

RECEIVER OPERATING CHARACTERISTICS (ROC). A SPREADSHEET TO ESTABLISH AND COMPARE DIAGNOSTIC PERFORMANCE OF MEASUREMENT RESULTS.

2018-07. Anders Kallner. Dept Clin Chem Karolinska university hospital, Stockholm Sweden.
anders.kallner@ki.se

Background

Signals may be true or false expressions of an event, and *vice versa*, events may emit true or false signals. Although this is taught to have been understood and practiced in the development of the radar, the thinking goes back to that of Bayes, formulated in the famous theorem in the mid-18th century. Four different situations may occur, as illustrated in a 2×2 or contingency table. \if applied to a clinical situation, we have to consider the health status of the patients and the outcome of a test or biomarker. In both ways, this is based on dichotomous reasoning – healthy or not healthy; positive or negative.

	Negative test	Positive test	
Diseased (DIS)	False negative (FN)	True positive (TP)	Sensitivity, TP/DIS
Not diseased (NDIS)	True negative (TN)	False positive (FP)	Specificity, TN/NDIS
	Predictive value PV(-)	Predictive value PV(+)	

The four outcomes are specified in the center of the table. The calculated quantities sensitivity, specificity, and predictive value of a negative and positive outcome, are defined in the marginal cells. The calculated quantities can assume values between 0 and 1. The sensitivity and specificity are also recognized as true positive rate (TPR, sensitivity) and true negative rate (TNR, specificity). The outcomes “negative” and “positive” of a test can be arbitrarily chosen by changing the cut-off, or value, that affects the inference of the test. The sensitivity and specificity are therefore related to the method, disease and cut-off, but not to the prevalence of disease in the population. The predictive values, however, depend on, or are correlated to, the prevalence of disease and can therefore rarely be transferred between sample populations. The essence of the table can be illustrated in a graph (Figure 1).

Short instruction

- There are two identical worksheets in the program, choose one or both!
- The example comprises data related to the patients, i.e. Patient ID, quantity value and diagnosis abbreviated H for healthy and D of diseased (*2018-07 ACB ROC analyses, method evaluation - example data 1.xlsx* and – *example data 2.xlsx*). There is a set of cut-offs or threshold values in both example data sets. Copy the patient-related data and paste it using “Paste special”, “Values (123)”, into columns C, D and E.
- Paste the cutoffs in column H.
- Enter the analyte name in P12. As default high values will be connected with disease but this may be changed in cell R8.
- Then draw the ROC curve and initiate all other calculations by moving the cursor to H1 and simultaneously press CTRL, Shift and “a”.

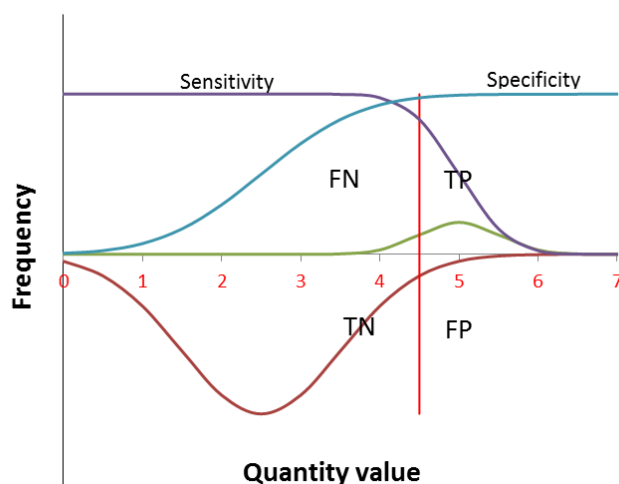


Figure 1. The absolute frequency of diseased are shown above the horizontal line and non-diseased below. The vertical line shows a tentative cut-off and the four outcomes in the respective quadrants. Clearly, the sensitivity will decrease as the cut-off moves towards higher quantity values and the specificity increase, provided high values are characteristic of the diseased. It is unlikely that the diseased population would be normal distributed, it is a simplification in the figure. The graph clearly shows that the sensitivity and specificity cannot both take the maximum value of one (1) at the same cut-off value.

A contingency table can be designed in different ways, i.e. interchanging the place in the table of diseased and non-diseased and the positive and negative outcomes. However, the point in the above design is to ensure that the true negative and positive outcomes are along the diagonal which coincides with how results will be displayed in a scattergram. In case the diseased are characterized by a low value, the true outcomes will still be represented on the diagonal but the “true positives” will be found in the lower left corner of the table and scattergram. The false results then fall along the negative diagonal; the FP represent the type I error and the FN the type II error.

Bayes’ theorem states that the post-test probability is proportional to the pretest probability, with the probabilities conveniently expressed as odds. The proportionality factor is often called the Bayes’ factor, and is equal to the LR(+), which is the ratio between the sensitivity (true positive rate) and (1-specificity) (false positive rate):

$$\text{odds} = \frac{\text{probability}}{1 - \text{probability}}; \text{probability} = \frac{\text{odds}}{1 + \text{odds}};$$

$$LR(+) = \frac{\text{sensitivity}}{1 - \text{specificity}}$$

$$\text{Post - test odds} = LR(+) \times \text{Pre - test odds}$$

A much-used way graphically illustrating the Bayes’ theorem is to plot the *sensitivity* against the *(1-specificity)*, for each conceivable and practically available cut-off, in a scattergram. This is the ROC plot (Figure 2). There are certain important features and characteristics of the ROC plot that need to be defined and pointed out.

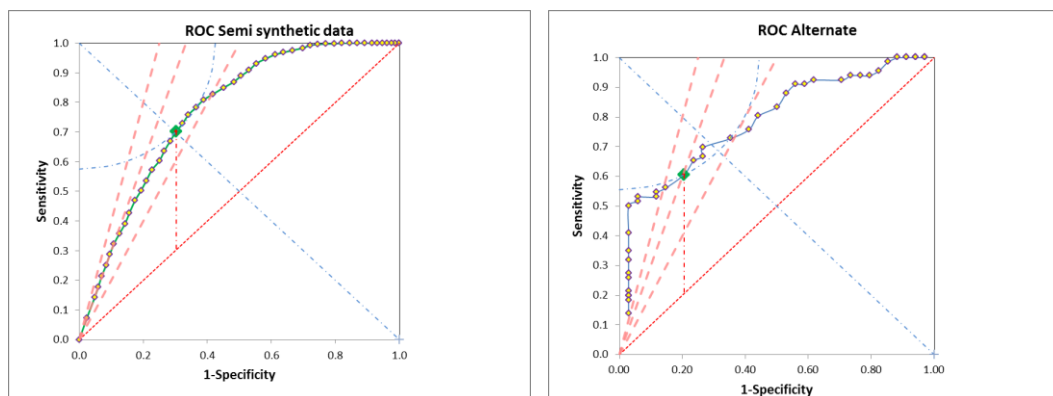


Figure 2. Two examples of ROC-curves. The left panel shows a semisynthetic dataset and the right is from a real study. The dotted red line is the equal line i.e. $LR(+) = 1$ and that perpendicular (blue dotted) the optimal combination provided the curve is symmetrical. The vertical hatched line is the J-index and the quarter of a circle the K-index. NB the periphery of the circle intersects the ROC curve twice. The sloping, hatched lines represent $LR(+) > 1$, i.e. 2, 3 and 4 from right to left. The available ROC space for biomarkers with a $LR(+) > 4$ is the right angle triangle in the left corner of the graph.

Since the maximum values of the sensitivity and specificity are one (1) the best possible outcome is in the upper left corner of the graph (sensitivity = 1 and 1 - specificity = 0; short-hand notation: 0/1). In cases where the sensitivity = 1 - specificity, i.e. $LR(+) = 1$, this can be translated to the “equal line” and thus the post-test odds = pre-test odds. Consequently, only $LR(+) > 1$ have any value for decisions. In practice, a minimum $LR(+) = 3$ is commonly used. Therefore, different “help-lines” corresponding to $LR(+) = 2, 3$ etc may be useful, in addition to the equal line. The distance in the plot between the ROC curve and the equal line is the Youden index (J-index), with larger values indicating better performance. The J-index can be understood and calculated from the plot and amounts to the $sensitivity_i - (1 - specificity_i) = sensitivity_i + specificity_i - 1$, where the index i refers to a particular point (cut-off) on the ROC curve. Another index can be calculated, the K-index, which is distance to a given point on the curve from the 0/1. This represent a quarter of a circle. If the ROC curve is symmetrical and centered on a line perpendicular to the equal line, i.e. $Y = -X + 1$, the J-index and the K-index intersect with this line. It can be regarded as the ultimately best-performing combination.

A ROC-plot lacks some very important information; one cannot immediately see the value of the cut-off in each point on the ROC-curve. This can be remedied by linking a CDA (Cumulative Data Analysis) graph (Figure 3) to the ROC-plot, in which the sensitivity and specificity are separately plotted together against the quantity value. The effect of changing the cut-off is immediately seen and can be chosen to optimize the choice of cut-off for a particular disease or diagnostic situation. Other supporting graphs can be designed e.g. the course of the predictive values at a given prevalence of events in relation to the quantity value.

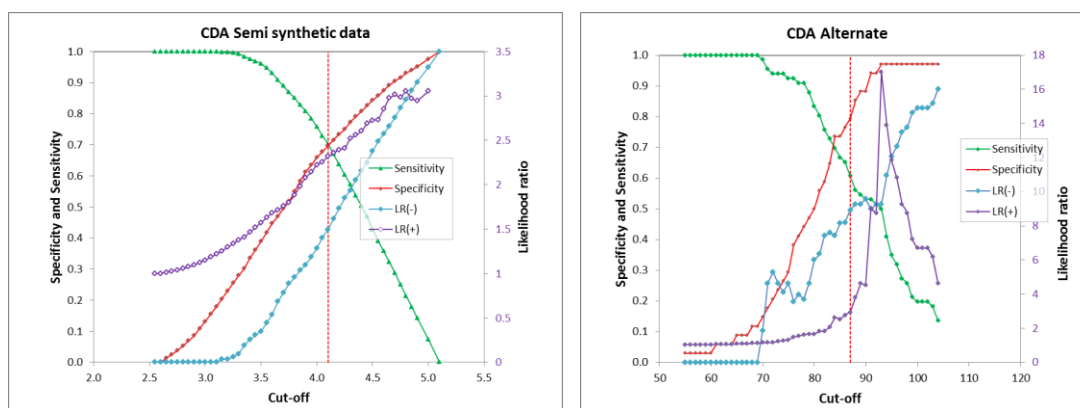


Figure 3. The CDA curves corresponding to the ROC- curves in figure 2. The scales for LR(+) and LR(-) are on the secondary axis, i.e. on the right hand side of the plots.

The diagnostic performance of a biomarker depends on the choice of the cut-off and this can be modified to either give advantage to either the sensitivity or specificity. Consequently, a reference interval or reference value can be adjusted to the diagnostic situation and offer a more dynamic reference than that based on conventional population studies.

The potential diagnostic value can also be quantitatively expressed as the area under the curve (AUC). There are different ways to calculate the AUC, a convenient and sufficiently accurate is to determine the sum of the trapezoids fitted under the curve. The uncertainty of this estimate can be calculated and therefore the performance of several biomarkers can be easily compared. It is often sufficient with a graphical presentation (Figure 4).

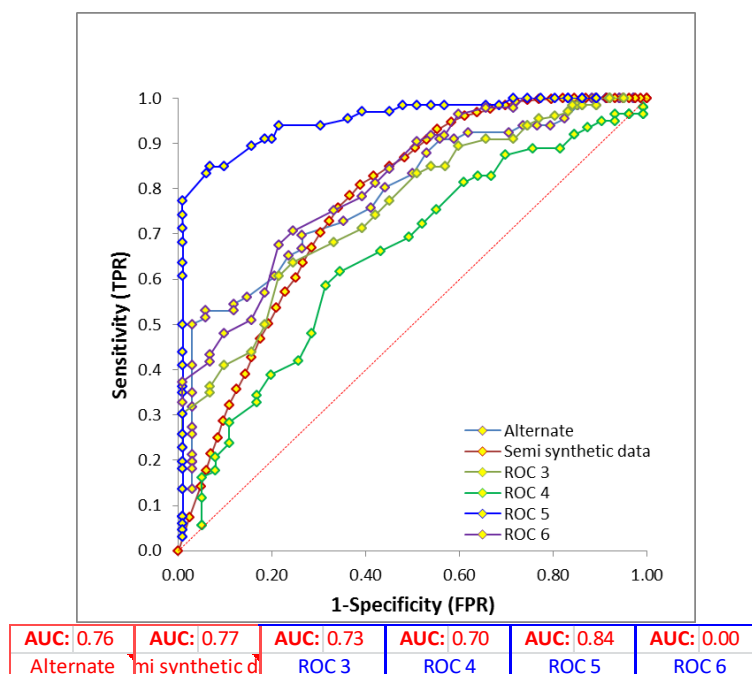


Figure 4. A group of ROC-graphs. TPR: True positive rate =sensitivity, FPR; false positive rate = 1-specificity The calculated AUCs are shown in the table.

ROC Spreadsheet

General

The workbook consists of five sheets, two of which are identical (ROC1 and ROC2) and offer the input of data, which is then calculated and used in the remaining three sheets. Common for all sheets is that the cells which are open for input are blue and text eventually appearing in them is also blue. All calculated results appear in red-lined cells. All sheets are protected to prevent involuntary modifying the formulas. You may not “cut and paste”, that will ruin the program but it is allowed to “copy and paste” and to delete the contents of cells. **NB: columns I and J in the sheets ROC1 and ROC2 are unprotected and should not be tampered with!**

Input and initiating the calculations

There are four input columns in the ROC1 and ROC2 sheets. The first column “C” is for sample ID and is optional; its contents are never used. Since sorting of data is allowed it might be useful to identify each sample to allow an original order to be restored or to find samples which need some correction or consideration. Results of measurements shall be entered in column D and in column E the status (characterization) of the sample (H for healthy and D for diseased). If other conditions than health are studied appropriate changes shall be made but only H and D are allowed as symbols for the characterization.

It is imperative for the entire procedure that the status of the sample (H or D) is determined by an independent method (endpoint). It need not be as drastic as death and autopsy but a relevant reference is necessary to fit the results in the contingency table.

Before calculations will begin the name of the measurand must be entered in cell P12 and it is necessary to indicate if the “diseased” (or otherwise characterized as D) are assumed to be found in the high end of the measuring interval (R8).

As the columns C and D are filled, the program will allocate the results to a contingency table, i.e. characterize each result as any of TP, TN, FP or FN and the corresponding cells L3 to N6 are populated. The prevalence of disease is calculated and reported in cell P3. Although calculated as default, the cell is unprotected and written in blue to allow exploring the effect of changing the prevalence in the inference of the results. If a value is entered manually the formula is replaced but is found in the text-box and can thus easily be restored. The code may need a translation if the Excel is in another language than English.

The maximum and minimum of all observations are given, thus defining the measuring interval. This information is valuable when the cut-off points shall be defined and the ROC curve drawn. The number of possible cut-off points is 50 but it is not necessary to use all and it is not necessary that they form a series with equally sized steps. It is necessary that they are sorted in an ascending order beginning in cell H3.

A strategic start in selecting cut-offs is to divide the measuring interval with 50, enter the minimum value in H3 and the minimum + a fiftieth (1/50) of the interval in the next cell. Then use the standard Excel function to activate cells H3 and H4 and “pull down” their contents by the handle in the lower right corner of the cell H4. The cells H3 to H52 will then be filled with equal increments as between cells H3 and H4. It may be desirable to change the value of the cut-offs in a particular interval and feel free to do that to any extent! Numbers can also be deleted and cells left empty i.e. use less than 50 cut-offs.

To draw the curve the included macro instruction needs to be activated. This is done by placing the cursor in cell H3 and press *control, shift and a*, all together and simultaneously. The curve is then drawn and all the figures are populated. The macro transfers the contents of cell H52 to this cell L6. The macro is safe i.e. it does not contain any toxic, or otherwise dangerous, codes. NB: you may be required to click “Enable content” by Excel, if prompted.

The maximum of the J-index is color-coded (green) in column J and the corresponding cut-off can be read in column H. The minimum K-index is also calculated and presented in cell N11.

Since the K-index is calculated by solving a second order equation There may be two different minimums, only that related to the lowest of the two cut-offs is reported. The J- and K-indexes are displayed in the graph as a vertical line and a quarter of a circle with the center in 0/1, respectively. All interesting and important quantities related to the contingency table are calculate and presented in cells L7 to N14.

The effect of choosing a particular cut-off is demonstrated by entering the chosen value in cell L6. This changes the positions and sizes of the J-index and the K-index.

In addition to the three main plots, which were described above in the Background section, a number of helpful graphs are available (histograms and box-plots) to allow the user to get acquainted with the dataset. Detailed and flexible graphs are presented in a separate sheet.

All functions of the workbook are interactive and immediately updated except the array of cut-offs. This, however, can be updated after any changes by again activating the macro function from cell H3.

Essentially, the ROC analysis is based on evaluating the effect of modifying the value of the cut-off, the choice of cut-offs is not too important.

Multiple comparisons

If two datasets are compared it is useful to enter the data in sheets ROC1 and ROC2 which are identical. A summary of the results is then transferred to the sheet Comparison and both ROC curves displayed. The AUC is calculated and the difference between them is evaluated using Student's t-test.

A regression may sometimes be interesting and is available in this sheet. You will need to adjust the scales of the axes to accommodate the data. This presentation is of limited value except in some special cases.

Finally, if there are more than two ROC curves which shall be compared the sheet Add ROC allows up to six parallel evaluations. The data which are evaluated in ROC1 and ROC2 are automatically transferred to columns B-E of the "Add ROC" sheet and can be further copied to any of the remaining columns. The additional and new data can be entered in the input sheets and again the columns copied. The AUC is calculated and presented in the head of the table.