



SDI Review Form 1.6

Journal Name:	Advances in Research
Manuscript Number:	2014_AIR_13219
Title of the Manuscript:	Investigation of Diagnostic Test Performance Using Receiver Operating Characteristic And Fundamental Concepts Of Information Theory
Type of the Article	Original Research Article

General guideline for Peer Review process:

This journal's peer review policy states that **NO** manuscript should be rejected only on the basis of '**lack of Novelty**', provided the manuscript is scientifically robust and technically sound.

To know the complete guideline for Peer Review process, reviewers are requested to visit this link:

(<http://www.sciencedomain.org/page.php?id=sdi-general-editorial-policy#Peer-Review-Guideline>)



SDI Review Form 1.6

PART 1: Review Comments

	Reviewer's comment	Author's comment (if agreed with reviewer, correct the manuscript and highlight that part in the manuscript. It is mandatory that authors should write his/her feedback here)
Compulsory REVISION comments	<p>The main point of the paper seems to be the claim in the abstract that “Based on these results, it can be verified that mutual information value is parallel to AUC value”. The meaning of “parallel to” is not clear. If “parallel to” means “indicates the same thing as”, then claim is wrong in two respects. First, any general claim concerning the relationship between mutual information and AUC could and should be proven or disproven by a mathematical proof, not by one example as the submitted manuscript offers. Second, I think that investigation of the mathematics will reveal that the claim is simply wrong. AUC measures whether a diagnostic test is accurate, summarized over multiple thresholds. I think mutual information does not measure accuracy. The mutual information when a test is perfectly correct is equal to the mutual information when a test is perfectly wrong. In other words, a perfectly correct test has the same mutual information as a perfectly wrong test. I consider myself an expert on ROC but not on information theory. So if the authors think that I am wrong in my second point, then they should explain why.</p> <p>For the reasons above, I completely disagree with the statement on lines 221-222 that “These results prove that, for the overall quality, neither sensitivity nor specificity but the results of mutual information should be examined.”</p>	<p>* ‘Parallel to’ means that Mutual Information value supports the result of AUC value for this analysis. AUC value and Mutual Information value are not based on the same mathematical background. Their quantitative result is totally different but the way of the interpretation of two tests’ performance is coherent.</p> <p>Mutual Information value is one of the criteria which reduce the uncertainty of the disease. For this reason, it is also the criteria which test is more effective than the other. There is only one disease for this study. Therefore, the uncertainty of the disease is constant ($H(D)$ can be seen in Table 3). However, their mutual information values are different ($I(D;T)$ can also be seen in Table 3).</p> <p>We can change the word of ‘parallel to’ if it is confusing.</p> <p>* Higher value of sensitivity or specificity does not require higher value of mutual Information value for overall quality.</p>



SDI Review Form 1.6

	<p>Line 86 is wrong that the area lies in the interval [0.5,1]. I am completely certain that the area lies in the interval [0,1].</p> <p>The equations fail to communicate clearly with respect to the use of the sigma summations. Authors must include arguments below and above all sigma summation symbols to show the variable of the index and how the index begins and ends.</p> <p>Much of the justification that the manuscript gives is based on the Efficiency (Eff), which the manuscript defines as the overall agreement, meaning true positives plus false positives. This seems to be a horrible criterion for diagnosis. If a disease is very common, then a test that is always positive will usually be correct. If a disease is very rare, then a test that is always negative will usually be correct. However such tests do not have diagnostic power. If a test always shows the same result, then the test has no diagnostic power.</p> <p>I do not buy the logic that “While mutual information can be measured for all threshold values, AUC isn’t measured for all threshold values. Because AUC is a single index value. Therefore, mutual information value has an advantage to AUC value.”</p> <p>The purpose and virtue of AUC is to summarize the accuracy over many threshold values. It is not the goal of AUC to describe all threshold values.</p> <p>The manuscript seems to never examine the results of a Nefelometric test, so I do not understand how they can make any claims that test I is more likely to be similar than test II to the</p>	<p>* We changed it as [0,1]</p> <p>* Notation was changed.</p> <p>* Either Sensitivity or Specificity could have been selected as diagnostic performance. But the best threshold value was determining with respect to the best efficiency value in pROC package of R Programme.</p> <p>* Mutual Information value measures both overall quality of test performance and efficiencies of all thresholds. AUC value is the criteria for overall quality.</p> <p>* Our study is about to examine the performance of two turbidimetric tests. In this study, Nefelometric test was used as gold standard (statement of the disease, not test value)</p> <p>* We converted ‘optimal’ to ‘largest percent correct’</p>
--	--	---



SDI Review Form 1.6

	<p>Nefelometric test.</p> <p>The conclusions depend on the definition of optimal. The manuscript never defines optimal sufficiently clearly. If the authors mean largest percent correct, then they should use the words “largest percent correct”, rather than optimal. The definition of optimal for any application should involve the cost of a false positive relative to the cost of a false negative. The manuscript never discusses this.</p> <p>The conclusions state “it can be deduced that 0-200 UI/ml reference interval which is mentioned in the medicine literature for Nefelometric test can be replaced with a “<i>new</i>” 0-173 UI/ml reference interval.” I think this is an irresponsible recommendation based on the manuscript’s research. A lower threshold will mean more people who have the disease will be diagnosed as not having the disease, thus will probably not be treated. The authors seem to assume the cost in human health of a false positive is equal to a false negative. The authors never consider that the cost of a false positive might be very different than the cost of a false negative. Furthermore, the manuscript is not particularly sophisticated in terms of medicine, so it should not make such bold recommendations concerning public health.</p> <p>I hope this feedback helps.</p>	<p>* As a matter of fact, lower threshold means that the true positive fraction and sensitivity will increase. On the other hand the true negative fraction and specificity will decrease. We disagree with the sentence of ‘A lower threshold will mean more people who have the disease will be diagnosed as not having the disease, thus will probably not be treated.’ Because [0,200] is coded as 0 (healthy) and 200+ is coded as 1 (ill).</p> <p>We totally agree with issue of cost analysis. We couldn’t have obtained the cost value of Table 1 from public hospital.</p> <p>* Thank you for your help.</p>
--	---	--



SDI Review Form 1.6

<p><u>Minor</u> REVISION comments</p>	<p>The manuscript must define ASO.</p> <p>The Introduction attempts to give a history of ROC. Any history of ROC must have the contribution of John Swets, who published many papers in Science concerning ROC.</p> <p>Please change the word in line 163 from “significant” to “practical”. The word “significant” should be used if and only if a p-value is less than the alpha-level for inferential statistics.</p> <p>Change “all” to “several” in line 86.</p> <p>In line 161, replace “disadvantages such as” to “disadvantages to the Nefelometric test such as”</p> <p>I had difficulty understanding Table 2 due to its confusing format. I think the authors could eliminate it, and instead label the threshold points in figure 2 according to the advice of Pontius and Parmentier (2014)</p>	<p>* Explanation of ASO is added.</p> <p>* We changed it as ‘practical’.</p> <p>* Line # 86 is changed.</p> <p>* Line # 161 is changed.</p> <p>* Threshold values in Table 2 are selected pertain to the best sensitivity value, the best specificity value and the best efficiency value. Explanation can also be seen between line #182 and #193.</p>
<p><u>Optional/General</u> comments</p>	<p>REFERENCES available for free from www.clarku.edu/~rpontius Pontius Jr, Robert Gilmore and Benoit Parmentier. 2014. Recommendations for using the Relative Operating Characteristic (ROC). Landscape Ecology 29(3): 367-382.</p> <p>Pontius Jr, Robert Gilmore and Kangping Si. 2014. The Total Operating Characteristics to measure diagnostic ability for multiple thresholds. International Journal of Geographical Information Science 28(3): 570-583. See https://www.youtube.com/watch?v=KKVC3GT5EPw</p>	