Receiver operating characteristic curve and its potential use quantifying cancer therapy outcomes

Rogers MJC*

Department of Otolaryngology, Raigmore Hospital Locum Bank, Highland, United Kingdom

Introduction

After a cancer has been diagnosed, TNM staging is used to classify tumours according to size, spread etc. for prognostication, treatment consideration and discussion and evaluation of outcome. TNM staging is not a diagnostic test as it is not employed in noncancerous cases. PET CT scanning is regarded by many as a gold standard diagnostic test for the presence of cancer. It is not perfect because it carries the risk of serious harm from the radioactivity dose and it is only able to detect tumours greater than 4mm diameter and also generates false positives as it lights up benign tissues e.g. polyps, Warthin's tumour [1]. Nonetheless, in the post-treatment setting, a complete absence of PET CT signal can be interpreted as having a 95% negative predictive value i.e. we are confident that 95% of those who are PET CT negative, truly are negative i.e. cured. This has huge emotional, quality of life and healthy living choice implications that have potential lasting long-term benefits. Cancer treatments are not yet 100% curative and come with consequences and complications.

Because there are a few treatment choices to be considered, weighing up the various pros and cons requires careful consideration, initially by a team discussion at a Multidisciplinary Team meeting and then later with the patient and relatives, friends and supporters. Comparing the relative merits of different treatment options may be likened to comparing apples and pears-with different arguable merits and demerits. This is made more problematic with small prevalence disease. This is due to the poor evidence base: patients are geographically and temporally dispersed, there are difficulties in recruiting to clinical trials and there is an absence of personalised medical biogenetic analysis at last at present. This situation will improve when the ability to assess multidimensional cancer datasets becomes more feasible and fruitful with further increases in computer RAM size or perhaps with the move to quantum computing. So a general method of multidimensional analysis of cancer outcome would be particularly helpful especially when it becomes possible to pool datasets globally.

Method- Using Receiver Operating Characteristic (ROC) Curves to Measure Outcome

If a group of patients, histologically diagnosed with cancer, were TNM staged and PET CT scanned and the results compared, we would expect both measures to indicate they had cancer. Thus there would be no disagreement between the two measures. Although TNM staging is not a diagnostic test, if we pretended it were a diagnostic test, then when comparing it with the PET CT scan, the complete agreement with the PET CT result might allow us to view it as a perfect diagnostic test. But obviously TNM staging is only performed in known cancer cases and it is not used as a diagnostic test in real life. Because there would be perfect agreement with the absolute PET CT diagnosis (although there may be some variation or disagreement between the two tests as to the exact staging of certain, individual patients) the test results, for an idealised perfect diagnostic test would be as follows. All the cases would be true positives, there would be no false negatives, true negatives or false positives. If one were to compare the sensitivity of TNM staging against the gold standard cancer diagnostic test that is PET CT, then the sensitivity (defined as the true positive fraction=true positive/(true positive+false negative)) is $100/(100+0)=1$ [2]. The sensitivity is conventionally mapped as the y-axis parameter of a receiver operating characteristic curve. Specificity is presented on a ROC curve as the false positive fraction (=1−specificity=false positive/(false positive+true negative)) and is $0/(0+0)=0$. The false positive fraction is conventionally mapped as x-axis parameter on a ROC curve. Thus the ROC curve of two tests or measures with perfect concordance will have a ROC curve passing through the upper left corner of the graph (=100% sensitivity and 100% specificity. The area under the ROC curve (AUROC) of such a test is 1.

In contrast, a useless diagnostic test has no discriminating ability and is as informative as flipping a coin to determine whether the diagnosis is present or not. Calculations under these circumstances would be in line with the following: true positive (TP)=25, false positive (FP)=25, true negative

(TN)=25 and false negative (FN)=25. So sensitivity=25/(25+25)=0.5 and false positive fraction=25/(25+25)=0.5. So the midpoint (0.5, 0.5) would be included on the ROC curve. The curve itself would on average conform to a straight line passing from bottom left to top right of the graph. The area under the curve approximates to 0.5.

The situation with a worse than useless diagnostic test is positively misleading as it refutes the diagnosis even when present and declares it present when absent. Under these circumstances: TP=0, FN=50, FP=50, TN=0. So the sensitivity is 0/(0+50)=0 and false positive ratio 50/(50+0)=1. This point (0,1) would lie on the ROC curve at the bottom right of the graph and the area under the graph approximates to 0. If this worse than useless test proved reliable, then it could conceivably be interpreted counter intuitively the diagnosis is the polar opposite to actual test result. But I know of no such tests being employed in current clinical practice.

Using Improvement in Test Scores to Demonstrate Benefit from an Intervention

The value of a teaching course or of online training is sometimes demonstrated by testing beginning students (when considered naive) before the teaching and again using the same question paper after the teaching; the improvement being a measure of the value added by the teaching. A similar approach may be adopted but using different but 100% concordant evaluation measures such as TNM staging (in already histologically confirmed cancer cases) and PET CT scanning to evaluate the presence or absence of treatment benefit. Under these circumstances, null treatment benefit would be demonstrated by a perfect diagnostic test ROC curve i.e. sensitivity=1 and false positive ratio=0. Whereas benefit from treatment would be highlighted by departure from the idealised perfect ROC curve and a loss of area under the curve.

By administering a course of treatment between TNM staging and PET CT scan, some or most of the PET CT scans would become negative because the cancer had been destroyed by the treatment. So the ROC curve comparing such measures would not conform to the idealised perfect diagnostic test ROC curve but instead tend to appear more like the useless or worse than useless ROC curves. The movement (and consequent loss of area under the curve) is a measure of how effective treatment is. The maximal area of AUROC loss is 1, corresponding to a completely ablative cancer treatment.

Discussion

This method may be used to compare cancer diagnostic or therapeutic outcomes by using different taxonomies and appropriate subgroup analysis e.g. different cancers using the same treatment or the same cancers using different treatments or against candidate causative genomic mutations. Taxonomies might be as small and homogeneous or large and heterogeneous as wished, offering the prospect of multidimensional analysis although clearly precision is constrained in smaller sample sizes. Such an approach offers the tantalising attraction of identifying orbits, stabilisers and centralisers and the possibility of treatment with micro-dosing interventional regimens.

Conclusion

The Receiver Operating Characteristic Curve and its potential use quantifying cancer therapy outcomes and hypothetical applicability to multidimensional cancer dataset analysis has been presented and briefly discussed. Further work is needed to demonstrate its actual worth.

References

1. YE Erdi. Limits of tumor detectability in Nuclear Medicine and PET. Mol Imaging Radionucl Ther. 2012; 21: 23-8.
2. S Ekelund. ROC curves- what are they and how are they used. 2011.

*Correspondence to

MJC Rogers

Department of Otolaryngology,

Raigmore Hospital Locum Bank,

Highland, United Kingdom

E-mail id: mjcrogers@yahoo.co.uk