**Search, lesion-classification and area under ROC curve – results from 236 fits using 3 proper ROC methods and implications for optimizing observer performance**

# ABSTRACT

This document presents an overview of the most recent results of three proper ROC-curve fitting methods. It is the subject of Chapter 18 of an upcoming book. It corrects erroneous negative statements regarding PROPROC published in an earlier document. The area under the ROC curve measures the ability of the observer to discriminate diseased from non-diseased cases. Perfect discrimination yields ROC-AUC = 1 while guessing performance yields ROC-AUC = 0.5. Search performance is the ability of the observer to find true lesions while avoiding false lesions. It is measured by a quantity denoted S. Whether a found lesion is marked depends on the reporting threshold. Having found a suspicious lesion, lesion classification performance is the ability to correctly classify it is malignant or benign. It ranges from 0.5 (chance level ability) to 1 (perfect). The primary conclusion of this document is that search performance is the "bottleneck", averaging about 17%, while lesion-classification performance is about 89%. Unless researchers recognize this fact and continue to focus on ROC performance, their efforts will not be productive. A fundamental rule of science is to first focus on the weak-link. There exists a strong inverse correlation between lesion-classification and search performance. Observers tend to compensate for deficiencies in search performance by greater performance in lesion-classification, and vice-versa.

# Three basic concepts

1. The **area under the ROC** curve measures the ability of the observer to **discriminate diseased from non-diseased cases**. Perfect discrimination yields ROC-AUC = 1 while guessing performance yields ROC-AUC = 0.5.
2. **Search performance** is the ability of the observer to find true lesions while avoiding false lesions. It is measured by a quantity denoted **S**. Whether a found lesion is marked depends on the reporting threshold.
3. Having found a suspicious lesion, **lesion classification** performance is the ability to **correctly classify it is malignant or benign**. It is denoted by , ranging from 0.5 (chance level ability) to 1 (perfect).

The primary conclusion of this document is that search performance is the "bottleneck", averaging about 17%, while lesion-classification performance is about 89%. Unless researchers recognize this fact, and continue to focus on ROC performance their efforts will not be productive. A fundamental rule of science is to first focus on the weak-link. The radiological search model is the only existing method of estimating search performance. It is implemented in RJafroc Version 1.0.0 (currently undergoing QC checks prior to upload to CRAN).

# A few Details

* Search performance is low (0.17) compared to lesion-classification performance (0.89). **Search performance is by far the weak link in observer performance**. It is not measurable under any conventional ROC model, i.e., those that generate decision variable samples on every image. It is measurable using RSM fitting model.
* There exists a **strong inverse correlation** between **C** and **S** and weaker positive correlations between **A** vs. **S** and **A** vs. **C**. Observers tend to compensate for deficiencies in search by greater performance in lesion-classification, and vice-versa.
* **All three methods yielded almost identical AUCs. On the average, PROPROC AUC was 1% larger than RSM AUC, while CBM AUC was 1% smaller than RSM AUC**.[[1]](#footnote-1)
* These findings are consistent with each proper-ROC method being a realization of an ideal observer1,2.
* RSM-and CBM-parameters are correlated, consistent with their physical meanings.
* RSM- and CBM-parameters are correlated, consistent with their physical meanings.
* For degenerate datasets[[2]](#footnote-2) PROPROC yields gross overestimates of performance 3. This problem is readily fixed but unfortunately, this important software due to Metz and colleagues is no longer being supported.

Due to its size, the detailed analysis, including plots, will be published separately.

# References

1. Macmillan NA, Creelman CD. *Detection Theory: A User's Guide.* New York: Cambridge University Press; 1991.

2. Barrett HH, Myers K. *Foundations of Image Science.* Hoboken, N.J.: John Wiley and Sons; 2003.

3. Zhai X, Chakraborty DP. A bivariate contaminated binormal model for robust fitting of proper ROC curves to a pair of correlated, possibly degenerate, ROC datasets. *Med. Phys.* 2017;44(3):in press.

1. This is the correction to the main conclusion of the earlier document mentioned above. [↑](#footnote-ref-1)
2. Defined as not having any interior points. [↑](#footnote-ref-2)