Problems with the partial AUC and standardized partial area resolved by the concordant partial AUC

André Carrington, Ph.D, P.Eng

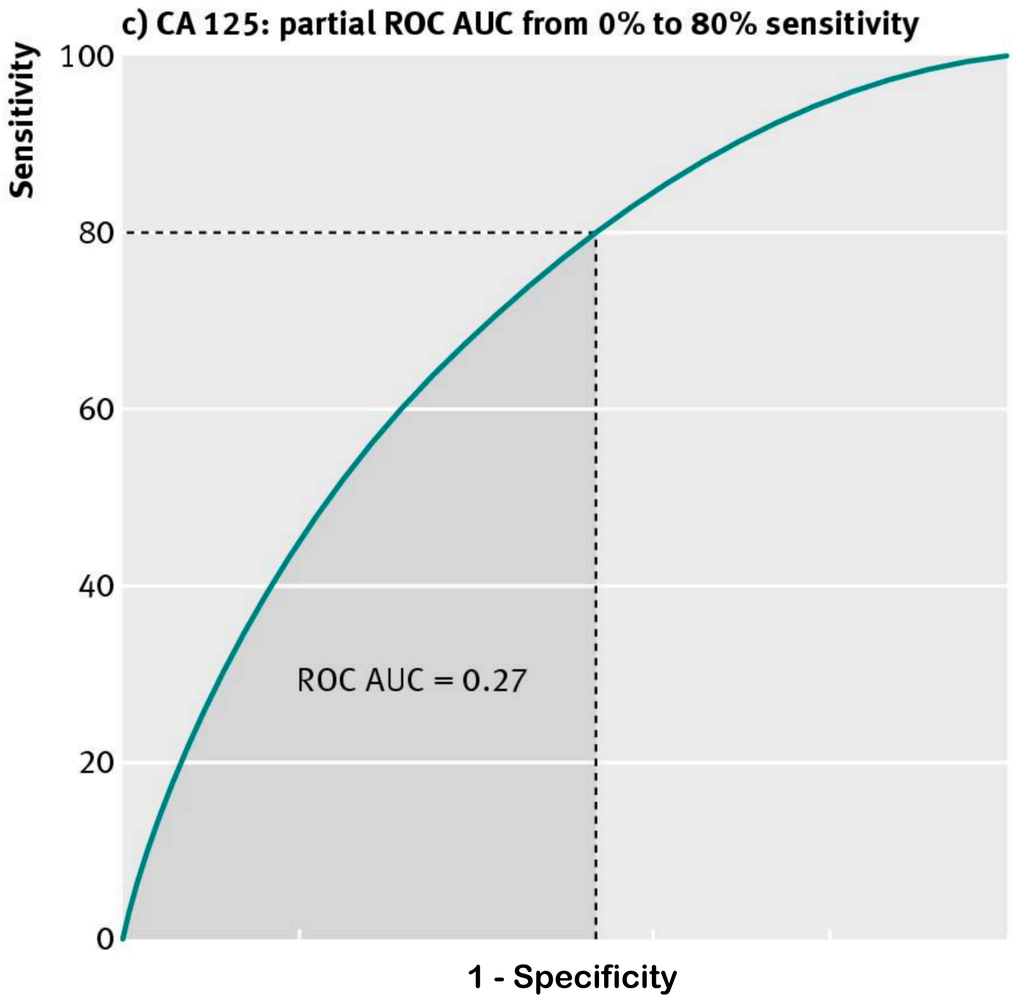
December 18, 2021

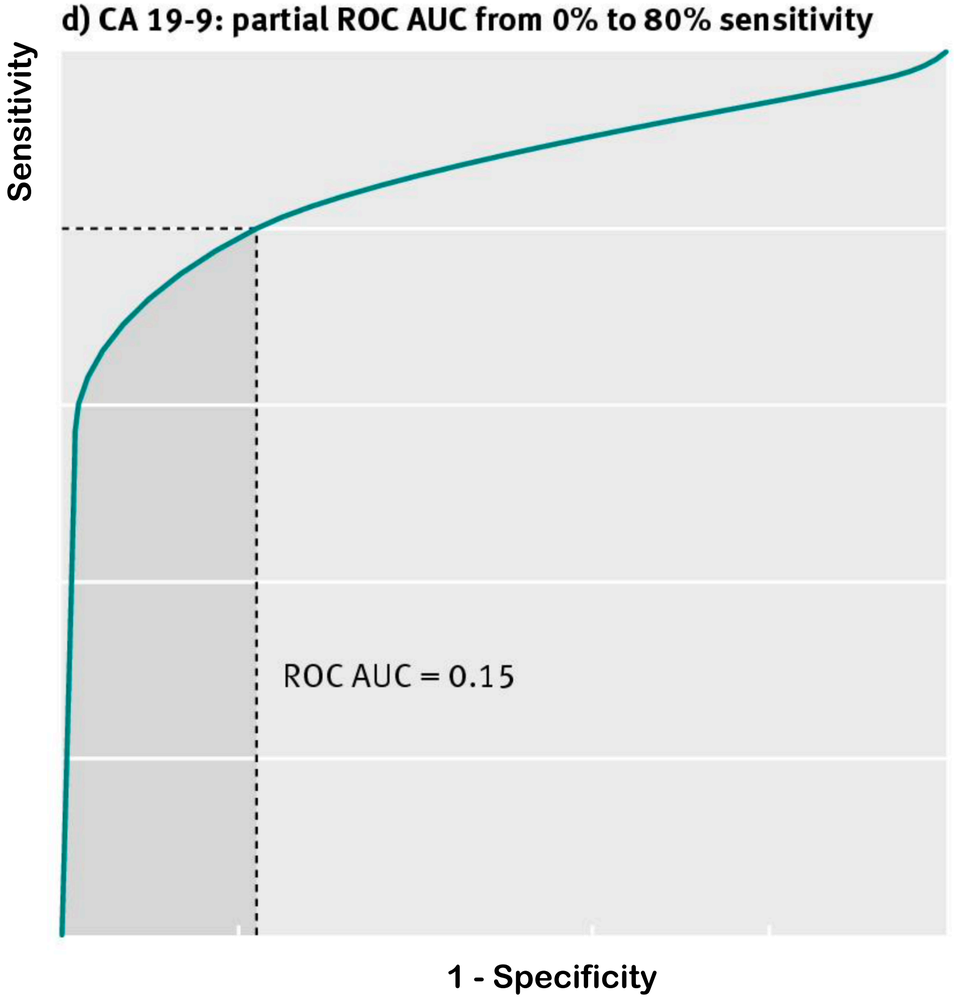
Problems with the Partial AUC[1,2]

- pAUC tries to fix flaws in AUC
- and its formula <u>looks like</u> a generalization of AUC
  - **but it only measures sensitivity* in the partial case![3]**
  - why? the definition for AUC is deceiving. For a whole ROC curve, the values of three different formulas converge[3,4]:
    - average sensitivity
    - average specificity
    - balanced average accuracy

*is average sensitivity when normalized

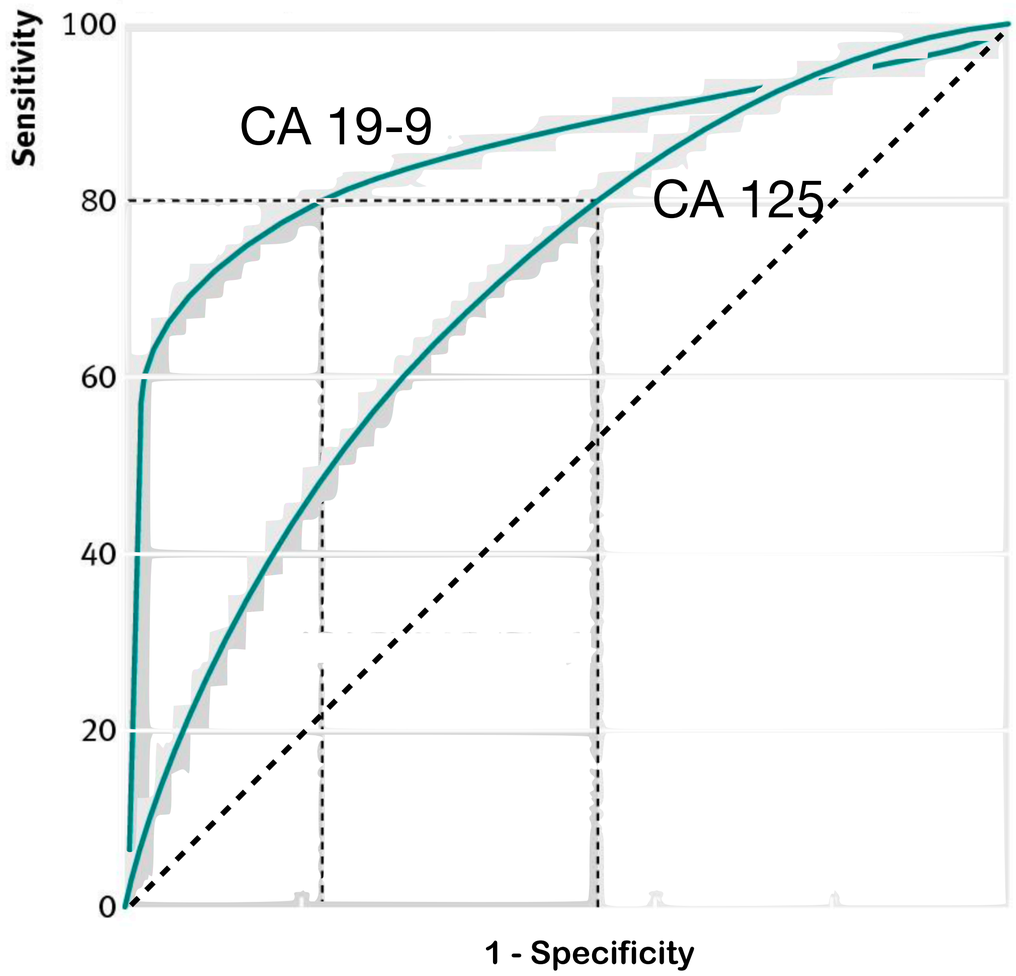
Mallett and Altman *et al.*[5] state: "by excluding sensitivity above 80%, the **partial...AUC** is 0.27 for CA 125 and 0.15 for CA 19-9, **suggesting that CA 125 is the superior test** (see fig 3c and d)"





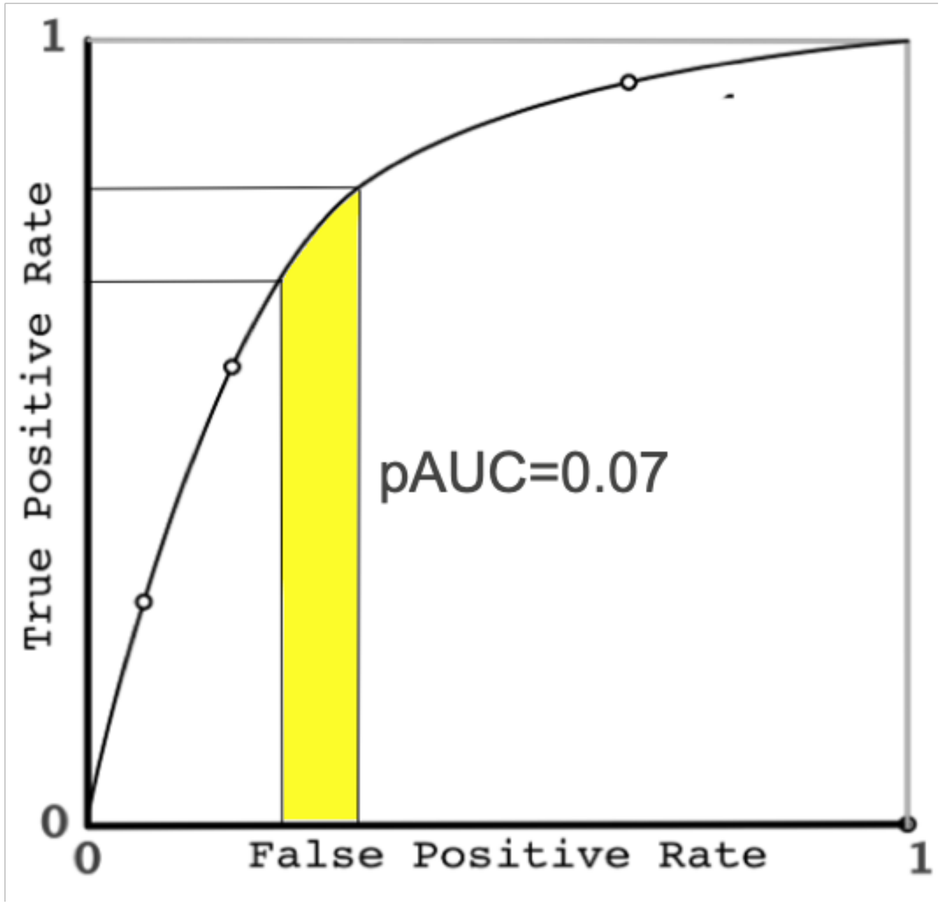
However, the next slide proves the opposite is true.

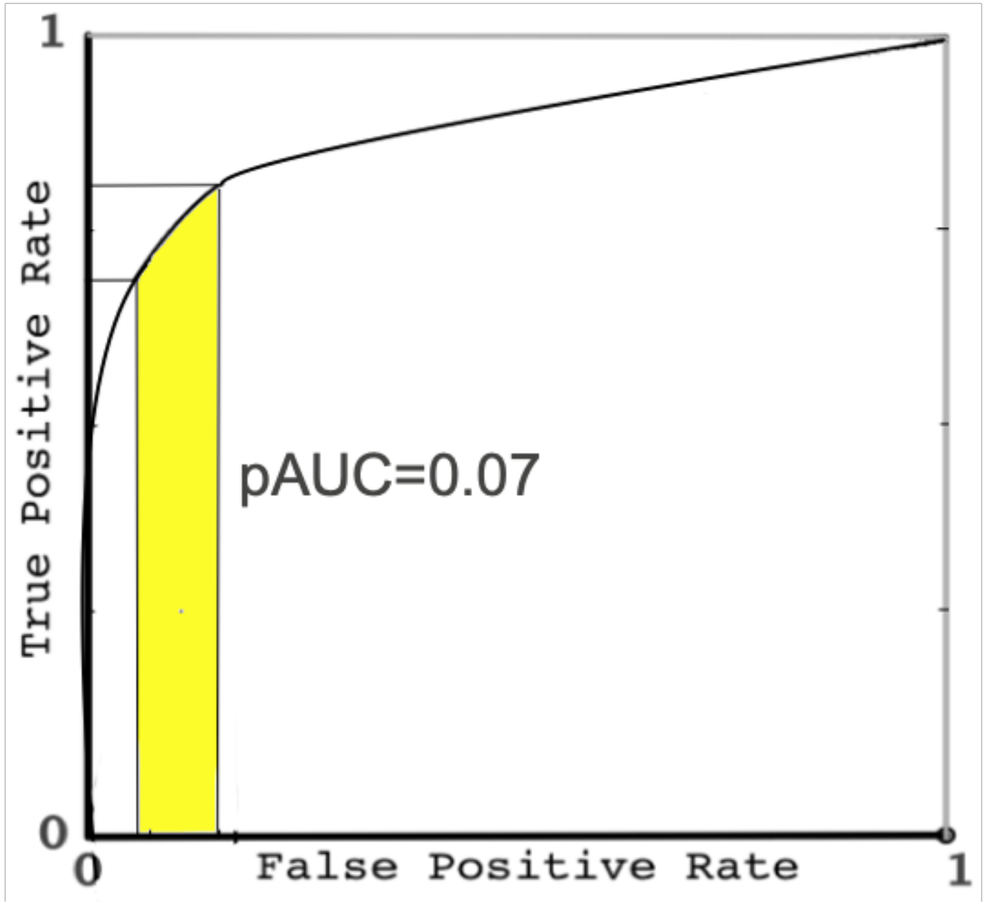
CA 19-9 is superior to CA 125 in performance, **contrary to pAUC**, because all CA 19-9 points are above or to the left of CA 125 (better performance) in the region of interest



Hence, pAUC does not represent AUC in a part. It is a misnomer!

Mallett and Altman *et al.*[5] also state: "A pAUC...produces values of 0.12 for CA 125 and 0.11 for CA 19-9, **suggesting the tests are equally effective** (fig 3e and f)."
We show a similar comparison below. The test at right is superior.





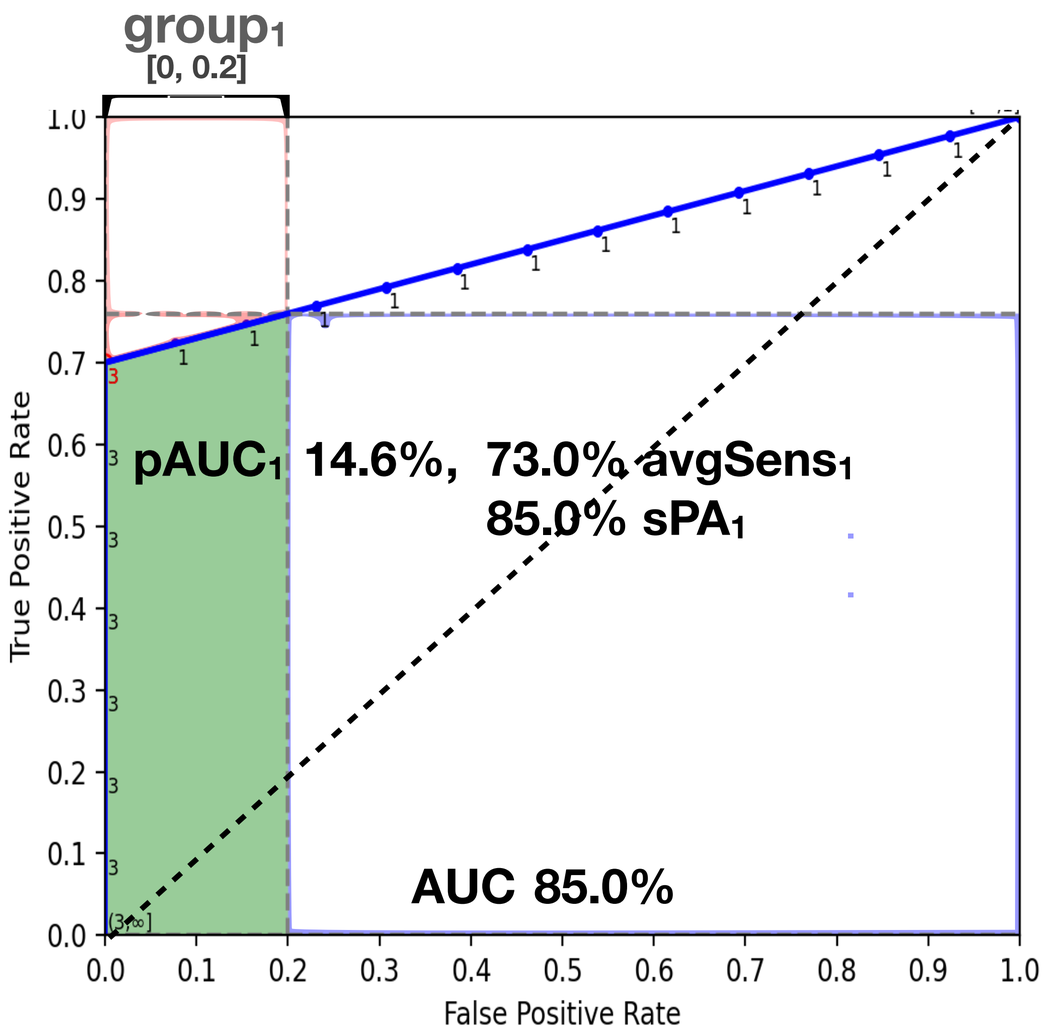
It is false to assert the tests are equally effective: the test at right is superior.

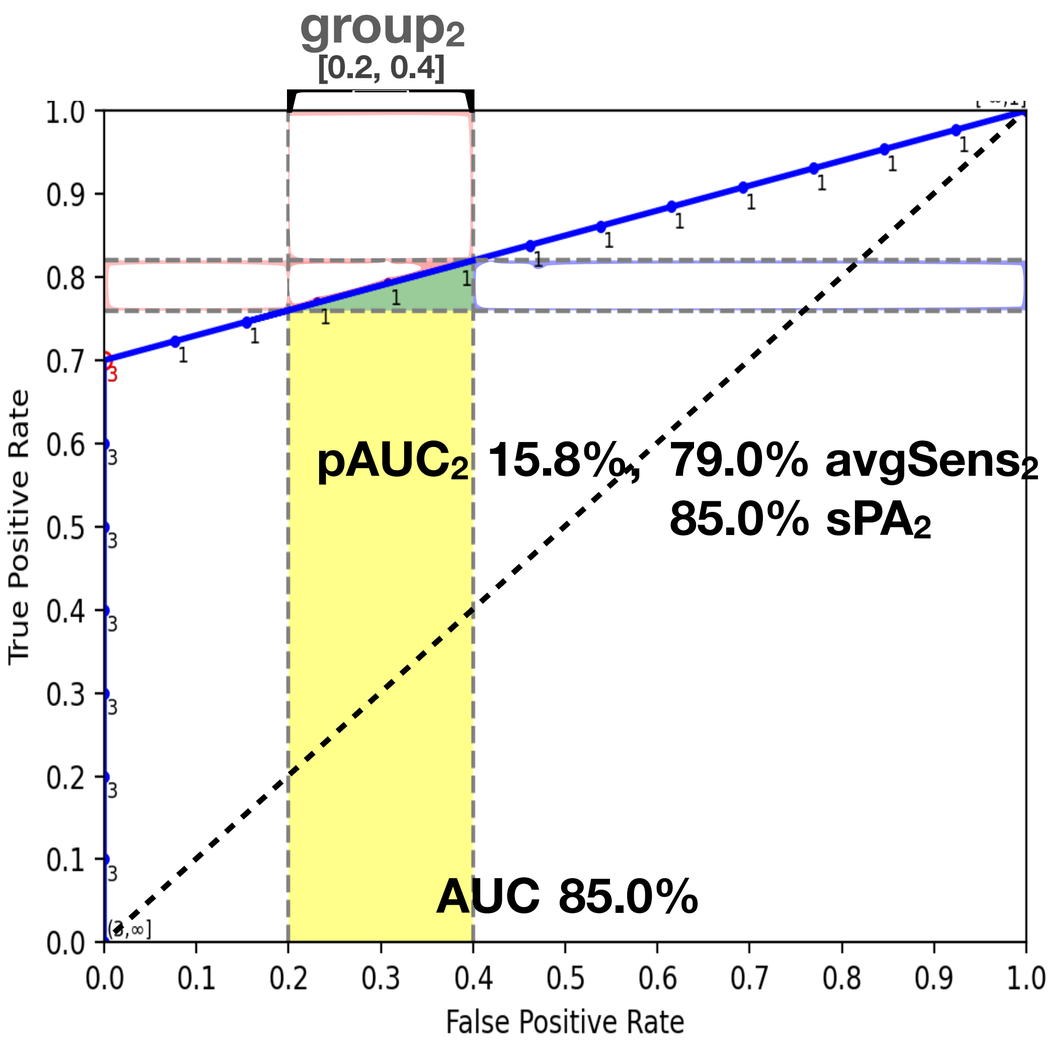
Problems with the Standardized Partial Area[6]

- sPA tries to fix flaws in AUC and partial AUC
- but it is an engineered measure that lacks probabilistic meaning
- it looks good at first, geometrically and empirically, but
  - **we demonstrate two new critical flaws (next)**
  - others also discuss flaws[3,7]

*normalized partial AUC is average sensitivity

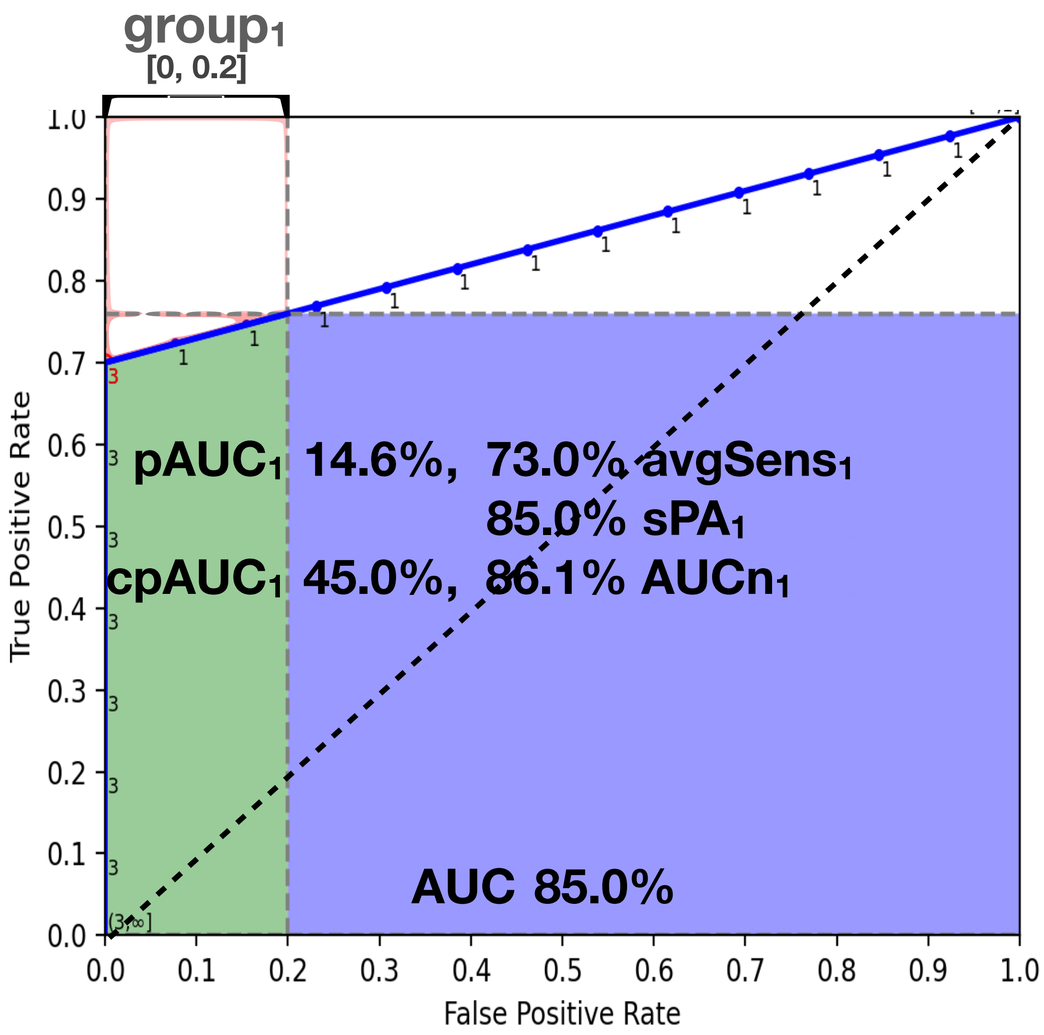
The ROC curve is further from the major diagonal (dashed line) in $group_1$ vs. $group_2$, hence $group_1$ **performs better**, however, $sPA_1 = sPA_2$ and $pAUC_1 < pAUC_2$.

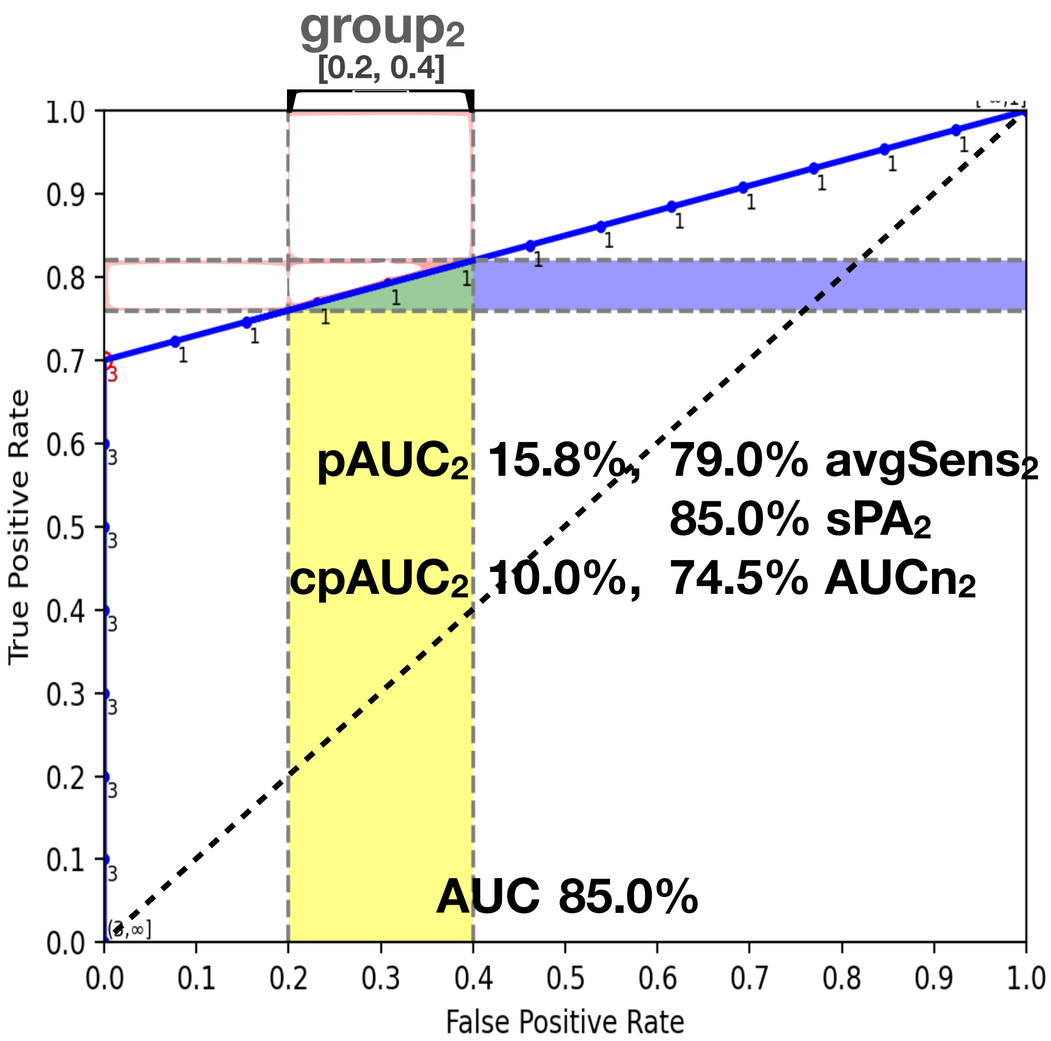




Legend: red = error; {yellow, green} = vertical performance re positives; {blue, green} = horizontal performance re negatives
*note: equal costs of errors are assumed by all of these performance measures

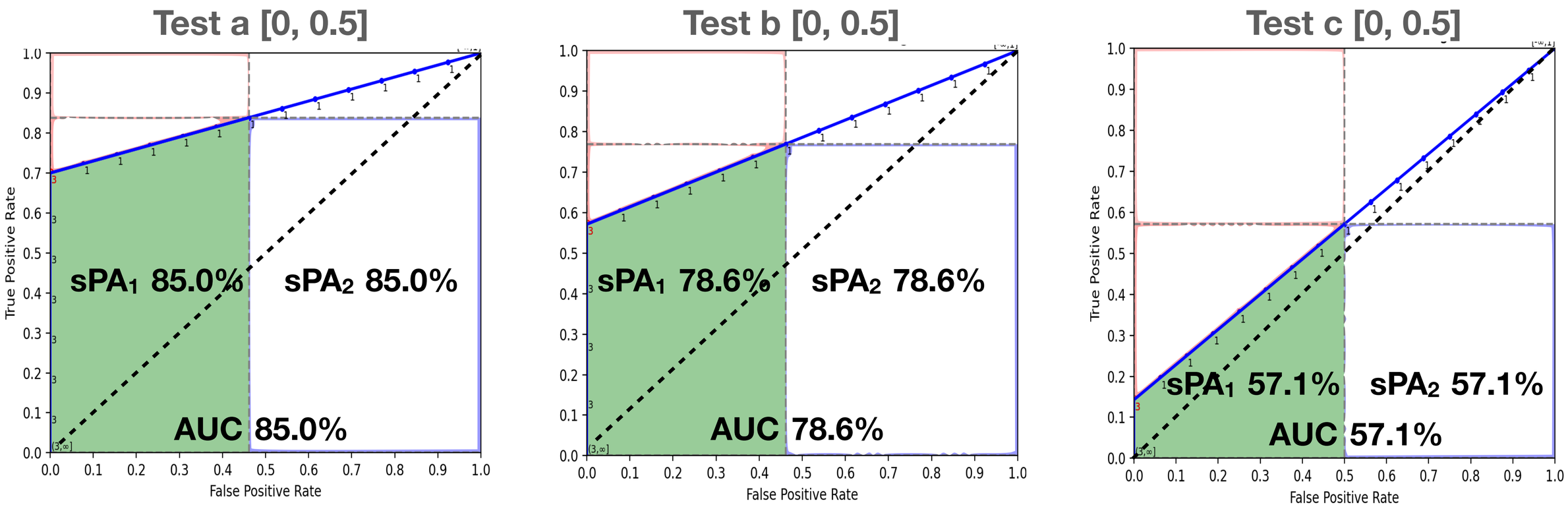
In contrast, the concordant partial AUC, plain[3] (cpAUC) and normalized[3,8] ($AUCn_i$) correctly show that $group_1$ **performs better**: $cpAUC_1 > cpAUC_2$, $AUCn_1 > AUCn_2$



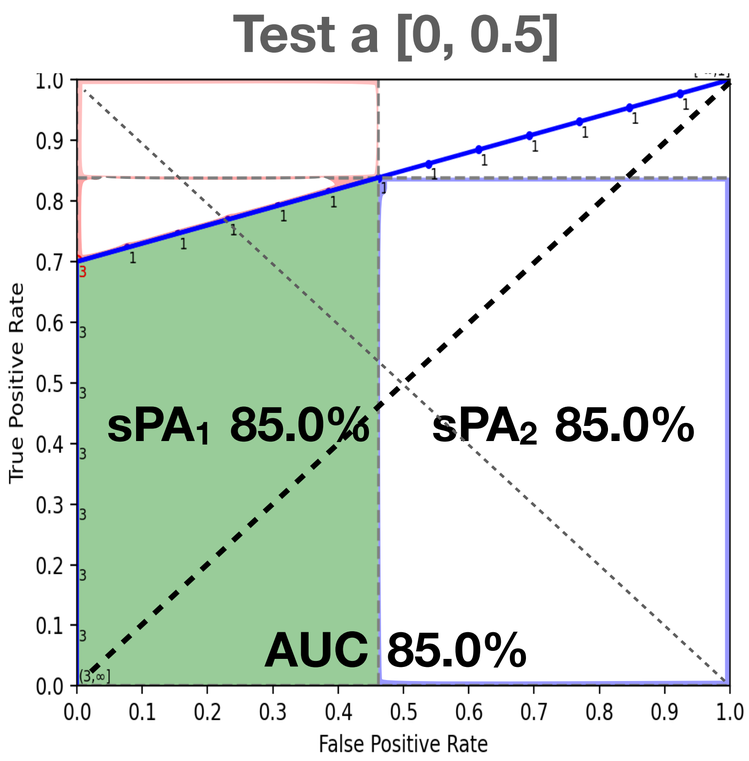


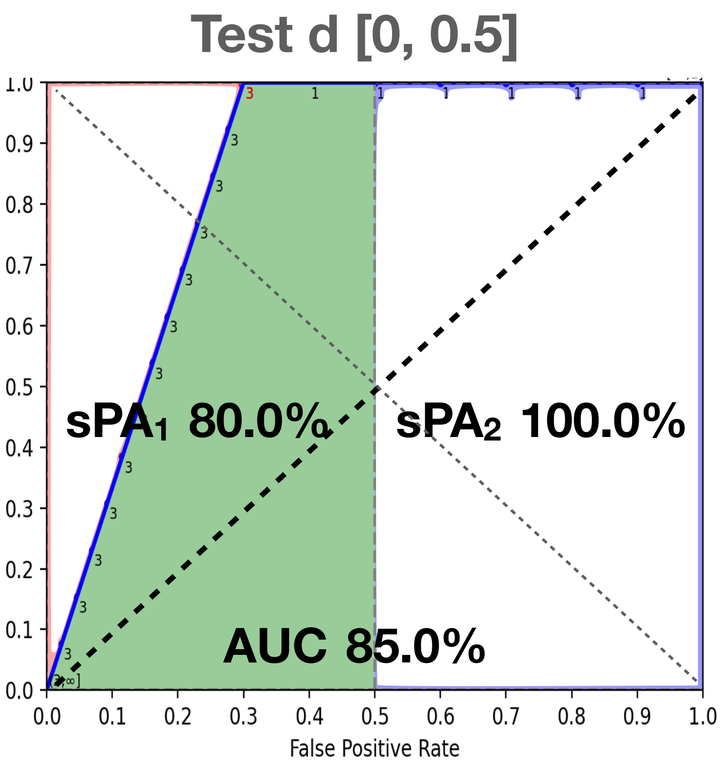
Legend: red = error; {yellow, green} = vertical performance re positives; {blue, green} = horizontal performance re negatives
*note: equal costs of errors are assumed by all of these performance measures

For ROC curves of this shape, any group/column of any width, has the same sPA value within a plot, as any other, and is equal to the AUC. sPA doesn't make sense.



The two ROC curves have the same area above the major diagonal because they are mirror images of each other (through the faint dashed line). sPA should have the same values (mirrored), but it does not.





References

1. McClish DK. Analyzing a portion of the ROC curve. Medical Decision Making. 1989 Aug;9(3):190-5.
2. Thompson ML, Zucchini W. On the statistical analysis of ROC curves. Stat Med. 1989;8:1277–90.
3. **Carrington AM,** Fieguth PW, Qazi H, Holzinger A, Chen HH, Mayr F and Manuel DG. A new concordant partial AUC and partial c statistic for imbalanced data in the evaluation of machine learning algorithms, BMC Medical Informatics and Decision Making 20, 4 (2020) doi:10.1186/s12911-019-1014-6.
4. Zhou XH, McClish DK, Obuchowski NA. Statistical methods in diagnostic medicine. John Wiley & Sons; 2009 Sep 25.
5. Mallett S, Halligan S, Thompson M, Collins GS, Altman DG. Interpreting diagnostic accuracy studies for patient care. Bmj. 2012 Jul 2;345.
6. McClish DK. Evaluation of the accuracy of medical tests in a region around the optimal point. Acad Radiol. 2012;19(12):1484–90. https://doi.org/10.1016/ j.acra.2012.09.004.
7. Vivo JM, Franco M, Vicari D. Rethinking an ROC partial area index for evaluating the classification performance at a high specificity range. Advances in Data Analysis and Classification. 2018 Sep 1;12(3):683-704.
8. **Carrington AM,** Manuel DG, Fieguth PW, Ramsay T, Osmani V, Wernly B, Bennett C, Hawken S, Magwood O, Sheikh Y, McInnes M, Holzinger A. Deep ROC Analysis and AUC as Balanced Average Accuracy for Improved Model Selection, Audit and Explanation (preprint). arXiv:0706.1234v2 [stat.ME].

The End

http://www.deeproc.org

https://github.com/Big-Life-Lab/deepROC

acarrington at toh.ca

André M. Carrington, Douglas G. Manuel, Paul W. Fieguth, Tim Ramsay, Venet Osmani, Bernhard Wernly, Carol Bennett, Steven Hawken, Olivia Magwood, Yusuf Sheikh, Matthew McInnes, Andreas Holzinger