6-Part Webinar Series: Research Methodology
Part V: ROC Curves

Questions Asked: 16 May, 2024

1. How about tests that are not quantitative and cannot be moved right or left like diabetes? If that is qualitative and yes/no. In that case, what would be the explanation? Will it be different if there is no threshold to move but just can be told as yes or no (disease or no diseases)?

   AUC is scale-invariant. It measures how well predictions are ranked, rather than their absolute values. AUC is classification-threshold-invariant. It measures the quality of the model's predictions irrespective of what classification threshold is chosen. So for something like diabetes – it would be the scale threshold used to classify the condition. So if you used fasting glucose – the thresholds might be > 140, > 130, > 100, > 90, > 70, > 50, etc.

2. Is it possible to have a 100% positive and 100% false possible? If yes, how so? (slide 21)

   AUC represents the probability that a random positive example is positioned to the right of a random negative example. AUC ranges in value from 0 to 1. A model whose predictions are 100% wrong has an AUC of 0.0 and one whose predictions are 100% correct has an AUC of 1.0.

3. Which one is the most important sensitivity, specificity, FN, or FP in selecting a diagnostic?

   It depends on the condition and the diagnostic test and your concerns about false positives and false negatives.

4. To say true positive or true negative, what would be our reference frame? How do we select this reference frame?

   <u>A fun example might be from An Aesop's Fable: The Boy Who Cried Wolf</u>

   A shepherd boy gets bored tending the town's flock. To have some fun, he cries out, "Wolf!" even though no wolf is in sight. The villagers run to protect the flock, but then get really mad when they realize the boy was playing a joke on them.

   After repeating this action for several nights when there was no wolf seen, the shepherd boy sees a real wolf approaching the flock and calls out, "Wolf!" The villagers refuse to be fooled again and stay in their houses. The hungry wolf turns the kills and eats the flock. The town goes hungry. Panic ensues.

Framework for The Boy Who Cried Wolf

True Positive (TP): • Reality: A wolf threatened. • Shepherd said: "Wolf." • Outcome: Shepherd is a hero.	False Positive (FP): • Reality: No wolf threatened. • Shepherd said: "Wolf." • Outcome: Villagers are angry at shepherd fo up.
False Negative (FN): • Reality: A wolf threatened. • Shepherd said: "No wolf." • Outcome: The wolf ate all the sheep.	True Negative (TN): • Reality: No wolf threatened. • Shepherd said: "No wolf." • Outcome: Everyone is fine.

5. How does having a low incidence of a disease increase the chance of having more false positives? How does this influence the false negatives?

   Incidence/Prevalence = true positives AND false negatives

6. How do we use and apply AUC in logistic regression?

   ROC curves in logistic regression are used for determining the best cutoff value for predicting whether a new observation is a "failure" (0) or a "success"

7. If the test has a high ROC value, does it mean it has both higher specificity and sensitivity?

   AUC is scale-invariant. It measures how well predictions are ranked, rather than their absolute values. AUC is classification-threshold-invariant. It measures the quality of the model's predictions irrespective of what classification threshold is chosen. The area under the curve (AUC) summarizes the entire location of the ROC curve rather than depending on a specific operating point. The AUC is an effective and combined measure of sensitivity and specificity that describes the inherent validity of diagnostic tests

8. Does each research study require an ROC curve? What are the pros and cons of including/excluding the ROC curve?

   Downsides for AUC include: (1) it ignores the predicted probability values and the goodness-of-fit of the model; (2) it summarizes the test performance over regions of the ROC space in which one would rarely operate; (3) it weights omission and commission errors equally. The main feature of AUC lies in its ability to measure the quality of a predictor irrespective of the decision threshold. It summarizes the trade-off between the true positive rate (TPR) and false positive rate (FPR) across all possible thresholds, thus providing a comprehensive metric for model performance.

   The inclusion or exclusion of AUC depends on the study and critical question.

9. Can I use the ROC in determining the predictor of mortality say, 30-day and 90day in-hospital mortality, e.g GCs , NIHSS, and Stroke subtypes?

   Receiver operating characteristic (ROC) curve analysis is a useful method to measure the ability of a clinical risk model to discriminate between hospital deaths and survivors.

10. Why is the ROC curve not reported in case-control studies?

    ROC curves are not typically used in case-control studies. But new methodologies are being developed for such. can in case control studies. Some research involves the analysis of correlated biomarker data. Correlated biomarker data arise from study designs including subjects that contain shared genetic or environmental factors. In a familial-matched design, the information about correlation might help to identify family members at increased risk of disease development and may lead to initiating treatment to slow or stop disease progression. Receiver operating characteristic (ROC) curves have been widely used in medical research to illustrate the performance of a biomarker in correctly distinguishing between diseased and non-diseased groups. Approaches appropriate to a familial-matched case-control design should accommodate inherent correlation in correctly estimating the biomarker's ability to differentiate between groups, as well as handle estimation from a matched case-control design. It is sensible to expect biomarkers to demonstrate improved ability to partition between groups from a paired case-control design. A new developing approach using conditional ROC curves may use ROC curves for correlated paired data. The proposed approach will use the information about correlation among biomarker values, producing conditional ROC curves that evaluate the ability of a biomarker to discriminate between diseased and non-diseased subjects in a familial-paired design. But all is at an early stage of development.

11. If the AUC of two tests are like 0.68 vs 0.71 but significantly different from statistical analysis, how should I consider it for a biological system? How does a unit increase/difference in Auc mean quantitatively?

    AUC basically has no clinical value, so the statical differences should be assessed based on the specific predictors being measured.

    The closer the ROC curve is to the upper left corner of the graph, the higher the accuracy of the test because in the upper left corner, the sensitivity = 1 and the false positive rate = 0 (specificity = 1). The ideal ROC curve thus has an AUC = 1.0.

    ROC values are generally inadequate for any real applications. As a rule of thumb, a ROC AUC score above 0.8 is considered good, while a score above 0.9 is considered great. However, the usefulness of the model depends on the specific problem and use case. There is no standard.

12. What is the difference between ROC curves and Calibration, I used to see that Calibration supports ROC to determine if "that test" is close to the gold standard. But I don't know how calibration works.

    The "model-based" ROC curve (mROC) can be used to assess model calibration and the effect of case mix during external validation. The mROC curve is the ROC curve that should be observed if the prediction model is calibrated in *the external population*. Calibration-in-the-large and the equivalence of mROC and ROC curves are together sufficient conditions for the model to be calibrated.
    mROC can easily be constructed and used to interpret the effect of case mix and calibration on the ROC plot. Given the popularity of ROC curves among applied investigators, this framework can further promote assessment of model calibration. The goal of model calibration is to ensure that the estimated class probabilities are consistent with what would naturally occur. If a model has poor calibration, we might be able to post-process the original predictions to coerce them to have better properties

13. How about 95%CI fit into the ROC and AUC methodologies?

    The 95% Confidence Interval is the interval in which the true (population) Area under the ROC curve lies with 95% confidence. ROC analysis is a powerful tool for assessing the diagnostic performance of index tests, which are tests that are used to diagnose a disease or condition. The AUC value is a summary metric of the ROC curve that reflects the test's ability to distinguish between diseased and nondiseased individuals. AUC values range from 0.5 to 1.0, with a value of 0.5 indicating that the test is no better than chance at distinguishing between diseased and nondiseased individuals. A value of 1.0 indicates perfect discrimination. AUC values above 0.80 are generally considered useful, while values below 0.80 are considered of limited utility. When interpreting AUC values, it is important to consider the 95% confidence interval. The confidence interval reflects the uncertainty around the AUC value. A narrow confidence interval indicates that the AUC value is likely accurate, while a wide confidence interval indicates that the AUC value is less reliable. ROC analysis can also be used to identify the optimal cutoff value for an index test. The optimal cutoff value is the value that maximizes the test's sensitivity and specificity.

14. During the COVID-19 outbreak, particularly in constrained environments, there arose a pressing demand for Antigen Tests (RDTs). While positive results from the RDTs were treated as definitive, negative outcomes necessitated confirmation through Polymerase Chain Reaction (PCR) testing. Can we consider this scenario an instance of prioritizing specificity over sensitivity?

    In the correct general direction. For specificity – there is the issue of false positives and unnecessary treatment. For sensitivity and the concern for false negatives. With COVID – there were some situations where treatment was advised in those individuals with symptoms – regardless of the test results.

2025 Congress related questions:

- What are the topics covered in the congress event?

Preparing, targeting a journal, paper submitting, and manuscript revision.

- What are the conditions for taking part in the 2025 conference?

Simply registering for the program

https://www.whleague.org/about-us/world-hypertension-congress-2025