Methodology of diagnostic tests

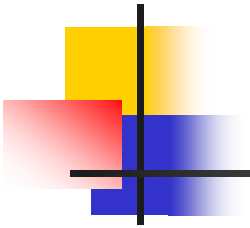
Erik Christensen

Bispebjerg Hospital, University of Copenhagen, Denmark

EASL-Monothematic Conference:

Portal hypertension: advances in knowledge, evaluation and management

Budapest, Hungary, January 24, 2009



Diagnostic tests - background

- A framework used to evaluate the information provided by symptoms, signs and investigational tests of any kind in the best possible way.
- This include sensitivity, specificity, positive and negative predictive values, likelihood ratios and ROC-curves.
- This lecture will review these and provide some suggestions for their extension and improvement.

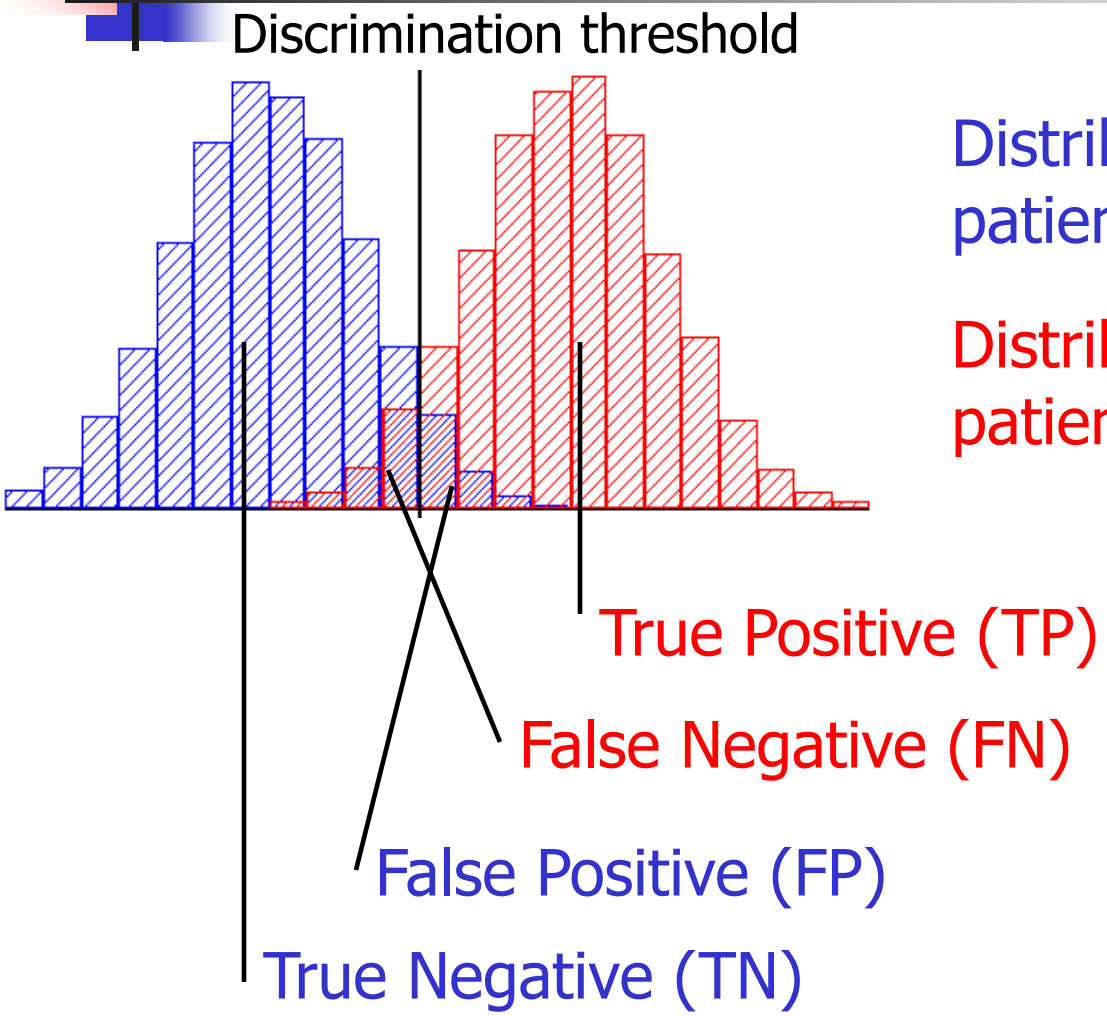
Diagnostic tests - background

- The decision of the doctor in regard to diagnosis and therapy is based on the characteristics (variables) of the patient.
- The doctor needs to know
  - 1) which variables hold the most information
  - 2) how to interpret the information in the best possible way
    - This depends on:
      - the type of the variable
      - the type of decision

Information – Decision Dilemma

- Many variables (e.g. liver function tests, hepatic venous pressure gradient (HVPG)) are continuous
- A doctor's decision cannot be continuous – it has to be binary (i.e. yes or no concerning diagnosis and treatment)
- Therefore – for the simple diagnostic tests, quantitative variables need to be made binary by introducing a threshold or cutoff to distinguish between ”normal” and ”abnormal”

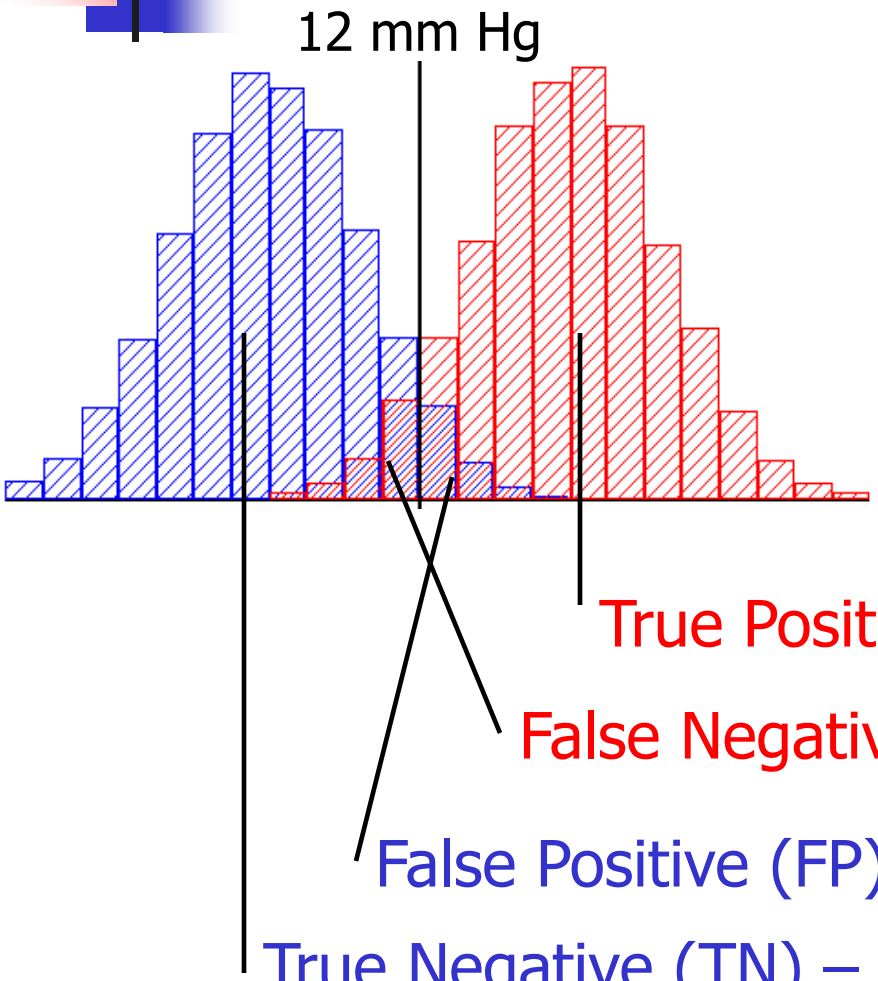
Discrimination between normal and abnormal



Distribution of test-values in patients without condition

Distribution of test-values in patients with condition

Example: Discrimination between bleeding and and non-bleeding by HVPG (hepatic venous pressure gradient)



Distribution of HVPG in patients without variceal bleeding

Distribution of HVPG in patients with variceal bleeding

True Positive (TP) – high HVPG and bleeding

False Negative (FN) – low HVPG and bleeding

False Positive (FP) – high HVPG and no bleeding

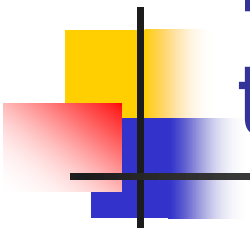
True Negative (TN) – low HVPG and no bleeding

Example: Discrimination between bleeding and non-bleeding by HVPG

Relation between
high (≥ 12 mm Hg) or low (< 12 mm Hg) **HVPG** (Hepatic Venous Pressure Gradient)
and occurrence of **variceal bleeding**

	Bleeding	No bleeding
High HVPG	True Positive (TP)	False Positive (FP)
Low HVPG	False Negative (FN)	True Negative (TN)

Performance of a binary classification test: sensitivity and specificity



- Sensitivity measures the proportion of actual positives, which are correctly identified as such. Also called the true positive rate.
- Specificity measures the proportion of actual negatives which are correctly identified as such. Also called the true negative rate.

Sensitivity and Specificity (example)

	Bleeding	No bleeding
High HVPG	True Positive (TP) = 70	False Positive (FP) = 30
Low HVPG	False Negative (FN) = 6	True Negative (TN) = 194
	True positive rate = **Sensitivity** = TP / (TP+FN) = 70 / (70+6) = 0.92	**True negative rate** = **Specificity** = TN / (FP+TN) = 194 / (30+194) = 0.87

True positive rate = Sensitivity: Probability of high HVPG in patients with bleeding

True negative rate = Specificity: Probability of low HVPG in patients with no bleeding

Sensitivity and specificity

- A sensitivity (true positive rate) of 100% means that the test recognizes all sick people as such. A negative test-result can thus rule out the condition.
- A specificity (true negative rate) of 100% means that the test recognizes all healthy people as healthy. A positive test-result can thus confirm the condition.

False negative rate (β) and False positive rate (α) (example)

	Bleeding	No bleeding
High HVPG	True Positive (TP) = 70	False Positive (FP) = 30
Low HVPG	False Negative (FN) = 6	True Negative (TN) = 194
	True positive rate = Sensitivity = TP / (TP+FN) = 70 / (70+6) = 0.92	True negative rate = Specificity = TN / (FP+TN) = 194 / (30+194) = 0.87
	False negative rate = FN / (TP+FN) = 6 / (70+6) = 0.08	**False positive rate** = FP / (FP+TN) = 30 / (30+194) = 0.13

False negative rate (β) is the probability of bleeding in patients with low HVPG. (Is equal to 1 – sensitivity)

False positive rate (α) is the probability of no bleeding in patients with high HVPG (Is equal to 1 – specificity)

Weaknesses of sensitivity and specificity

- They do not take the prevalence of the condition into consideration
- They give the probabilities of test-outcomes in patients with or without the condition
- The doctor needs the opposite information: The probabilities of the condition in patients with a positive or negative test-outcome

Positive predictive value and Negative predictive value

- **The positive predictive value** is the proportion of patients with positive test results who are correctly diagnosed as having the condition (precision rate or post-test probability of disease).
- **The negative predictive value** is the proportion of patients with negative test results who are correctly diagnosed as not having the condition.
- Both depend on the prevalence of the condition, which may vary.

Positive predictive value and Negative predictive value (example)

	Bleeding	No bleeding	
High HVPG	True Positive (TP) = 70	False Positive (FP) = 30	**Positive predictive value** = TP / (TP+FP) = 70 / (70+30) = 0.70
Low HVPG	False Negative (FN) = 6	True Negative (TN) = 194	**Negative predictive value** = TN / (FN+TN) = 194 / (6+194) = 0.97
	True positive rate = Sensitivity = TP / (TP+FN) = 70 / (70+6) = 0.92	True negative rate = Specificity = TN / (FP+TN) = 194 / (30+194) = 0.87	
	False negative rate = FN / (TP+FN) = 6 / (70+6) = 0.08	False positive rate = FP / (FP+TN) = 30 / (30+194) = 0.13	

Positive predictive value (PV_{pos}) is the probability of bleeding in patients with high HVPG

Negative predictive value (PV_{neg}) is the probability of no bleeding in patients with low HVPG

Influence of prevalence on predictive values

- The predictive values PV_{pos} and PV_{neg} <u>depend</u> on bleeding prevalence

Prevalence of bleeding = 25 %

	Bleeding	No bleeding	
High HVPG	TP = 70	FP = 30	**Positive predictive value** = TP / (TP+FP) = 70 / (70+30) = 0.70
Low HVPG	FN = 6	TN = 194	**Negative predictive value** = TN / (FN+TN) = 194 / (6+194) = 0.97

Prevalence of bleeding = 3.3 %

	Bleeding	No bleeding	
High HVPG	TP = 70	FP = 300	**Positive predictive value** = TP / (TP+FP) = 70 / (70+300) = 0.19
Low HVPG	FN = 6	TN = 1940	**Negative predictive value** = TN / (FN+TN) = 1940 / (6+1940) = 0.997

<u>With decreasing prevalence</u> of variceal bleeding

the **positive predictive value (PV_{pos}) decreases**

and the **negative predictive value (PV_{neg}) increases**

- conversely with increasing prevalence

The likelihood ratio

- The likelihood ratio incorporates both the sensitivity and specificity of the test and provides a direct estimate of how much a test result will change the odds of having the condition.
- **The likelihood ratio for a positive result (L+)** tells you how much the odds of the condition increase when the test is positive.
- **The likelihood ratio for a negative result (L-)** tells you how much the odds of the condition decrease when the test is negative.

Positive likelihood ratio (L+) and Negative likelihood ratio (L-) (example)

	Bleeding	No bleeding	
High HVPG	True Positive (TP) = 70	False Positive (FP) = 30	Positive predictive value = TP / (TP+FP) = 70 / (70+30) = 0.70
Low HVPG	False Negative (FN) = 6	True Negative (TN) = 194	Negative predictive value = TN / (FN+TN) = 194 / (6+194) = 0.97
	True positive rate = Sensitivity = TP / (TP+FN) = 70 / (70+6) = 0.92	True negative rate = Specificity = TN / (FP+TN) = 194 / (30+194) = 0.87	**Positive likelihood ratio** = TP-rate / FP-rate = 0.92 / 0.13 = 6.9
	False negative rate = FN / (TP+FN) = 6 / (70+6) = 0.08	False positive rate = FP / (FP+TN) = 30 / (30+194) = 0.13	**Negative likelihood ratio** = FN-rate / TN-rate = 0.08 / 0.87 = 0.09

Positive likelihood ratio (L+) is the probability of high HVPG among bleeders divided by the probability of high HVPG among non bleeders.

Negative likelihood ratio (L-) is the probability of low HVPG among bleeders divided by the probability of low HVPG among non bleeders.

Advantage of likelihood ratios

- Do not vary in different populations or settings because they are based on ratio of rates
- Can be used directly at the individual level
- Allow the clinician to quantitate the probability of bleeding for any individual patient
- The interpretation is intuitive:
  - The larger the L+, the greater the likelihood of bleeding
  - The smaller the L-, the lesser the likelihood of bleeding

Calculation of post-test probability from likelihood ratio using Bayes' theorem

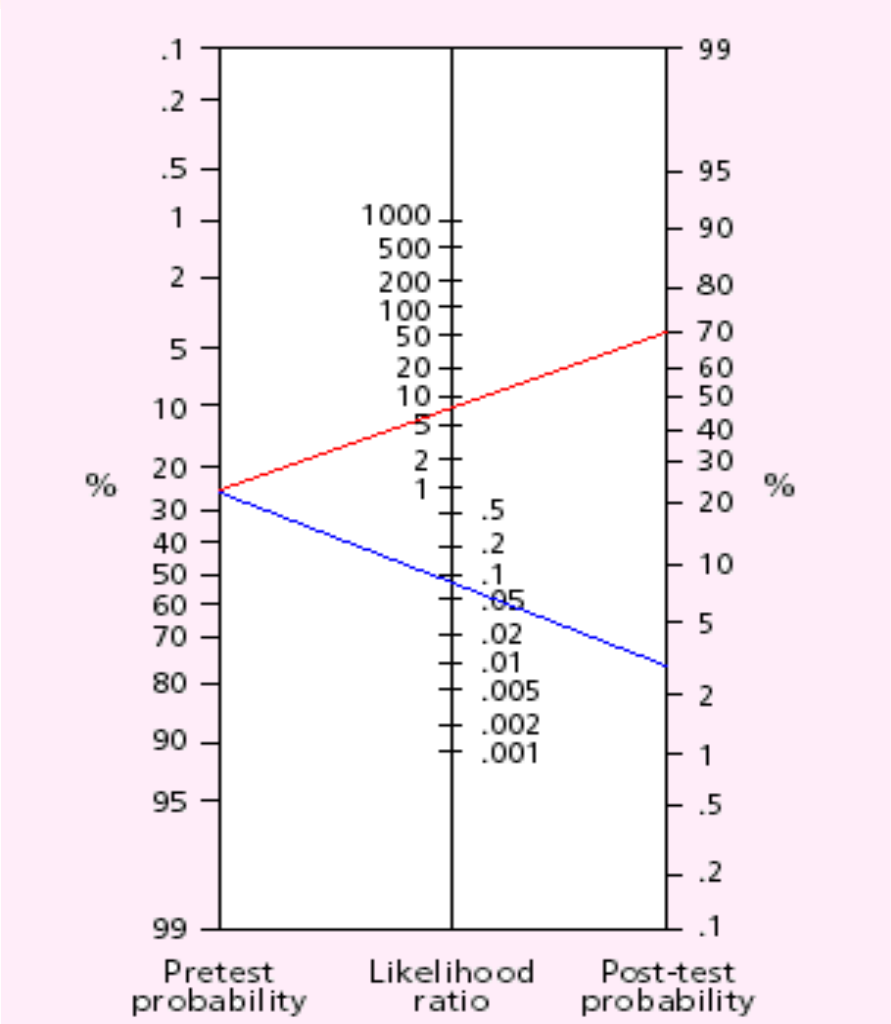
Bayes' theorem:
Post-test odds = pretest odds × likelihood ratio

Example:
Pretest probability = $p_1 = 0.25$
Pretest odds = $p_1/(1-p_1) = 0.25/0.75 = 0.34$

Probability of bleeding with high HVPG (L+ = 6.9):
Post-test odds = pretest odds × L+
Post-test odds = $o_2 = 0.34 \times 6.9 = 2.34$
Post-test probability = $o_2/(1+ o_2) = 2.34/3.34 = 0.70$ $(= PV_{pos})$

Probability of no bleeding with high HVPG (L- = 0.09):
Post-test odds = pretest odds × L-
Post-test odds = $o_2 = 0.34 \times 0.09 = 0.03$
Post-test probability = $o_2/(1+ o_2) = 0.03/1.03 = 0.03$ $(= 1\text{-}PV_{neg})$

Use of nomogram for easy calculation of post-test probabilities from likelihood ratio



Pretest probability (prevalence of bleeding) = 25%

Probability of bleeding with high HVPG (L+ = 6.9):

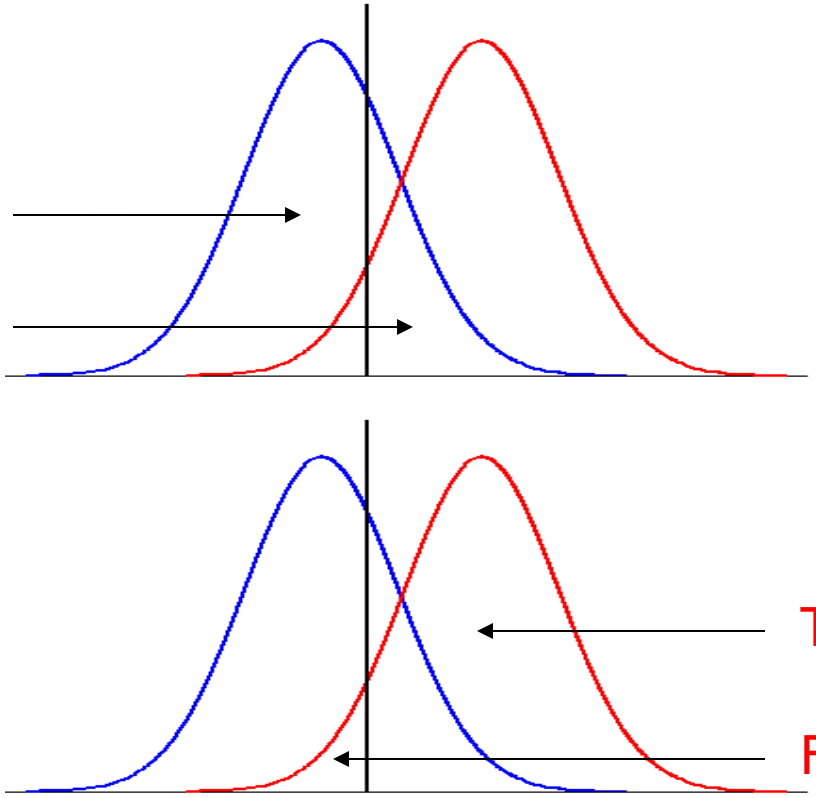
Post-test probability = 70%

Probability of no bleeding with high HVPG (L- = 0.09):

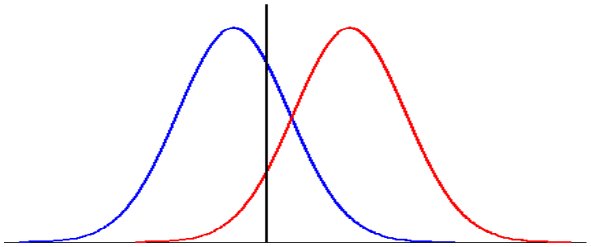
Post-test probability = 3%

BMJ. 2004;329:168–169.

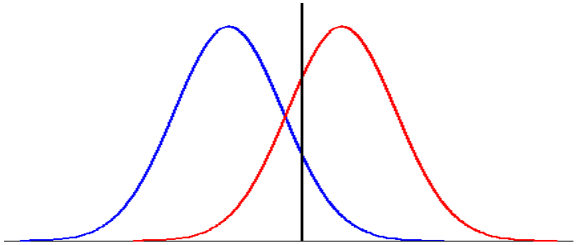
Role of discrimination threshold



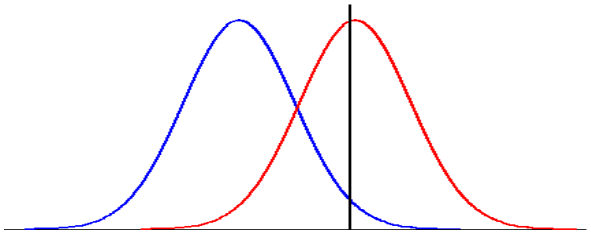
Effect of shifting the discrimination threshold



Sensitivity = True positive rate = 0.91
Specificity = True negative rate = 0.75



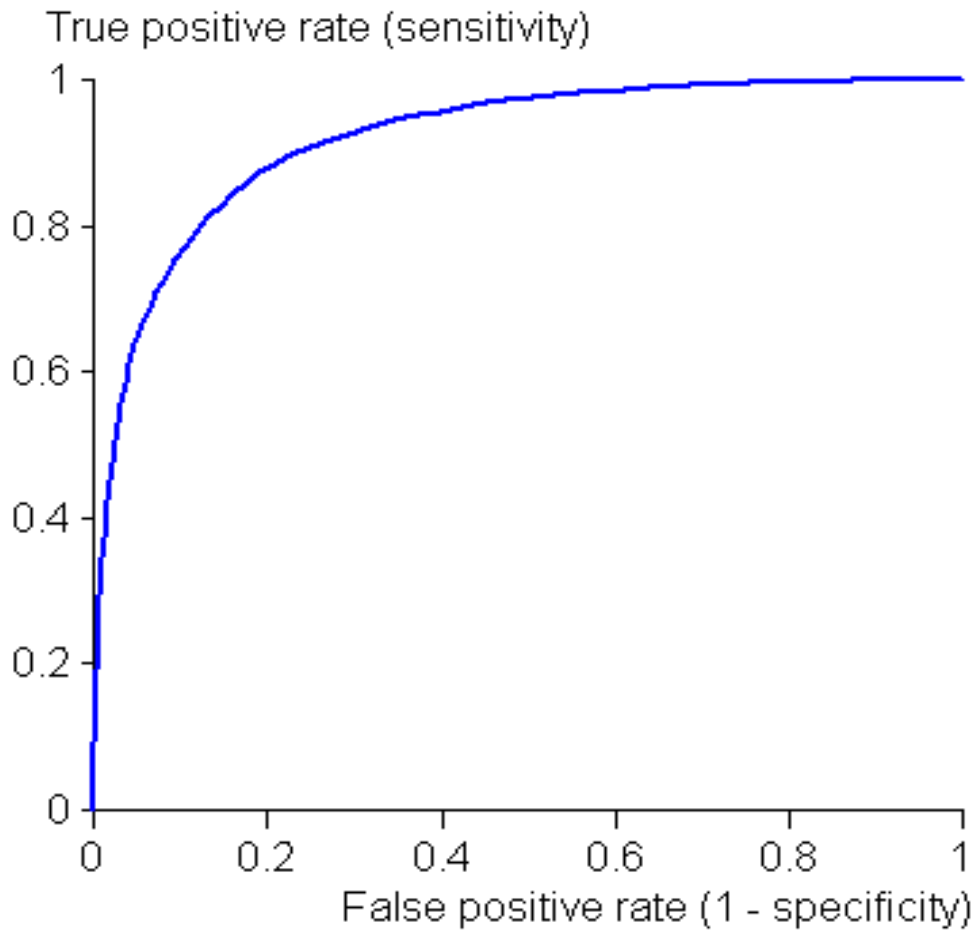
Sensitivity = True positive rate = 0.73
Specificity = True negative rate = 0.89



Sensitivity = True positive rate = 0.51
Specificity = True negative rate = 0.97

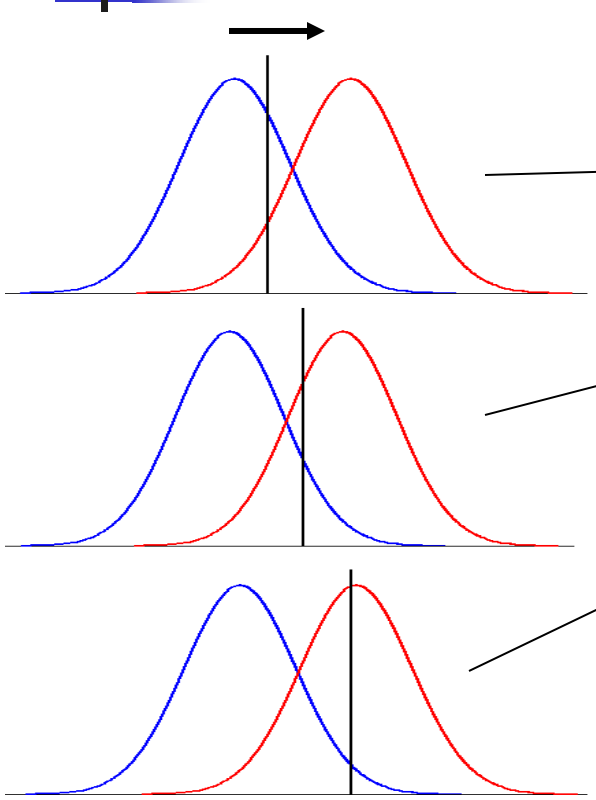
With increasing discrimination threshold
the true positive rate (sensitivity) decreases,
whereas the true negative rate (specificity) increases.

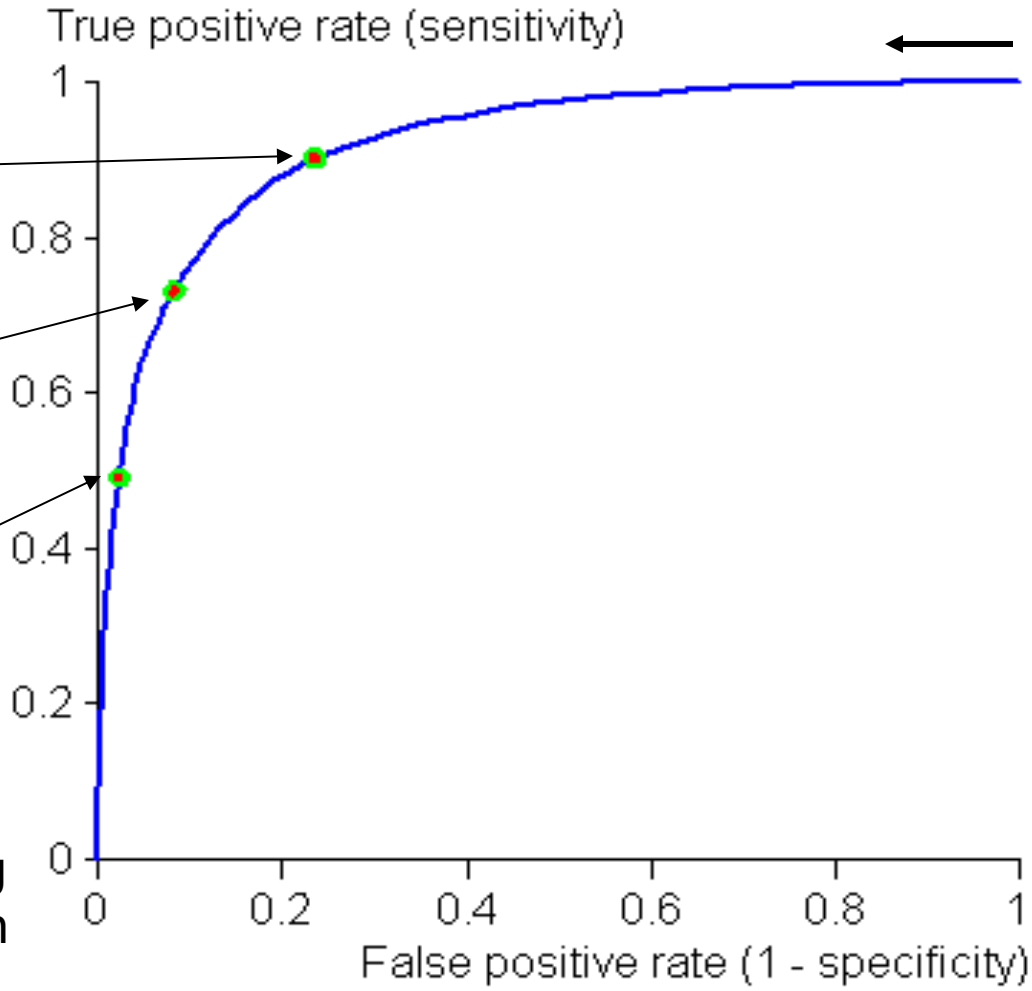
Receiver operating characteristic (ROC) curve



- A graphical plot of the true positive rate (TPR = sensitivity) versus the false positive rate (FPR = 1 - specificity) for a binary classifier system as its discrimination threshold is varied through the whole range of variation.

Receiver operating characteristic (ROC) curve

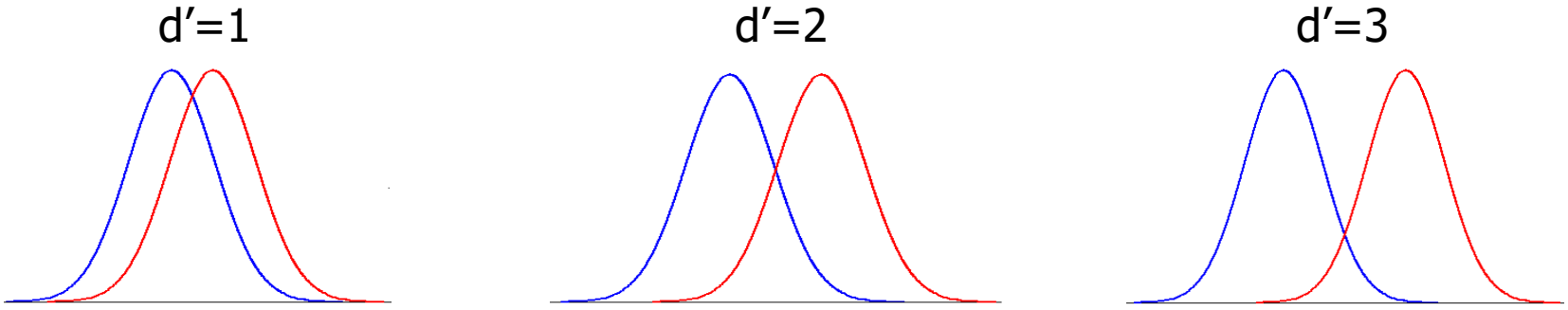




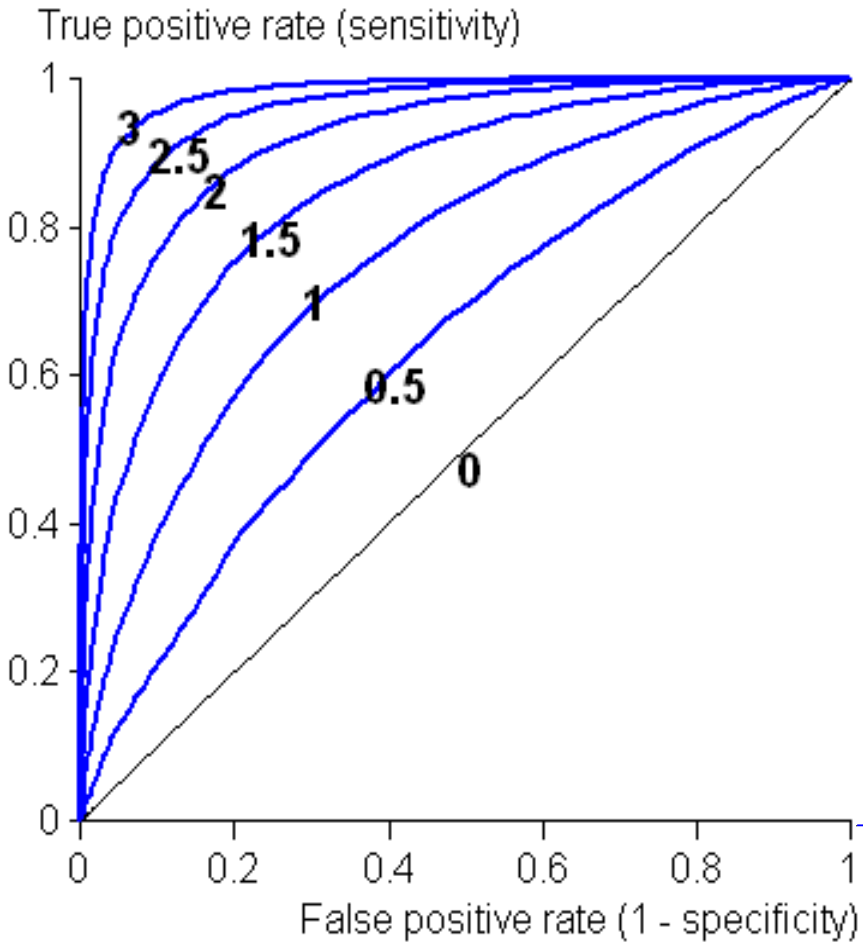
The points are calculated by moving the discrimination threshold through the whole range.

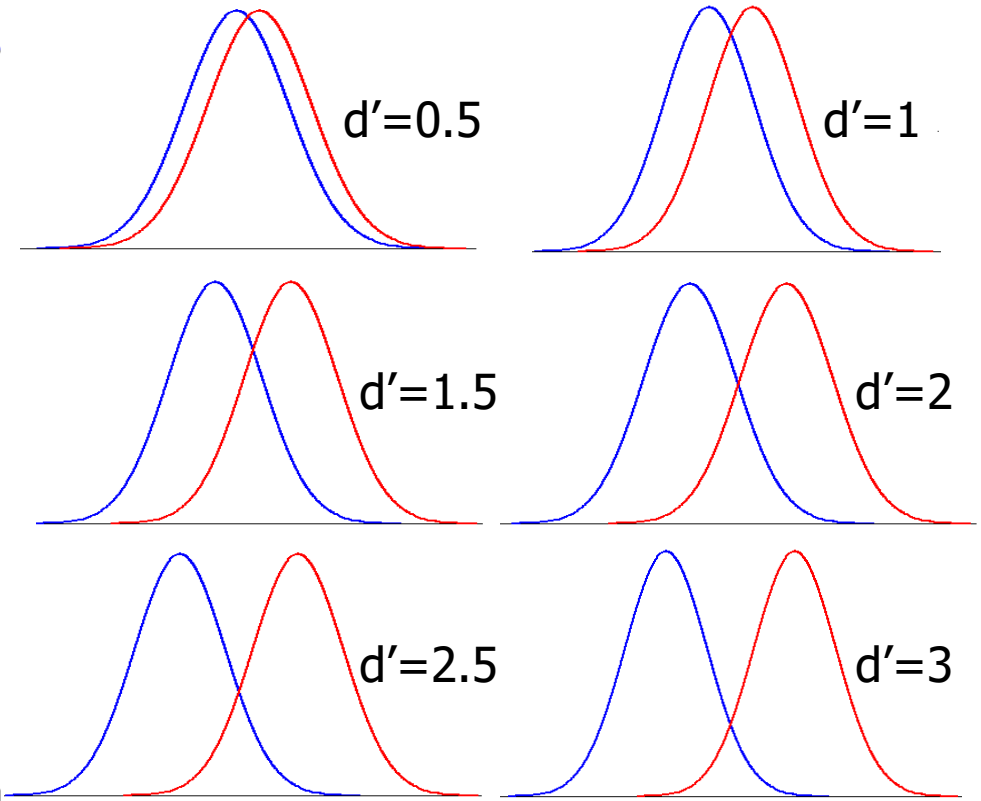
The separtion between distributions is given by the Discriminability index d'

- d' is the difference in means of the two distributions divided by their standard deviation

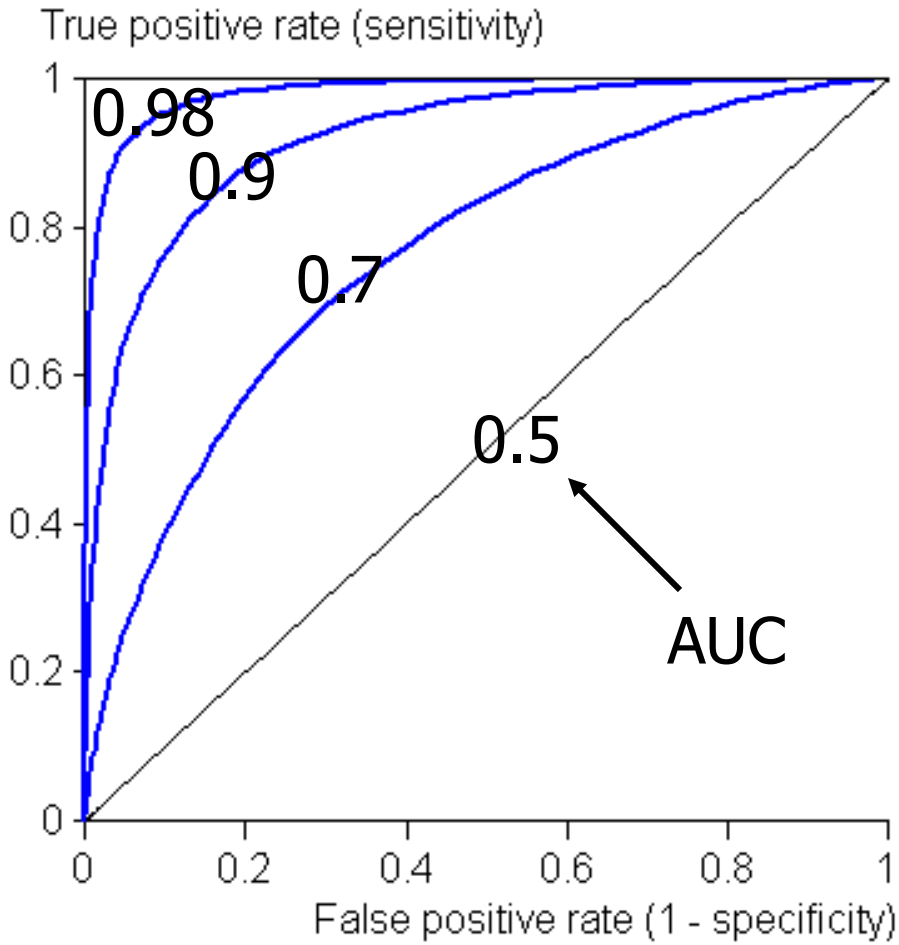


ROC-curves with different values of discriminability index d'





ROC-curves with different Areas Under the Curve (**AUC**) or c-statistic

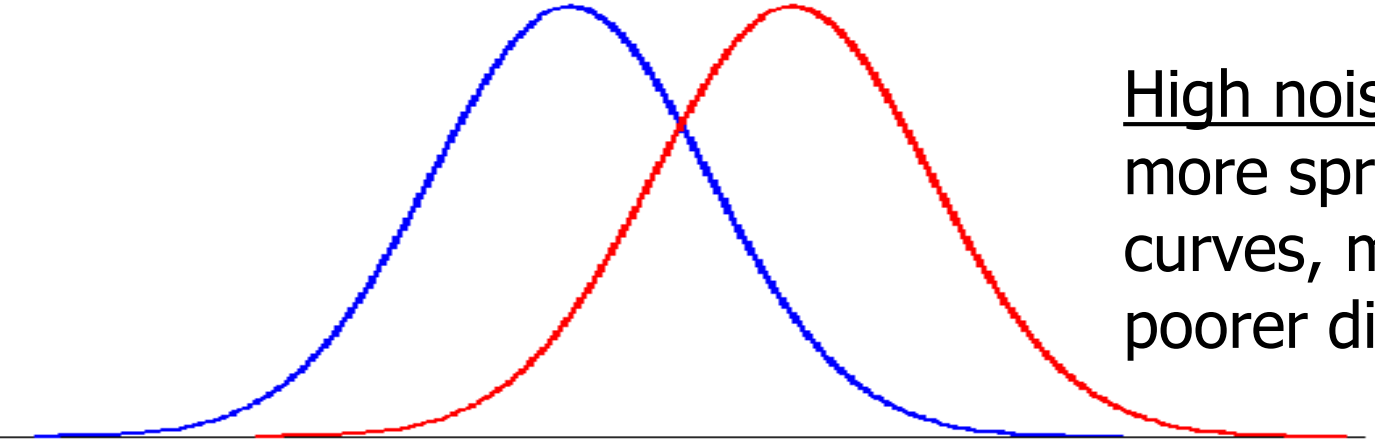


The better the discrimination, the larger the AUC or c-statistic:

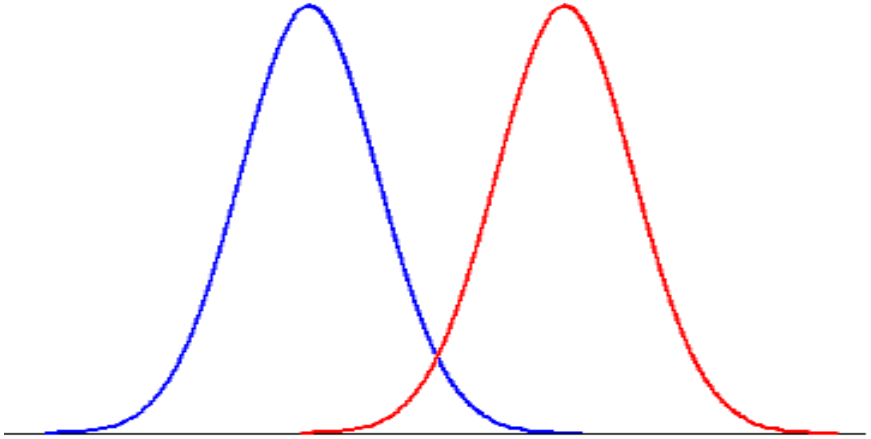
AUC = 0.5 $\Rightarrow$ no discrimination
AUC = 1 $\Rightarrow$ perfect discrimination

Different ROC-curves derived from the same cases can be compared statistically.
(Hanley JA et al. Radiology 1983;148:839-43)

The influence of **noise** in test-values



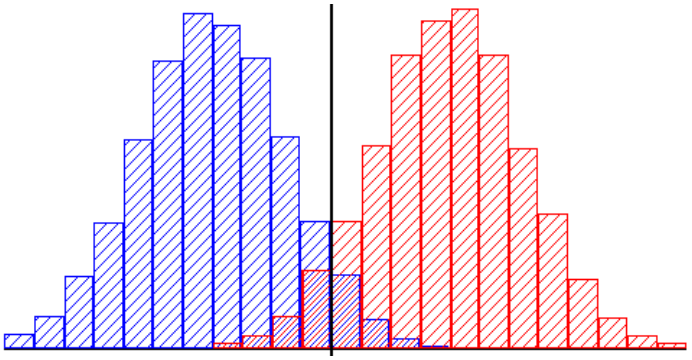
High noise ⇒
more spread, wider curves, more overlap, poorer discrimination



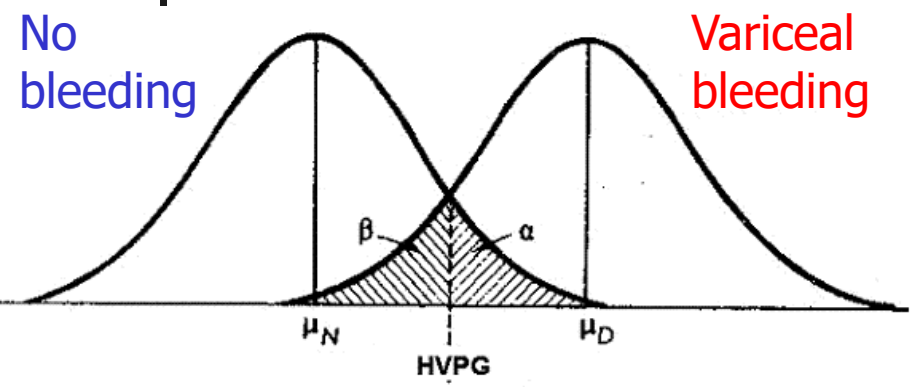
Low noise ⇒
less spread, narower curves, less overlap, better discrimination

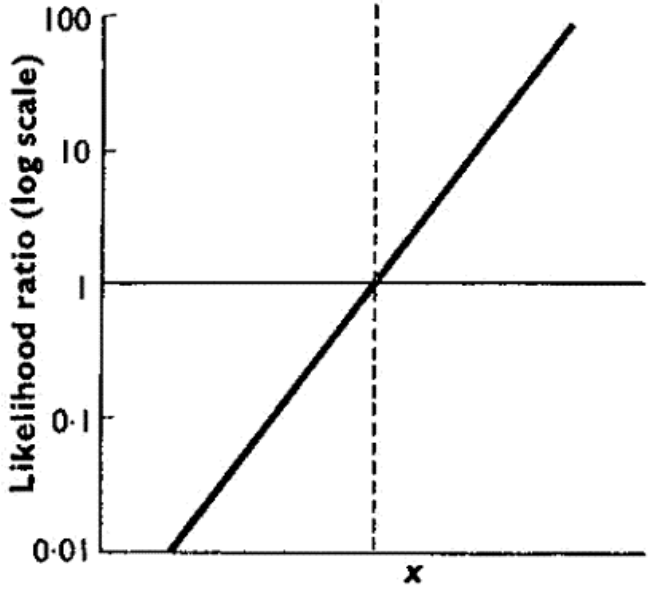
Weaknesses of dichotomization

- The quantitative information within each group (normal, abnormal) is not utilized
- All test-values smaller than the cutoff are considered equal
- All test-values larger than the cutoff are also considered equal
- By disregarding the actual value of the test-variable within each of the groups (normal, abnormal), information is lost



From all-or-none classification to strength of evidence based on the quantitative test value without dichotomization





There is a relation between the risk of bleeding and the actual level of HVPG irrespective a defined threshold.

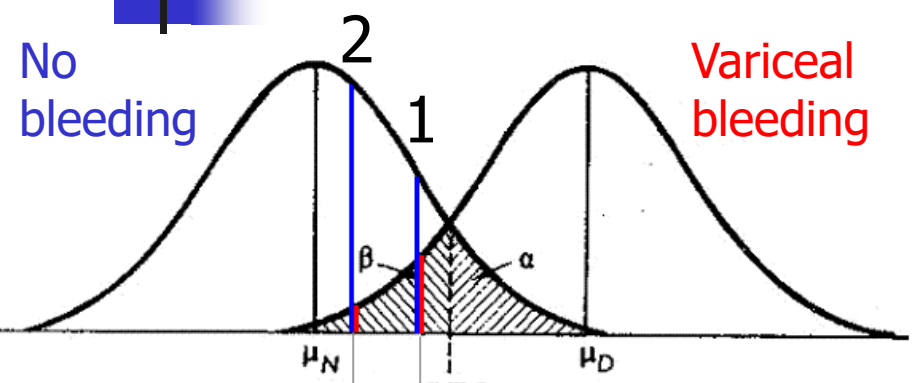
The smaller the HVPG, the lesser the risk of bleeding.

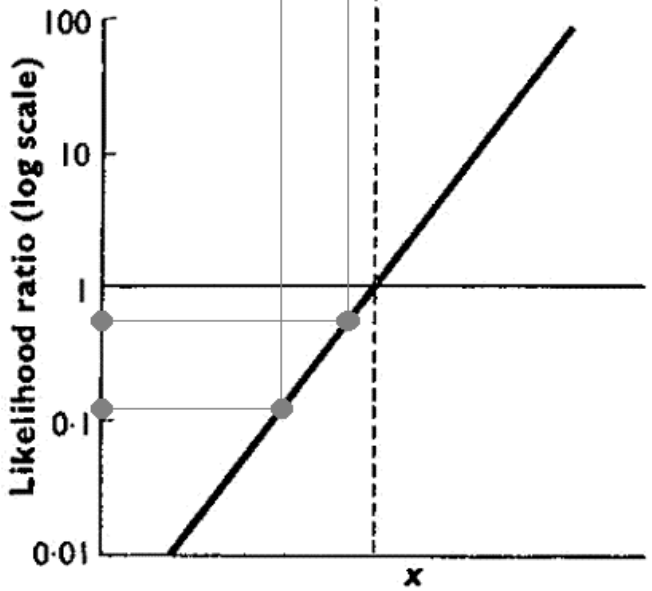
The larger the HVPG, the greater the risk of bleeding.

The risk can be expressed as the likelihood ratio (the ratio between the probability densities or heights) of the two curves at the actual HVPG level.

Using Bayes' theorem the probability of bleeding can be estimated for a patient.

Likelihood ratio based on the quantitative test value without dichotomization (examples)





Likelihood ratio based on probability densities or heights of the two curves at the actual HVPG level.

Example 1: Likelihood ratio (bleeding/non bleeding) is 0.5

Example 2: Likelihood ratio (bleeding/non bleeding) is 0.12

Thus for a given patient the risk of bleeding can be estimated from his/her HVPG level.

Utilizing the combined information from more variables

- Besides the key variable HVPG other descriptive variables (e.g. symptoms, signs and liver function tests) may influence the risk of bleeding from varices.
- By utilizing such information, estimation of the risk of bleeding may be improved.
- Such predictive models may be developed using multivariate statistical analysis like logistic regression or Cox regression analysis.

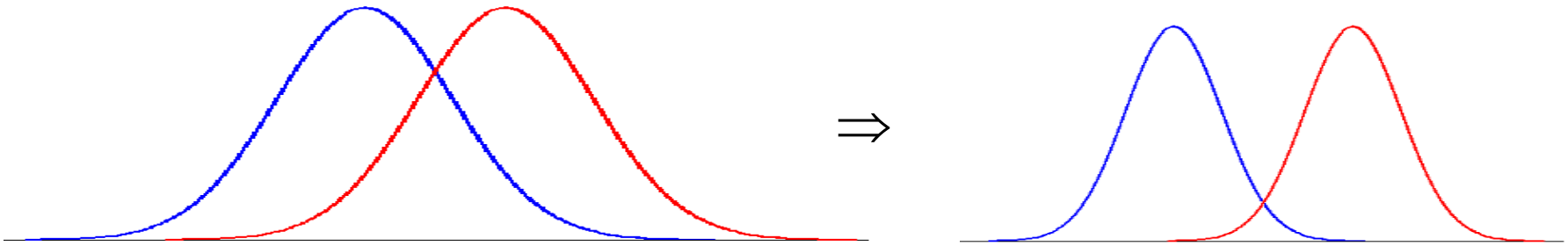
Combined information from more variables (example)

Cox model for prediction of bleeding in cirrhosis:

PI-b = 0.103 × HVPG +
0.288 × Pugh score +
0.419 × size of varices +
0.927 [if previous variceal bleeding occurred]

This model had more predictive power than HVPG alone.

From: Merkel C, et al. Gastroenterology 1992;102:973-9.



Conclusion

- The simple diagnostic tests are important tools in the evaluation of patients.
- They nevertheless have limitations, which are a consequence of dichotomization of quantitative variables, whereby information is lost.
- Quantitative variables should be kept as such whenever possible.
- Combination of more variables may improve prediction.
- Dichotomization of variables should preferably only be used in the last step, when a binary decision (i.e. yes/no in regard to diagnosis or therapy) has to be made.