

## **EBM Methods Verdict**

### **How to reduce diagnostic error: the 'neutral zone' approach**

(Based on: Jeske DR, Zhang Z, Smith S. Construction, visualization and application of neutral zone classifiers. Stat Methods Med Res 2019. doi: 10.1177/0962280219863823.)

**Thomas R Fanshawe & Maria Vazquez-Montes**

**Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK**

*Many diagnostic tests report only positive or negative diagnoses, as if these results were definitive. Should more tests allow a 'neutral zone', corresponding to an uncertain diagnosis?*

Diagnostic tests often use single continuous biomarker measurements to return a positive or a negative result. However, in clinical practice a diagnosis might often be reached based on multiple signs, symptoms and test results rather than using only a single measurement. These multiple individual characteristics are sometimes incorporated into a statistical model that aims to estimate the probability of disease, which helps in making a diagnosis. Guidelines have been published for optimal development, validation and report of such tools.<sup>1,2</sup>

The Receiver Operating Characteristic (ROC) curve remains a popular technique for showing diagnostic performance in terms of sensitivity and specificity, i.e. how accurately the test can correctly classify people with and without disease. Traditionally, statistical methods that follow this approach assume that individuals can be classified into just two categories: either a positive or a negative diagnosis.

The recent paper by Jeske et al.<sup>3</sup> considers methods for identifying what has been termed a 'neutral zone' – that is, the scenario in which the prediction of an individual's diagnostic category is uncertain, and cannot be definitively classified as either positive or negative. Their approach is a two-stage modelling process that aims to improve false positive and false negative rates, and ultimately to reduce costs by identifying the subset of individuals for whom potentially more expensive laboratory tests are needed.

This idea is appealing and appears to carry several conceptual advantages. Firstly, it explicitly recognises the uncertainty inherent in many diagnostic procedures by only offering a firm diagnosis to those individuals for whom the predicted probability of disease is sufficiently high or low. This may help to reduce diagnostic error for individuals in the mid-range. Secondly, it gives a more accurate reflection of the true diagnostic performance in the groups of individuals classified as either positive or negative, without diluting the estimated sensitivity and specificity. Thirdly, it might

indicate that individuals with a 'neutral' diagnosis require additional subsequent tests, while reducing the need for unnecessary further testing for those in whom a confident diagnosis can already be made. In this sense, the proposed analytical procedure reflects the way a diagnosis is often reached in clinical practice.

Multi-stage, or sequential, approaches to diagnostic testing are not new in the diagnostic methodology literature – see for example the 2003 paper by Thompson<sup>4</sup> for a discussion of relevant issues. The method proposed by Jeske et al.<sup>3</sup> is closely related to simple 'AND/OR' rules for combining diagnostic tests that have been explored elsewhere,<sup>5,6</sup> but formalises the way in which they are constructed, illustrated using an example for predicting prostate cancer recurrence. Based on cross-validation results presented for this example, it appears possible that overfitting to the data may be an issue, and this may require further investigation.

In common with all statistical methods for developing diagnostic or prediction tools for use in clinical practice, the method requires some consideration of the context in which the prediction will be used. In the case of a simple positive/negative classifier, this usually means evaluation of the relative costs (broadly defined) of making false positive and false negative diagnoses, which are not always equal. The two-stage method would additionally require information about the costs of performing further tests beyond the first classification stage. If a large number of individuals are classified into the 'neutral zone', performing these extra tests may increase the total financial cost of the diagnostic test sequence, and this must be balanced against the potential gain in diagnostic accuracy, while recognising that a certain level of misdiagnosis is unavoidable in many clinical scenarios. It may therefore be necessary to complement these methods with a health economic analysis.

The paper by Jeske et al.<sup>3</sup> also introduces a 4-way plot that combines the standard ROC plot, showing how sensitivity and specificity vary together as the threshold for test positivity changes, with the predictive ROC (pROC) plot, which provides similar information based on positive and negative predictive values. Each single point in the plot gives information about the four performance measures at a glance. This plot also makes it possible to read off the probability of a classification falling inside the neutral zone for disease positive and disease negative cases. Although useful for illustrating the development of the method, the standard ROC plot is so well established in the diagnostic research literature that it seems unlikely that the new plot will become widely used in the near future without additional dissemination and implementation in an easily accessible statistical package.

**VERDICT:** When developing or evaluating diagnostic test sequences, formally allowing the possibility of a 'neutral zone' to reflect an uncertain diagnosis may help to reduce rates of misdiagnosis and the amount of testing required. Following a two-stage method could be helpful but prediction modelling methodology guidelines, which could help to avoid model overfitting, and the effect on patient benefit and cost-effectiveness should be considered.

#### References

1. Bossuyt PM, Reitsma JB, Bruns DE et al. STARD 2015: An updated list of essential items for reporting diagnostic accuracy studies. *BMJ* 2015;351:h5527.
2. Collins GS, Reitsma JB, Altman DG et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *Ann Intern Med* 2015;162(1):55-63.
3. Jeske DR, Zhang Z, Smith S. Construction, visualization and application of neutral zone classifiers. *Stat Methods Med Res* 2019. doi: 10.1177/0962280219863823.
4. Thompson ML. Assessing the diagnostic accuracy of a sequence of tests. *Biostatistics* 2003;4(3):341-51.
5. Macaskill P, Walter SD, Irwig L, Franco EL. Assessing the gain in diagnostic performance when combining two diagnostic tests. *Stat Med* 2002;21:2527-46.
6. Zou KH, Bhagwat JG, Carrino JA. Statistical combination schemes of repeated diagnostic test data. *Acad Radiol* 2006;13(5):566-72.