated.

USING A ROC CURVE TO UNDERSTAND THE DIAGNOSTIC VALUE OF A TEST

A ROC curve that follows the diagonal line $y = x$ produces false positive results at the same rate as true positive results. Therefore we expect a diagnostic test with reasonable accuracy to have a ROC curve in the upper left triangle above the $y = x$ line ('reference line') of Figure 1.

The area under the ROC curve (AUC) is a global measure of a test's ability to discriminate whether a specific condition is present or not present. An AUC of 0.5 represents a test with no discriminating ability (i.e. no better than chance) while an AUC of 1.0 represents a test with perfect discrimination. The (unpublished) ROC curve for ACS (Figure 1) which was generated from the T-MACS scores calculated for the derivation set has an AUC of 0.94 after correction for in-sample optimism by cross-validation, which would suggest that T-MACS score is a very good discriminator of ACS vs no ACS.

When selecting an optimal threshold (or cut-off point) we need to consider the aims of the diagnostic test, considering the significance and costs of a false positive or false negative interpretation. A commonly used approach when selecting a cut-off point is to give equal weight to the importance of sensitivity and specificity by choosing the point nearest to the top-left most corner of the ROC curve.[3] This point is also known as the Youden Index.[4] We illustrate how this is done using Figure 1. If the interest was to simply find the optimal trade-off between sensitivity and specificity then a T-MACS score of ≥ 0.34 represents this point of balance. We can read from the ROC curve that sensitivity is 83.4% and specificity is 94.6% at this threshold. If we use this threshold, then the T-MACS score has an accuracy of 92.2% in detecting ACS among adults presenting to ED with chest pain in the study. Accuracy is defined as the proportion of correct decisions (i.e. true positive + true negative) over the number of total tested.[5] The AUC can be thought of an indicator of overall "accuracy", and there is only one AUC value for each ROC curve. On the other hand, 'accuracy' will vary from point-to-point along the ROC curve, as there is variation in sensitivity and specificity from point-to-point (see Table 1).

Table 1: The sensitivity and specificity of the various T-MACS thresholds for ACS (derivation set)

Threshold for a positive test	Sensitivity (95% CI)	Specificity (95% CI)	Accuracy (95% CI)
Low risk (i.e. very low risk = negative test)	98.7% (95.3% – 99.8%)	47.6% (43.4% – 51.9%)	58.6 (55.0 – 62.3)
Moderate risk (i.e. very low risk & low risk = negative test)	94.0% (89.0% – 97.2%)	68.5% (64.4% – 72.3%)	74.0 (70.7 – 77.2)
High risk (i.e. very low risk, low risk & moderate risk = negative test)	47.0% (38.9% – 55.3%)	100.0% (99.3% – 100.0%)	88.6 (86.3 – 91.0)

Sensitivity is the proportion of true positives that are correctly identified by the test
Specificity is the proportion of true negatives that are correctly identified by the test
Accuracy is the proportion of correct decisions (i.e. true positive + true negative) over the number of total tested

By selecting the threshold based on the trade-off between sensitivity and specificity, we are assuming that the balance between false positive and false negative rates is not clinically important.[6] This may not be true in two common scenarios. The first scenario is when the clinical risk of missing a particular diagnosis is disproportionate to the risk of over-diagnosis (or vice versa). The second scenario is when the primary intended use of a diagnostic test is for either "ruling-in" or "ruling-out" a target condition. In these scenarios, it is important to understand the balance between the two types of errors and to select the threshold accordingly.

Body et al selected a threshold of 0.02 for T-MACS and classify patients below this value as 'ruled-out' (i.e. very low risk) using the derivation set. This threshold has very high sensitivity (98.7%) – so if a patient score is this low, we are very confident they are at very low risk of ACS. However, scores above this threshold are only 47.6% specific, hence the accuracy of detecting ACS with this threshold is only 58.6%. As noted above, a T-MACS score threshold of 0.34 represents the optimal balance between sensitivity and specificity, thus has a higher accuracy of 92.2%. However, this threshold only has a sensitivity of 83.4% for detecting ACS. This means around 16.6% of adults presenting to ED with chest pain that have ACS will be undetected if a T-MACS score of < 0.34 is considered a 'negative' test result. Therefore, despite the lower accuracy of with a threshold of 0.02, it is a more useful threshold for 'ruling out' ACS since less than 2% of ACS cases will be undetected.

Body et al also selected a higher threshold of 0.95 for 'ruling in' ACS. This threshold also has a modest accuracy of 88.6%, due to the low sensitivity of 47.0%. However, this threshold has a specificity of 100%, which means there is no false positive result. We can therefore be confident that a patient at this higher threshold is at a very high risk for ACS. In contrast, a T-MACS score of ≥ 0.34 will result in a false positive rate of 5.4% despite its superior accuracy of 92.2%.

Table 1 represents the sensitivity and specificity of the various T-MACS thresholds for ACS using the derivation set, linking to the results tabulated in Table 4 of the paper by Body et al. Table 1 illustrates the rising specificity and declining sensitivity with higher threshold values.

A strength of the study by Body et al is the re-evaluation of the selected thresholds using a validation set. This yielded very similar results to the derivation set (e.g. the sensitivity for the 'ruled out' threshold for ACS remains very high at 98.1%), suggesting that the thresholds selected are indeed robust.

TAKE HOME POINTS

The ROC curve can be used to determine an appropriate cut-off, affecting the sensitivity and specificity of the test. The selection of a test threshold depends on the purpose of the test and not necessarily just by giving equal weight to sensitivity and specificity to achieve higher accuracy.

COMPETING INTERESTS: None declared

REFERENCES

1. Perrier A, Desmarais S, Goehring C, et al. D-dimer testing for suspected pulmonary embolism in outpatients. *Am J Respir Crit Care Med* 1997;156:492-6.
2. Body R, Carlton E, Sperrin M, et al. Troponin-only Manchester Acute Coronary Syndromes (T-MACS) decision aid: single biomarker re-derivation and external validation in three cohorts. *Emerg Med J* 2016 Aug 26. pii: emermed-2016-205983. doi: 10.1136/emered-2016-205983. [Epub ahead of print]
3. Altman DG, Bland JM. Diagnostic tests 3: receiver operating characteristic plots. *BMJ* 1994;309:188.
4. Schisterman EF, Perkins NJ, Liu A, et al. Optimal cut-point and its corresponding Youden Index to discriminate individuals using pooled blood samples. *Epidemiology* 2005;16:73-81.
5. Metz CE. Basic principles of ROC analysis. *Semin Nucl Med* 1978;8:283-98.
6. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem* 1993;39:561-77.

Figure 1: T-MACS ROC curve for ACS using the data from derivation set

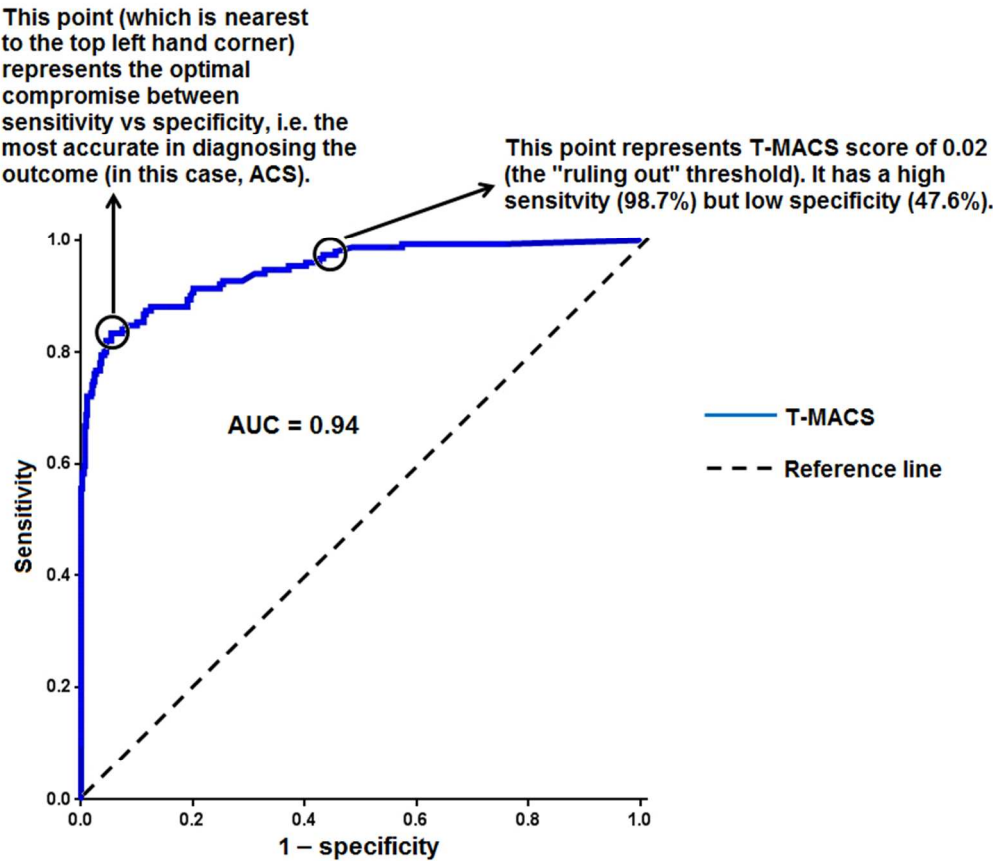


Figure 1: T-MACS ROC curve for ACS using data from derivation set
Figure 1
221x207mm (96 x 96 DPI)