Supplementary Material

**Manuscript title:**

Clinical Decision Support in Primary Care for better Diagnosis and Management of Retinal Disease: A review

**Authors:**

Sharon Ho\*† BOptom (Hons), BSc

Michael Kalloniatis\*† PhD, MScOptom, GradCertOcTher, BSc (Optom), FAAO

Angelica Ly\*† PhD, GradCertOcTher, BOptom (Hons), FAAO

**Author institutions:**

\*Centre for Eye Health, The University of New South Wales, Sydney, Australia

†School of Optometry and Vision Science, The University of New South Wales, Sydney, Australia

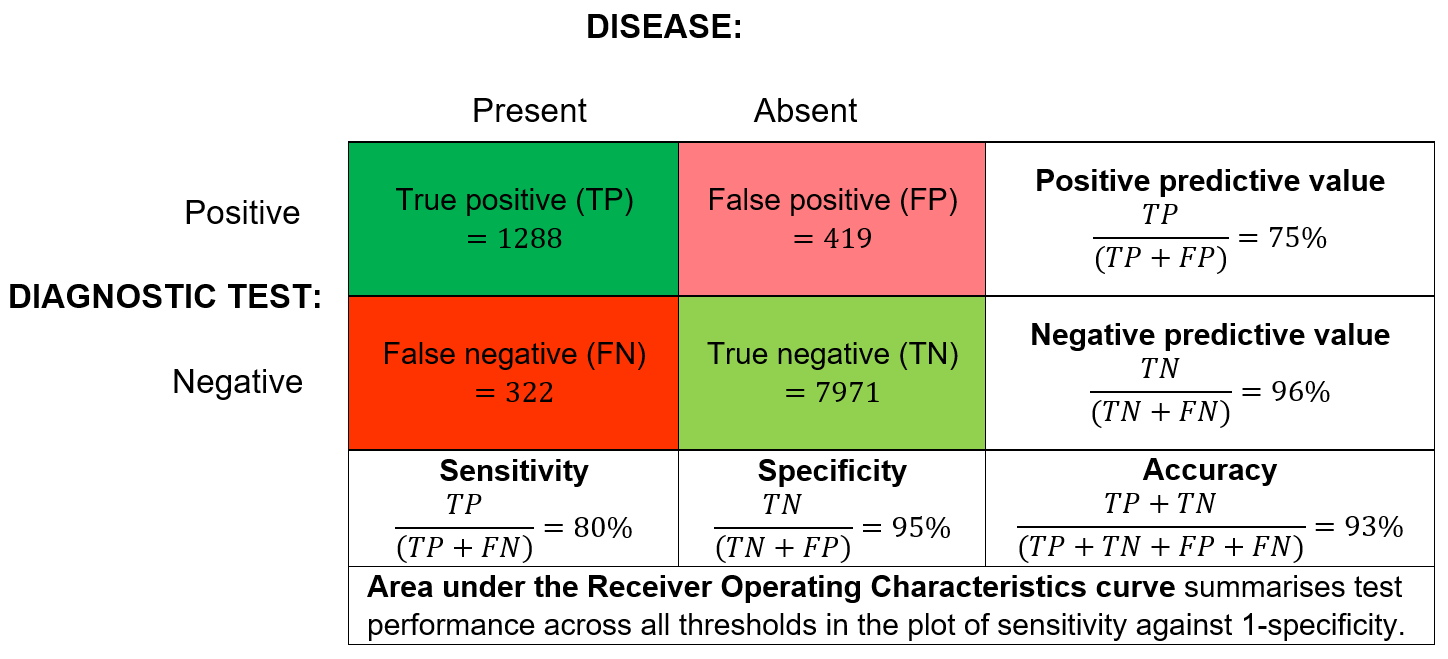
**Running title:** Clinical decision support

**Keywords:** clinical decision support; evidence based medicine; information technology; primary care; retinal disease.

**Statistical measures of diagnostic accuracy**

Diagnostic test performance is commonly described through four key parameters: true positive (TP), true negative (TN), false negative (FN), and false positive (FP). They can be reflected graphically in a Receiver Operating Characteristics (ROC) curve. As illustrated in Supplementary Figure 1, if a disease is present in a patient, and the given diagnostic test is also positive for disease, the test result is TP. Likewise, if a disease is absent in a patient, and the diagnostic test is negative, the test result is TN.

Accuracy is calculated as the sum proportion of true results, either true positive or true negative, in a population.1 Sensitivity is the proportion of true positives that are correctly identified, and specificity is the proportion of true negatives that are correctly identified.1 Although these are widely applied and familiar concepts to clinicians, they are not without their drawbacks. The calculated example in Supplementary Figure 1 depicts a pool of 10,000 patients with diabetes, where an assumed prevalence of diabetic retinopathy of 16.1 per cent (average across two studies)2,3 indicates 1,610 patients will have diabetic retinopathy. Assuming 80 per cent sensitivity and 95 per cent specificity as has been reported suitable criteria for diabetic retinopathy screening,4,5 1,288 cases and 7,971 normal patients will be correctly identified. Yet 419 normal patients will be falsely identified as having diabetic retinopathy and 322 cases with diabetic retinopathy will be missed. Therefore, actual diagnostic accuracy remains substandard despite high performance measures.

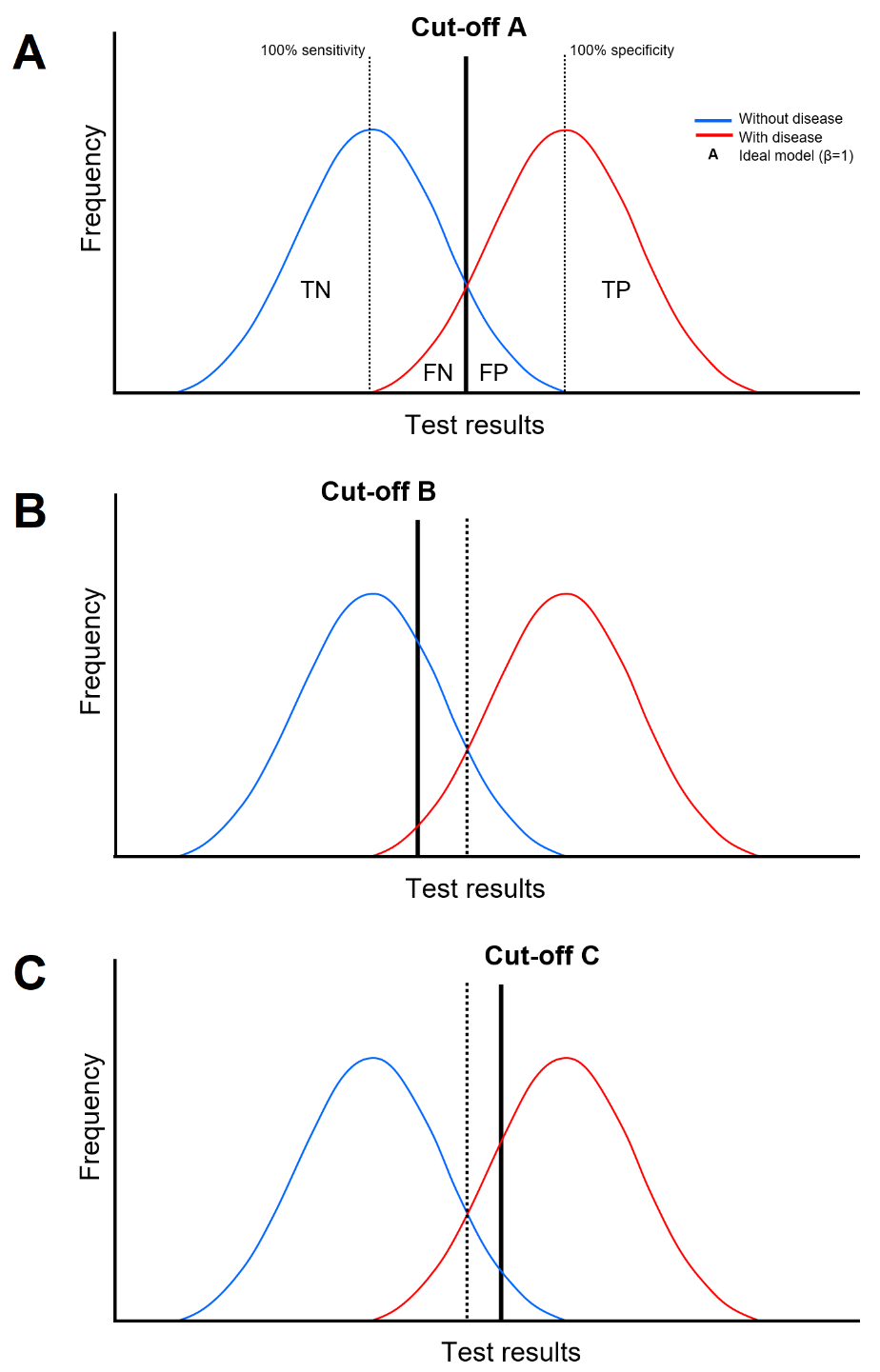
****

**Supplementary Figure 1.** Common measures of diagnostic test performance using principles of signal detection theory: Accuracy, Sensitivity, Specificity, Positive predictive value, Negative predictive value, Area under the Receiver Operating Characteristics curve. The calculated numerical values reference a hypothetical example of diabetic retinopathy screening with 80% sensitivity and 95% specificity in a patient population of n=10,000, described in-text.

**Signal detection theory as continuous Gaussian distributions**

The relationships between TP, TN, FN and FP may be portrayed using Gaussian distributions comparing the ‘normal’ population and the ‘disease’ population. The ideal diagnostic model (Supplementary Figure 2A) is represented by the criterion, or cut-off, with the highest sensitivity (highest TP rate) together with the highest specificity (lowest FP rate). Theoretically this would imply 100 per cent sensitivity and 100 per cent specificity, provided the two distributions are separated by a sufficient amount. However, such perfect tests do not exist in real life since we expect some failure to separate normal and abnormal patients due to the overlap in the ‘normal’ and ‘disease’ distributions (Supplementary Figure 2A).6 Therefore systems are reported to err towards either higher sensitivity or higher specificity based upon the criterion chosen.

Different disease conditions should be managed with different criteria for test performance that reflect the clinical consequences of misdiagnosis.6 For example in comparing neovascular age-related macular degeneration against pachychoroid pigment epitheliopathy, the visual prognosis in the event of misdiagnosis is typically much worse in the former than the latter. A diagnostic system for neovascular age-related macular degeneration should therefore aim for higher sensitivity to minimise FN so that fewer patients with disease are missed. Shifting the criterion cut-off to the left will achieve this (Supplementary Figure 2B). Simultaneously, FP will increase as a necessary compromise. Conversely, shifting the criterion cut-off to the right will result in a higher specificity but lower sensitivity diagnostic system (Supplementary Figure 2C). In this way, sensitivity and specificity are often described to “*move in different [opposite] directions*”.6 A higher sensitivity test would be most reliable and useful when the result is negative, or when ruling out the presence of disease, since it is unlikely to diagnose a diseased patient as negative. On the other hand, a higher specificity test would be most useful when the result is positive, or when ruling in patients with disease.



# Supplementary Figure 2. A) Gaussian distribution showing the relationships between TP/TN/FN/FP. Cut-off A represents the ideal model where β=1. This criterion provides the maximum TP with minimum FP. B) Gaussian distribution showing cut-off B to the left of β=1 (dotted line). This diagnostic system will have a higher sensitivity but lower specificity. C) Gaussian distribution showing cut-off C to the right of β=1 (dotted line). This diagnostic system will have a higher specificity but lower sensitivity.

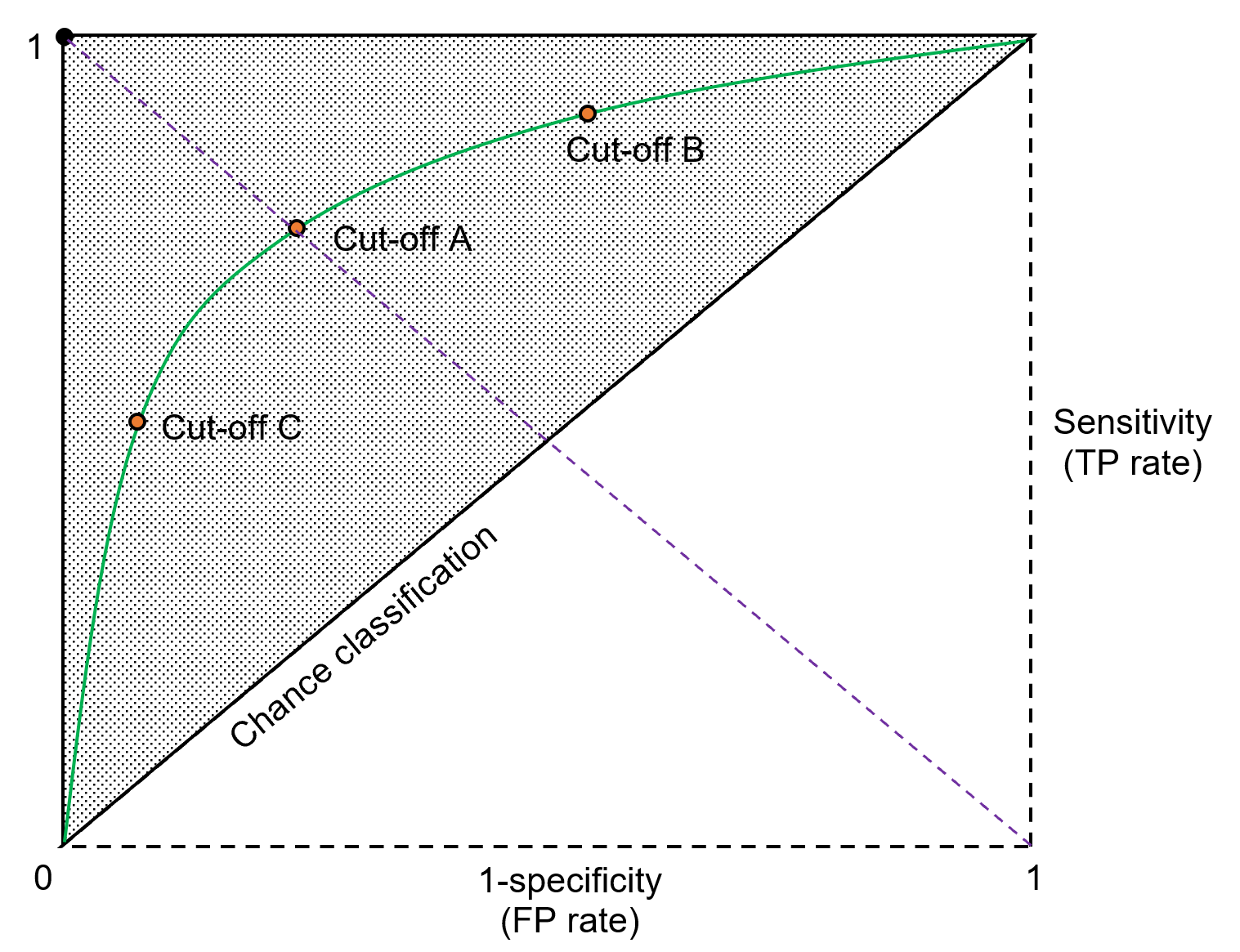
**Area under the ROC curve**

Where Gaussian distributions illustrate performance for a single cut-off, the ROC curve plots the connection or trade-off between sensitivity and specificity for every possible cut-off for a diagnostic test.1 ROC curves convey the overall diagnostic ability of a system, and computing the area under the ROC curve (AUC) is one way to summarise it into a single value.7 Supplementary Table 1 reports the general accuracy classification by AUC for a diagnostic system.

**Supplementary Table 1.** Accuracy classification by AUC for a diagnostic system.8

|  |  |
| --- | --- |
| *AUC Range* | *Classification* |
| 0.5 | No discrimination (diagnostic ability as expected by chance) |
| 0.7 to 0.8 | Acceptable |
| 0.8 to 0.9 | Excellent |
| 0.9 to 1.0 | Outstanding |

While AUC is frequently used in the literature to summarise test performance, the metric falls short in terms of its clinical meaning for practitioners. Assume the AUC is 0.8 for the ROC curve in Supplementary Figure 3, where the cut-offs A, B and C correspond to those depicted in the Gaussian distributions in Supplementary Figure 2. Since the points lie on the same curve, all three cut-offs will produce a system of AUC=0.8. However, clearly the clinical performance of these systems will vary as they are each represented by different sensitivity and specificity values. An AUC value on its own must be interpreted with caution.



**Supplementary Figure 3.** The ROC graph. i) ROC curve plots sensitivity vs. (1-specificity) for all possible cut-off points, ii) Green curve represents AUC=0.8 in the in-text hypothetical example, iii) Diagonal (purple dotted line) joining coordinates (1,0) and (0,1) represents cut-offs where β=1, iv) Diagonal joining coordinates (0,0) and (1,1) represents “chance classification” or AUC=0.5, v) Shadow area represents “better than chance” classification or AUC>0.5, vi) Coordinate (0,1) represents the perfect model where AUC=1.0 (100% sensitivity, 100% specificity).

**References**

1 Zhu W, Zeng N, Wang N. Sensitivity, Specificity, Accuracy, Associated Confidence Interval and ROC Analysis with Practical SAS ® Implementations. *NorthEast SAS users group, health care and life sciences* 2010.

2 Keel S, Xie J, Foreman J et al. The Prevalence of Diabetic Retinopathy in Australian Adults with Self-Reported Diabetes: The National Eye Health Survey. *Ophthalmology* 2017; 124: 977-984.

3 Tapp RJ, Shaw JE, Harper CA et al. The prevalence of and factors associated with diabetic retinopathy in the Australian population. *Diabetes Care* 2003; 26: 1731-1737.

4 Marshall S. The Exeter BDA Meeting—a Synopsis. *Diabetic Medicine* 1988; 5: iii-iv.

5 Squirrell DM, Talbot JF. Screening for diabetic retinopathy. *J R Soc Med* 2003; 96: 273-276.

6 Halligan S, Altman DG, Mallett S. Disadvantages of using the area under the receiver operating characteristic curve to assess imaging tests: A discussion and proposal for an alternative approach. *European Radiology* 2015; 25: 932-939.

7 Hajian-Tilaki K. Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. *Caspian J Intern Med* 2013; 4: 627-635.

8 Mandrekar JN. Receiver Operating Characteristic Curve in Diagnostic Test Assessment. *Journal of Thoracic Oncology* 2010; 5: 1315-1316.