Chapter 1– Basics and Statistics of Analytical Biochemistry

Biochemistry and Molecular Biology (BMB)

1.1 Biochemical Studies

1.2 Units of Measurements

1.3 Weak Electrolytes

1.4 Buffer Solution

1.6 Quantitative Biochemical Measurements

1.7.1-1.7.2 Principle of Clinical Biochemical Analysis

Others:

- Receiver Operating Characteristic Curve
- Diagnosis Sensitivity and Specificity

Basic principles

- **Molarity** : Number of moles of the substances in 1 dm^3 of solution.

- **One mole**: equal to molecular mass of the substance

- **Molecular mass**:

  Da: *daltons*

  kDa: *Kilodaltons* =1000 Da

  M_r: *no unit*

  Relative molecular mass

  = the molecular mass of a substance relative to 1/12 of the atomic mass of the ^{12}C .

Units for Different Concentrations

Table 1.5 **Interconversion of mol, mmol and μmol in different volumes to give different concentrations**

Molar (M)	Millimolar (mM)	Micromolar (μM)
1 mol dm^{-3} 1 mol l^{-3}	1 mmol dm^{-3}	1 μmol dm^{-3}
1 mmol cm^{-3}	1 μmol cm^{-3}	1 nmol cm^{-3}
1 μmol mm^{-3}	1 nmol mm^{-3}	1 pmol mm^{-3}

Biological substances are most frequently found at relatively low concentrations and in *in vitro* model systems the volumes of stock solutions regularly used for experimental purposes are also small. The consequence is that experimental solutions are usually in the mmol dm^{-3}, μmol dm^{-3} and nmol dm^{-3} range rather than molar. Table 1.5 shows the interconversion of these units.

Ion Strengths

Reason of deviation:

Presence of electrolytes will result in electrostatic interaction with other ions and solvents

Total ion charge in solution

$$M=1/2 *(c_1z_1^2 + c_1z_1^2 +\ldots+ c_nz_n^2)$$

c_1, c_2, …c_n: concentrations of each ion in *molarity*

z_1, z_2, …z_n: charge on the individual ion

Example 2 CALCULATION OF IONIC STRENGTHS

Question

Calculate the ionic strength of (i) 0. 1 M NaCl, (ii) 0. 1 M NaCl + 0.05 M KNO_3 + 0.01 M Na_2SO_4.

Answer

Ionic strength can be calculated using the equation $\mu = \frac{1}{2}\sum cz^2$.

(i) Calculating cz^2 for each ion:

$$Na^+ = 0.1 \times (+1)^2 = 0.1\,M$$

$$Cl^- = 0.1 \times (-1)^2 = 0.1\,M$$

Hence

$$\frac{1}{2}\sum cz^2 = 0.2/2 = 0.1\,M$$

(ii)

$$Na^+ = 0.1 \times (+1)^2 + 0.02 \times (+1)^2 = 0.12\,M$$

$$Cl^- = 0.1 \times (-1)^2 = 0.10\,M$$

$$K^+ = 0.05 \times (+1)^2 = 0.05\,M$$

$$NO_3^- = 0.05 \times (-1)^2 = 0.05\,M$$

$$SO_4^{2-} = 0.01 \times (-2)^2 = 0.04\,M$$

Hence

$$\frac{1}{2}\sum cz^2 = \frac{1}{2}(0.36) = 0.18\,M$$

Activity and Activity Coefficients

Activity : the effective concentration in solution

A_x = [Concentration] γ_x

γ_x : Activity coefficient

- The coefficient establish the relationship between activity and concentration.
- It will **decrease** when the **ionic strength increases** (include concentration, charge and ion mobility)

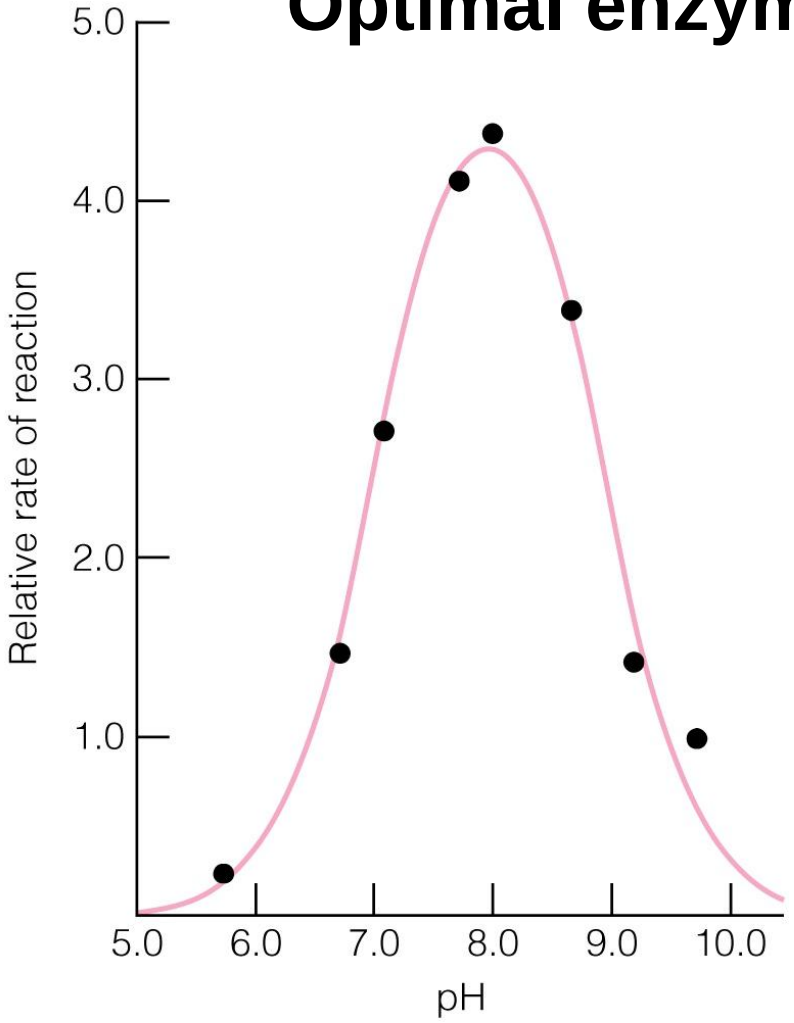
e.g. 0.001 M Mg^{2+} 0.872

Fe^{3+} 0.738

Except for very diluted solution, the effective concentrations are usually less than the actual concentration

Preparation of Buffer Solution

Optimal enzyme activity pH 8



α-Chymotrypsin: catalyzed cleavage of the C-N bond

Henderson-Hasselbalch Equation

For a weak acid, which dissociates as follows:

$$HA \rightarrow H^+ + A^-$$

$$\text{equilibrium constant} = K_{eq} = K_a = \frac{[H^+] \times [A^-]}{[HA]}$$

$$\log_{10}K_a = \log_{10}[H^+] + \log_{10}[A^-] - \log_{10}[HA]$$

$$-\log_{10}[H^+] = -\log_{10}K_a + \log_{10}[A^-] - \log_{10}[HA]$$

$$pH = pK_a + \log_{10}\left(\frac{[A^-]}{[HA]}\right)$$

$$pH = pK_a + \log_{10}\left(\frac{[\text{conjugate base}]}{[\text{conjugate acid}]}\right) = pK_a + \log_{10}\left(\frac{[\text{proton acceptor}]}{[\text{proton donor}]}\right)$$

Why is pKa useful?

$$pH = pK_a + \log_{10}\left(\frac{[A^-]}{[HA]}\right)$$

Perhaps it is useful to look at this in another way: if we consider the situation where the acid is one half dissociated, in other words **where [A-] is equal to [HA],** then, substituting in the Henderson-Hasselbalch Equation

pH = pKa + log10(1)

pH = pKa + 0

pH = pKa

This means that an acid is half dissociated when the pH of the solution is numerically equal to the pKa of the acid.

$$pH = pK_a + \log_{10}\left(\frac{[A^-]}{[HA]}\right)$$

$$\mathbf{HA \rightarrow H^+ + A^-}$$

Acid	K_a		pK_a
Trichloroacetic	2×10^{-1}	$=10^{-0.7}$	0.7
Dichloroacetic	5×10^{-2}	$=10^{-1.3}$	1.3
Monochloroacetic	1.6×10^{-3}	$=10^{-2.8}$	2.8
Formic	2.1×10^{-4}	$=10^{-3.7}$	3.7
Benzoic	7.8×10^{-5}	$=10^{-4.1}$	4.1
Acetic	1.9×10^{-5}	$=10^{-4.7}$	4.7
H_2CO_3	2.9×10^{-7}	$=10^{-6.5}$	6.5
H_2S	5.8×10^{-8}	$=10^{-7.2}$	7.2
HCN	1.3×10^{-9}	$=10^{-8.9}$	8.9

Acids with the lowest pKa values are able to dissociate in solutions of low pH, i.e. even where the hydrogen ion concentration is high.

Acids with higher pKa values dissociate only in solutions of high (more alkaline) pH.

Example 3 CALCULATION OF pH AND THE EXTENT OF IONISATION OF A WEAK ELECTROLYTE

Question

Calculate the pH of a 0.01 M solution of acetic acid and its fractional ionisation given that its K_a is 1.75×10^{-5}.

Answer

To calculate the pH we can write:

$$K_a = \frac{[\text{acetate}^-][\text{H}^+]}{[\text{acetic acid}]} = 1.75 \times 10^{-5}$$

Since acetate and hydrogen ions are produced in equal quantities, if x = the concentration of each then the concentration of unionised acetic acid remaining will be $0.01 - x$. Hence:

$$1.75 \times 10^{-5} = \frac{(x)(x)}{0.01 - x}$$

$$1.75 \times 10^{-7} - 1.75 \times 10^{-5}x = x^2$$

This can now be solved either by use of the quadratic formula or, more easily, by neglecting the x term since it is so small. Adopting the latter alternative gives:

$$x^2 = 1.75 \times 10^{-7}$$

hence

$$x = 4.18 \times 10^{-4}\text{M}$$

hence

$$\text{pH} = 3.38$$

Note that this solution has ignored the activity coefficients of the acetate and hydrogen ions. They are 0.90 and 0.91 respectively at 0.01 M and 25 °C. Inserting these values into the above expression and assuming that the activity coefficient of acetic acid is unity gives:

$$1.75 \times 10^{-5} = \frac{(x)0.90(x)0.91}{0.01 - x}$$

Solving this equation for x gives a value of 4.61×10^{-4}M, and hence a pH of 3.33. This illustrates the relatively small influence of activity coefficients in this case.

The fractional ionisation (α) of the acetic acid is defined as the fraction of the acetic acid that is in the form of acetate and is therefore given by the equation:

$$\alpha = \frac{[\text{acetate}]}{}$$

Quantitative Biochemical Measurements

- What to study? **Model**
- How to study **Method**
- Is the results correct? **Performance**
- How to interpret results? **Report**

Quantitative Biochemical Measurements

- Analytical Considerations:

(I) Test Model :

in vivo v.s. ***in vitro***

Material: urine, serum/plasma/blood

Matrix v.s Analyte

Sampling v.s population

in vivo v.s. *in vitro*

In vivo: In a living cell or organism

In vitro: Biological or chemical work done in the test tube

(in glass)

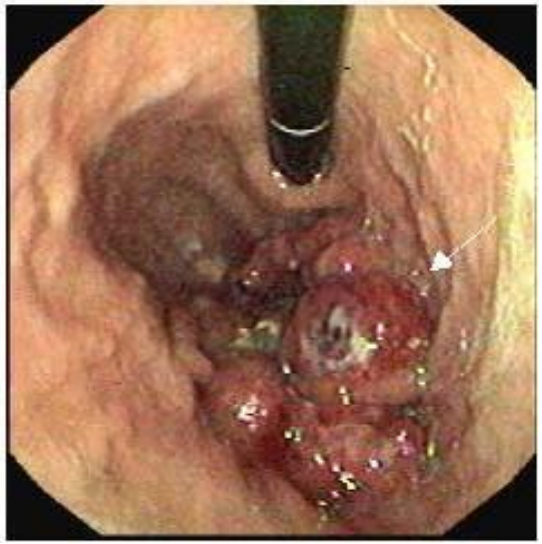
Sampling v.s Population

Population: Representative portion of analyte

Heterogeneous v.s Homogeneous

Extraction Methods:

- Liquid extraction
- Solid-phase extraction
- Laser microdisection (cancer cell)
- ……etc



Quantitative Biochemical Measurements

(II) Selection of Analytical Methods

- Qualitative v.s Quantitative analysis
- Chemical and physical properties of analyte
- Precision, accuracy and detection limit
- Interference from matrix
- Cost and value
- Possible hazard and risk

NOTE

Precision v.s. Accuracy for Quantitative or Numerical data

Accuracy— a measure of rightness.

Accuracy can be defined how closely a measured value agrees with the correct value.

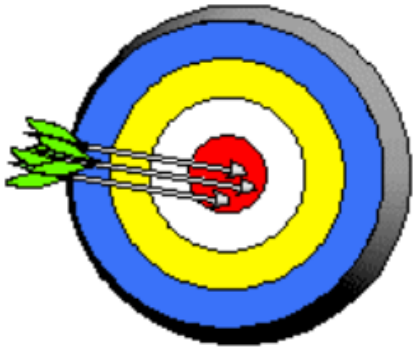
Accuracy is determined by comparing a number to a known or accepted value.

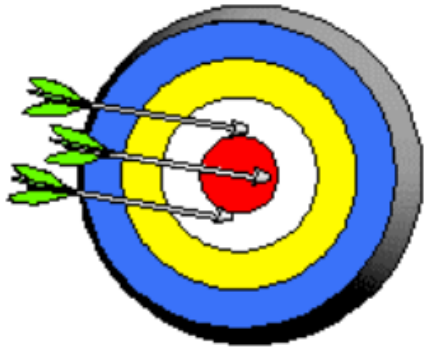
Precision — a measure of exactness.

Precision can be defined how closely individual measurements agree with each other.
It is sometimes defined as reproducibility

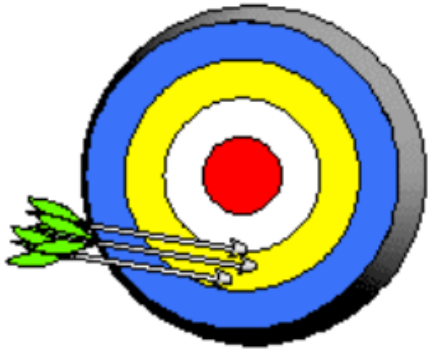
Accuracy	Precision
√	√

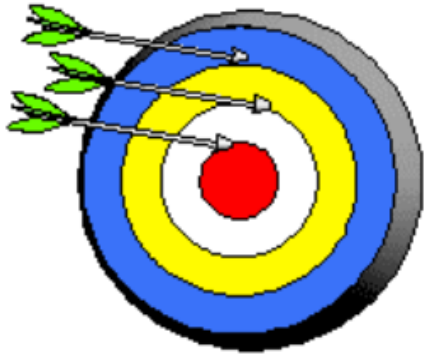
Accuracy	Precision
√	X





The average is close to the center but the individual values are not similar





Accuracy	Precision
X	√

Accuracy	Precision
X	X

Physical Basis of Analytical Methods

Physical properties that can be measured with some degree of precision	Examples of properties used in the		
	Protein	Lead	Oxygen
Extensive			
Mass	+	+	
Volume			+
Mechanical			
Specific gravity	+		
Viscosity	+		
Surface tension	+		
Spectral			
Absorption	+	+	
Emission			
Fluorescence			
Turbidity	+		
Rotation			
Electrical			
Conductivity			
Cuurent/voltage			+
Half-cell potential			+
Nuclear			
Radioactivity	+		

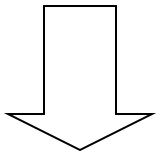
Major manipulative steps in a generalized method of analysis

Purification of the test substance
↓
Development of a physical characteristic by the formation of a derivative
↓
Detection of an inherent or induced physical characteristic
↓
Signal amplification
↓
Signal measurement
↓
Computation
↓
Presentation of result

Quantitative Biochemical Measurements

(III) Experimental Errors

- Systematic error
- Random error



Standard Operation Procedures (SOP)

Systematic Error

- Constant or proportional (**Bias)**
- Also called

Overestimation /underestimation

(1) **Analyst error**: pipette, calibration, solution preparation, method design

(2) **Instrumental error**: contamination of instrument, power fluctuation, variation in T, pH, electronic noise

(3) **Method error**: side reaction, incomplete reaction

Identification of Systematic Errors

- Blank sample
- Standard reference sample
- Alternative methods
- External quality assessment sample

Random Error

- Variable, either positive or negative
- also called

  Indeterminate error

(1) **Instrumental error**: random electric noise

Standard Operating Procedures (SOP)

Detailed, written instructions to achieve uniformity of the performance of a specific process;

Include:

- Quantity/quality of reagent
- Preparation of standard solution
- Calibration of instrument
- Methodology of actual analytical procedures

Assessment of Performance of Analytical Method

The NEW ENGLAND JOURNAL of MEDICINE

Question:

1. What is the correlation of the **memory of immune cell** and cancer metastasis?
2. Will it affect the survival rate?

（大腸直腸癌）

Franck Pagès, M.D., Ph.D., Anne Berger, M.D., Ph.D., Matthieu Camus, M.Sc.,
Fatima Sanchez-Cabo, Ph.D., Anne Costes, B.S., Robert Molidor, Ph.D.,
Bernhard Mlecnik, M.Sc., Amos Kirilovsky, M.Sc., Malin Nilsson, B.S.,
Diane Damotte, M.D., Ph.D., Tchao Meatchi, M.D., Patrick Bruneval, M.D., Ph.D.,
Paul-Henri Cugnenc, M.D., Ph.D., Zlatko Trajanoski, Ph.D.,
Wolf-Herman Fridman, M.D., Ph.D., and Jérôme Galon, Ph.D.

NEJM, 353, 2654-2666, 2005

Background

The role of tumor-infiltrating (浸潤) immune cells in the early metastatic invasion (轉移性侵犯) of colorectal cancer（直腸癌）is unknown.

Methods

We studied pathological signs of early metastatic invasion (venous emboli 靜脈栓塞 and lymphatic 淋巴 and perineural invasion(神經旁間隙) in 959 specimens of resected colorectal cancer. The local immune response within the tumor was studied by flow cytometry (39 tumors), low density-array real-time polymerase-chain-reaction assay (75 tumors), and tissue microarrays (415 tumors).

Table 1. Disease-free and Overall Survival among 959 Patients with Colorectal Cancer.

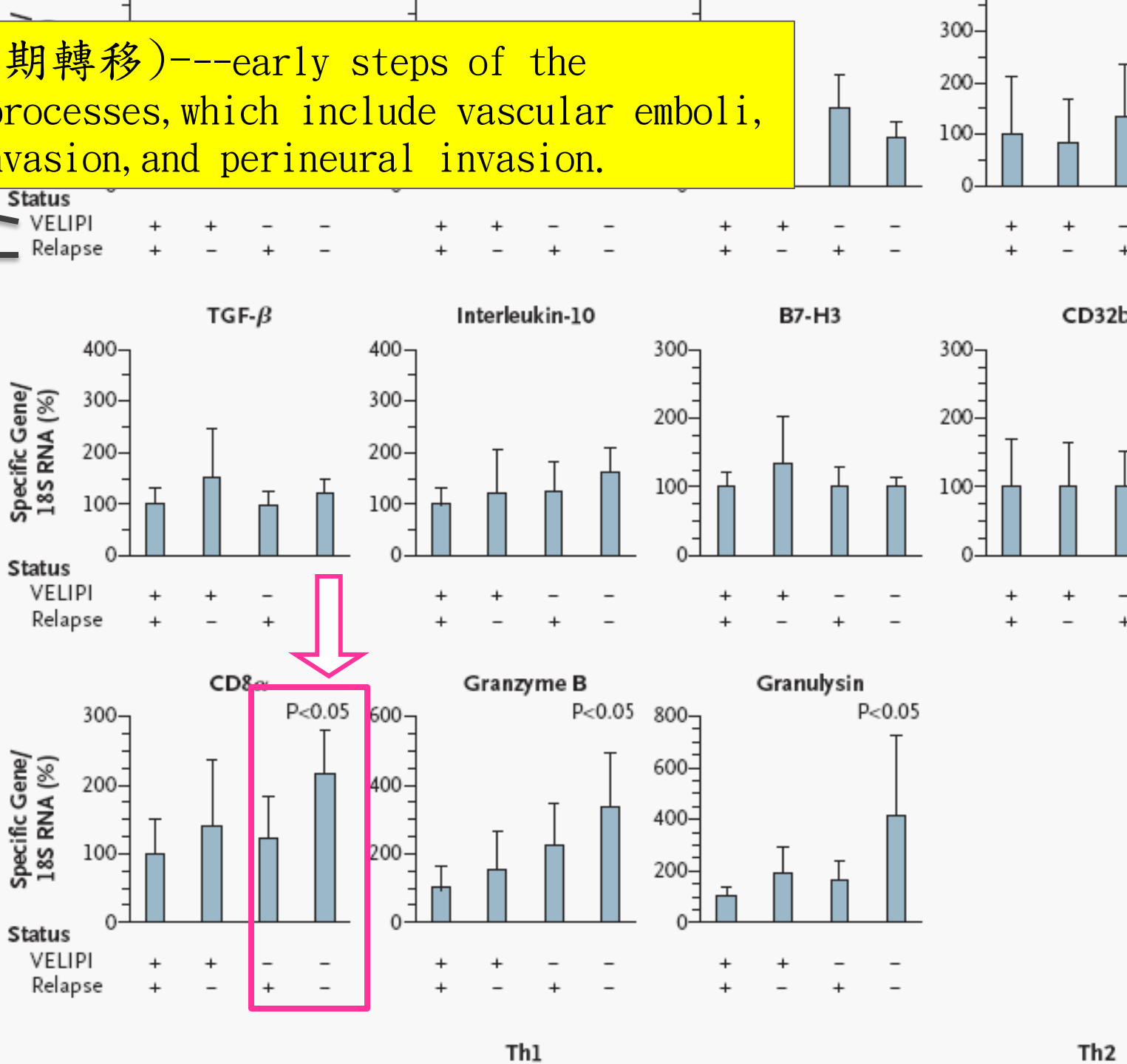
Characteristic	No. of Patient	Disease-free survival			Overall survival		
		5 yr %	Median mo	P value	5 yr %	Median *mo*	P Value*
Tumor (T) stage†				<0.001			<0.001
pTis	39	48.7	55.7		48.7	55.7	
pT1	54	42.6	52.2		44.4	53.8	
pT2	156	40.4	43.6		44.2	49.1	
pT3	502	23.7	16.5		26.7	25.8	
pT4	208	16.8	1.6		17.8	16.8	
Nodal (N) status				<0.001			<0.001
Negative	568	35.4	34.6		38.6	43.1	
Positive	384	15.1	4.3		16.7	16.9	
Nx‡	7						

■ **Disease-free survival** (DFS) denotes the **chances of staying free of disease** after a particular treatment for a group of individuals suffering from a cancer.

■ **Overall survival** is a term that denotes the **chances of staying alive** for a group of individuals suffering from a cancer.

VELIPI（早期轉移)---early steps of the metastatic processes, which include vascular emboli, lymphatic invasion, and perineural invasion.

Relapse
復發



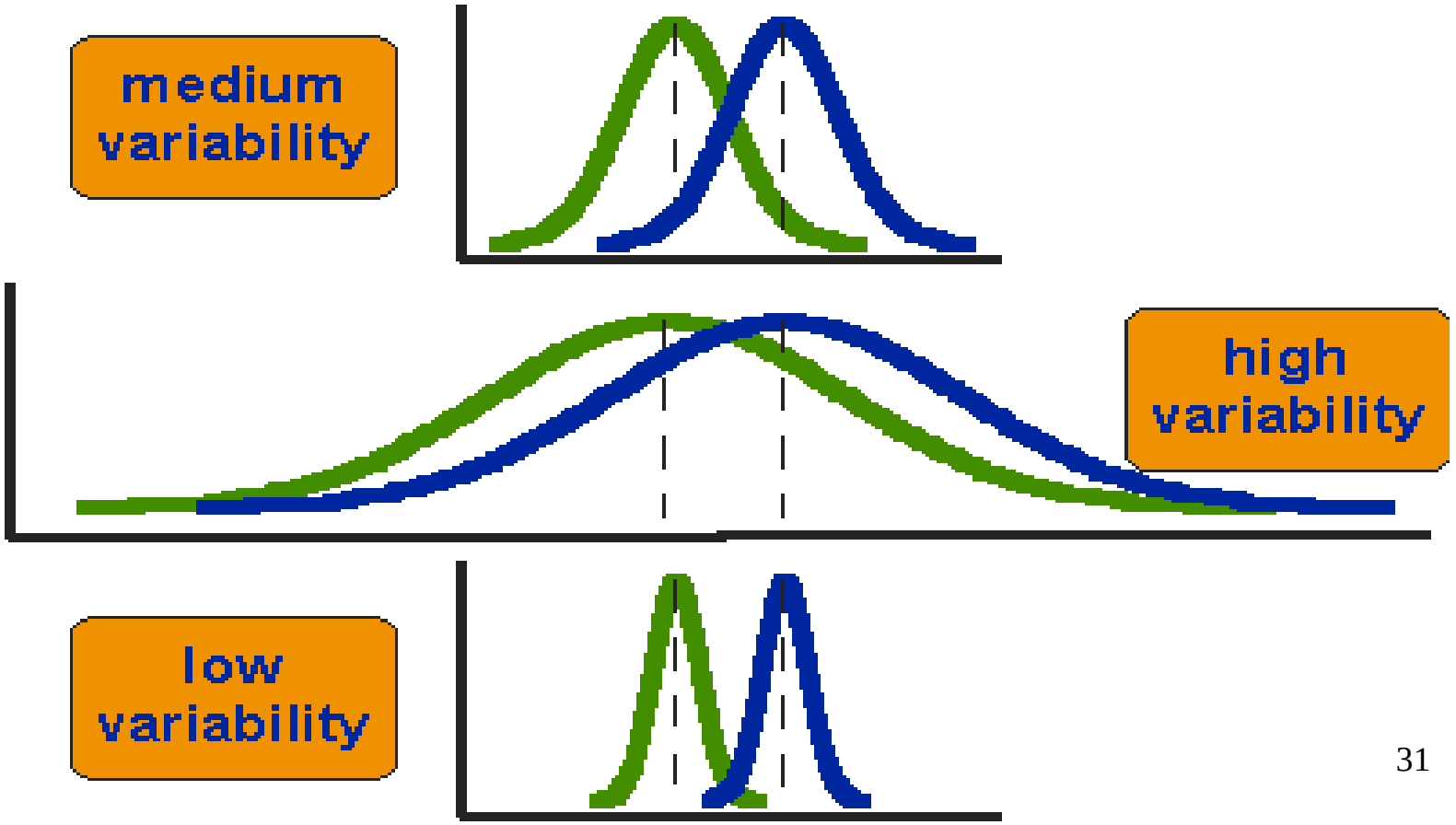
Interpretation of Quantitative Data

Table I			
Levels of LDE in the CSF of Administrators and Controls			
Group	Number	Mean	SD
Administrators	25	25.83	5.72
Controls	25	17.25	4.36

Is the difference of measured mean values from the two groups significantly different ?

How do we evaluate the data ?
Are the two groups different?

Normal control (健康) **52 54** **Cancer Patient (癌症)**



Normal v.s Patient?

A. Discrimination - Comparison of Data Groups

1. 2 groups with equal variances
2. 2 groups with unique variances

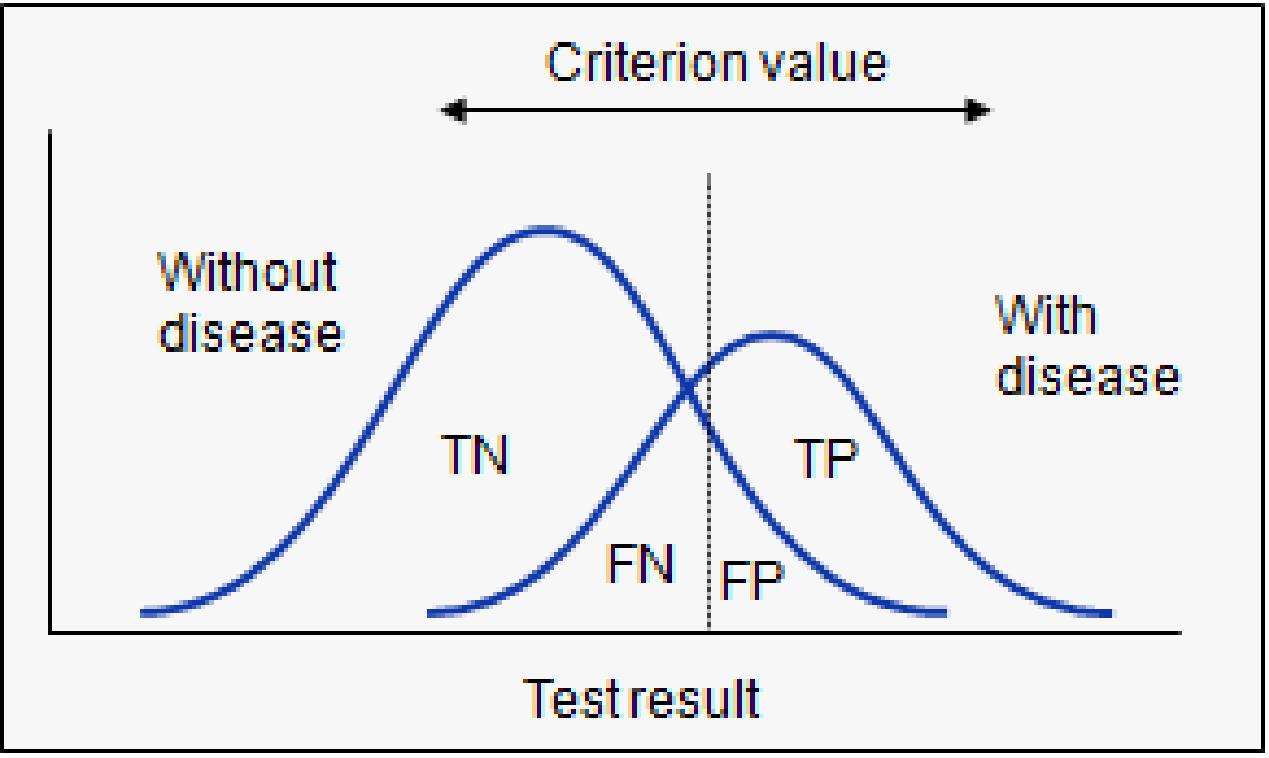
B. Receiving Operating Characteristic (ROC) curve

1. 2 X 2 contingency table
2. sensitivity & specificity
3. plotting ROC curve
4. uses of ROC curve

When the two study groups do have statistically significant difference, how do evaluate the correlation of any new data with the two groups?

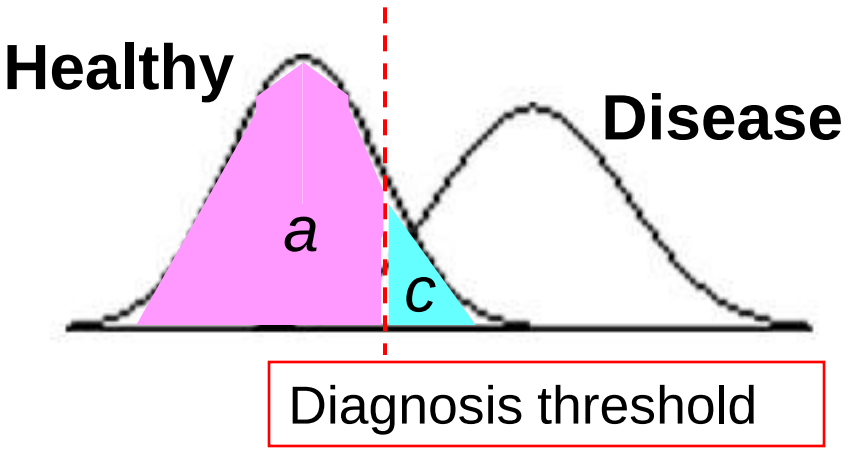
Receiver Operating Characteristics Curve (ROC curve analysis)

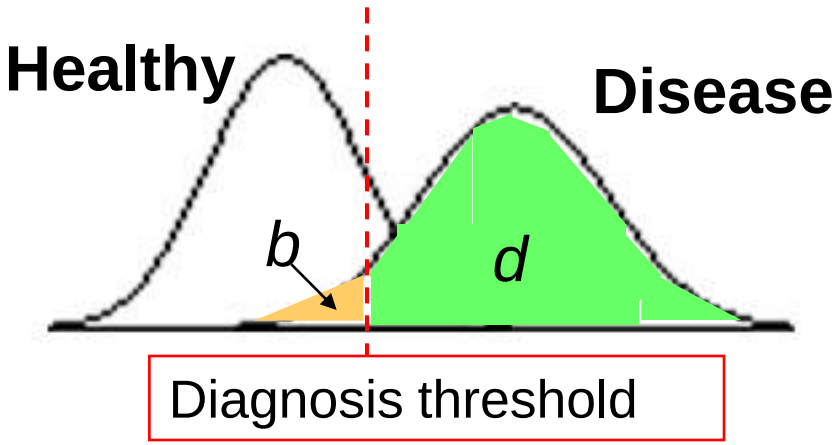
The diagnostic performance of a test, or the accuracy of a test to discriminate diseased cases from normal cases is evaluated using Receiver Operating Characteristic (ROC) curve analysis



TN: true negative
FN: false negative
TP: true positive
FP: false positive

2 x 2 Contingency Table

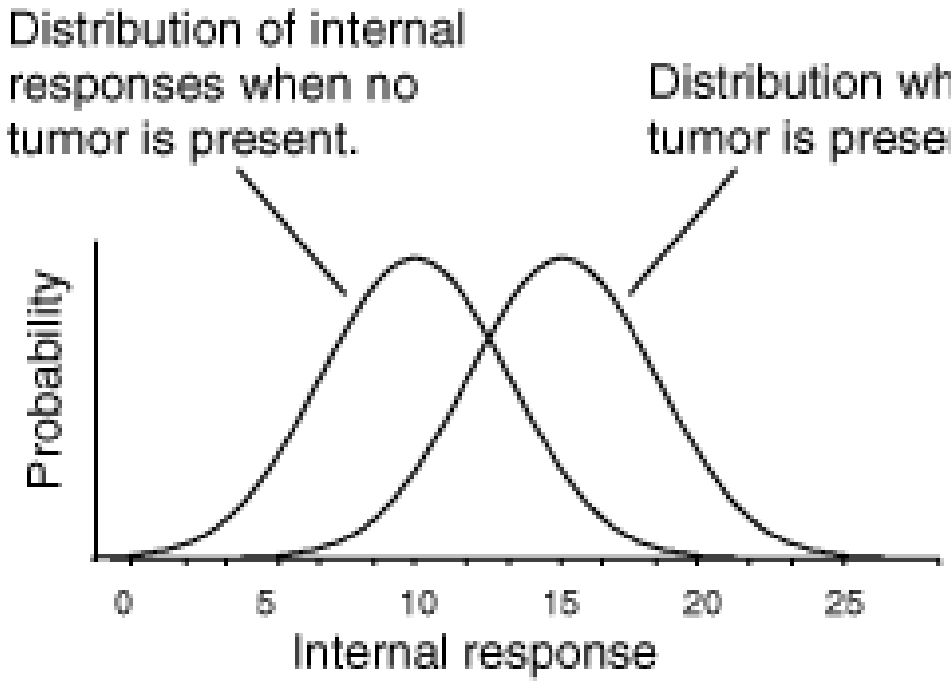


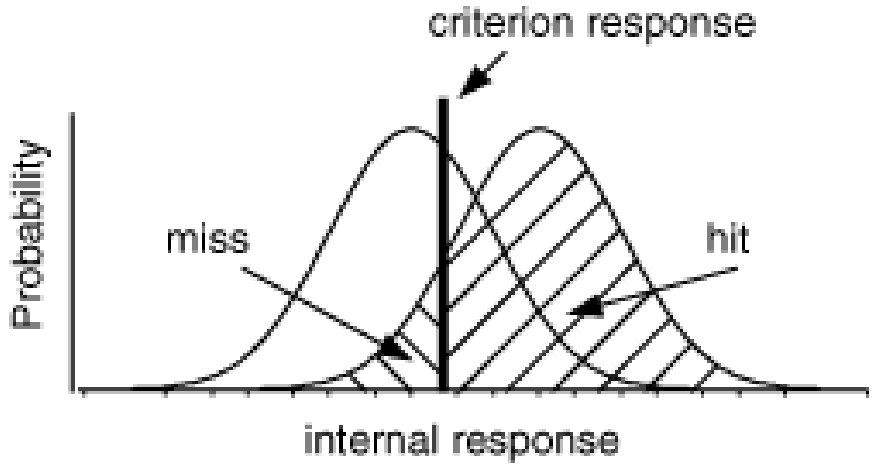


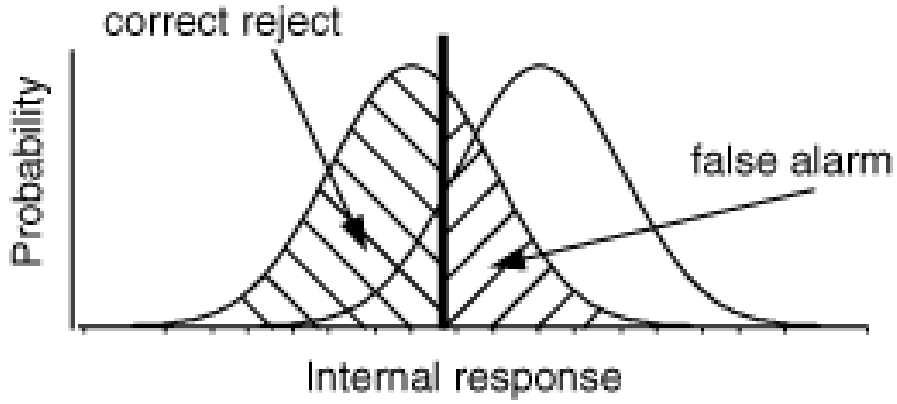
Result	Disease (true)		Total
	Absent	Present	
Normal (negative)	*a*	*b*	*a*+*b*
Disease (positive)	*c*	*d*	*c*+*d*
total	*a*+*c*	*b*+*d*	*a*+*b*+*c*+*d*

Correct

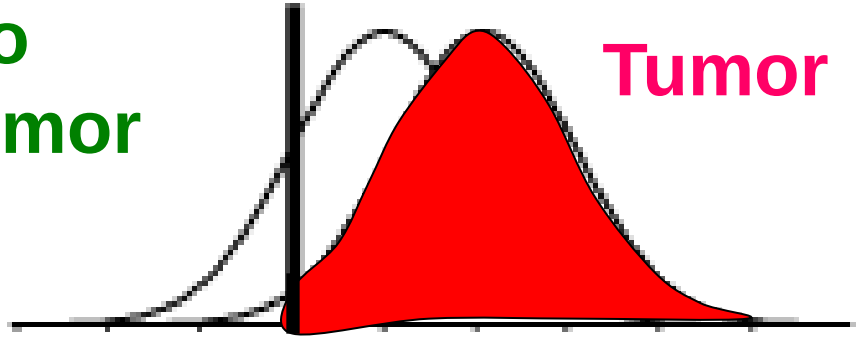
Wrong


Distribution of internal responses when no tumor is present.
Distribution when tumor is present.
Probability
0
5
10
15
20
25
Internal response

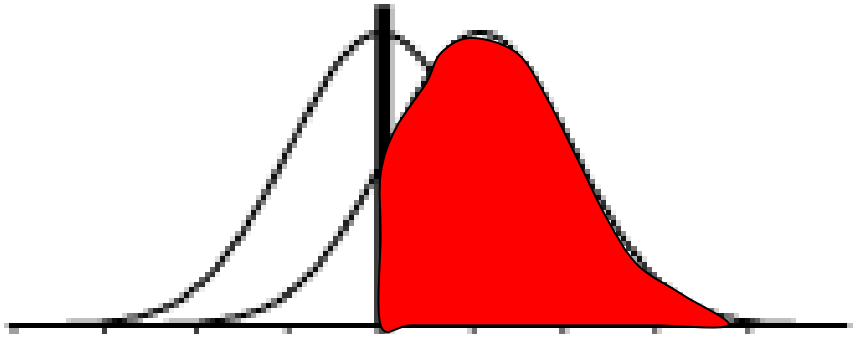

criterion response
Probability
miss
hit
internal response


correct reject
Probability
false alarm
Internal response

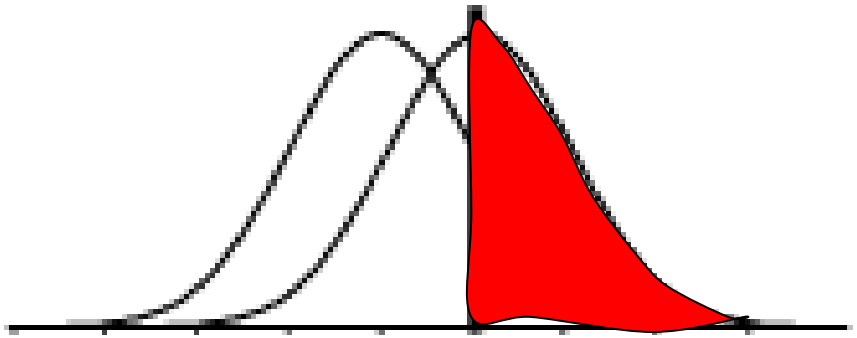
d' = 1



Hits = 97.5%
False alarms = 84%

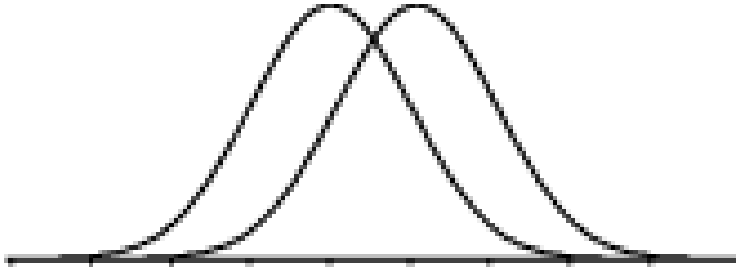


Hits = 84%
False alarms = 50%

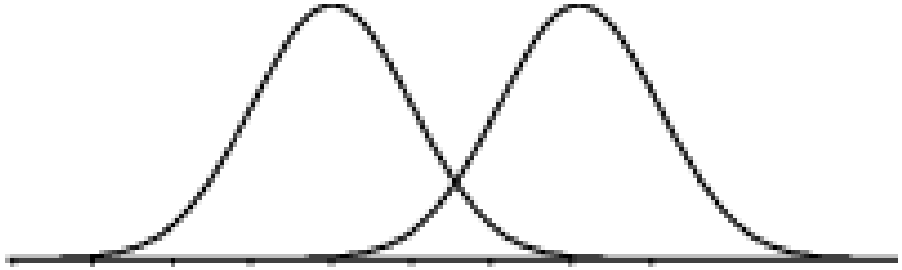


Hits = 50%
False alarms = 16%

Receiver Operating Characteristics (ROC) Curve



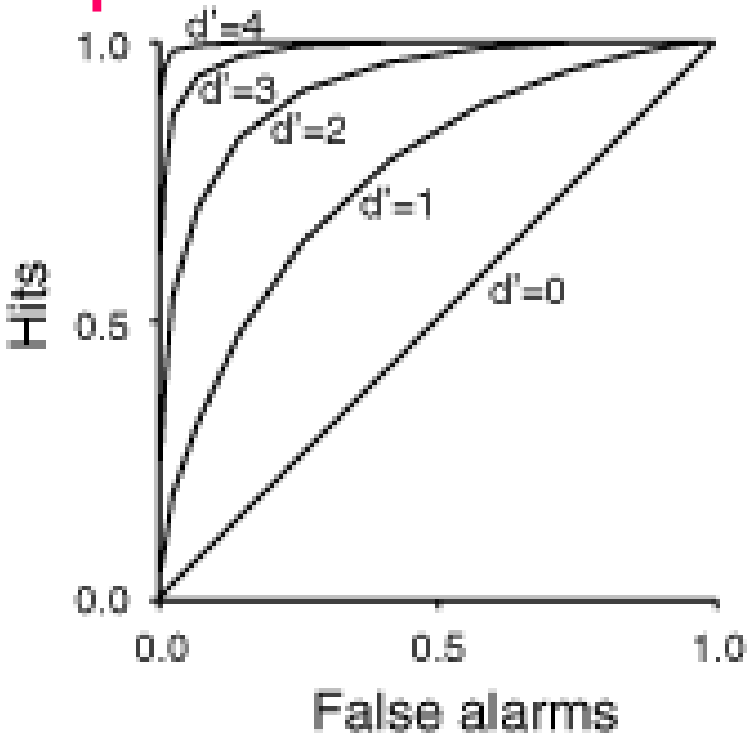
d' = 1 (lots of overlap)



d' = 3 (not much overlap)

High noise,
Lots of overlap

Low noise,
Not much overlap



ROC curves

Sensitivity & Specificity

Sensitivity

- probability that a test result will be positive when the disease is present (true positive rate, expressed as a percentage).

Sensitivity = P(disease positive | **disease**)

= *d* / (b+*d*)

– **True Positive**

(1-sensitivity) : False Negative

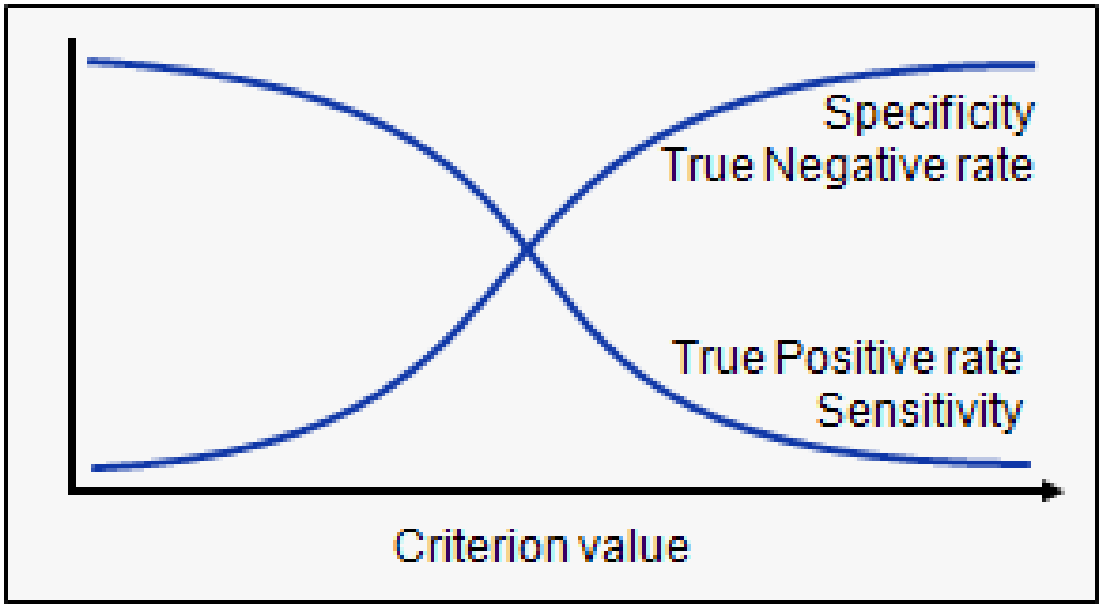
Sensitivity & Specificity

Specificity

- probability that a test result will be negative when the disease is not present (true negative rate, expressed as a percentage)
- Specificity = P(disease negative | **noraml**)
- = $a / (a+c)$
  - **True negative**

(1-specificity) : False positive

Sensitivity and Specificity versus Criterion Value



When you select a higher criterion value, the false positive fraction will decrease with increased specificity but on the other hand the true positive fraction and sensitivity will decrease.

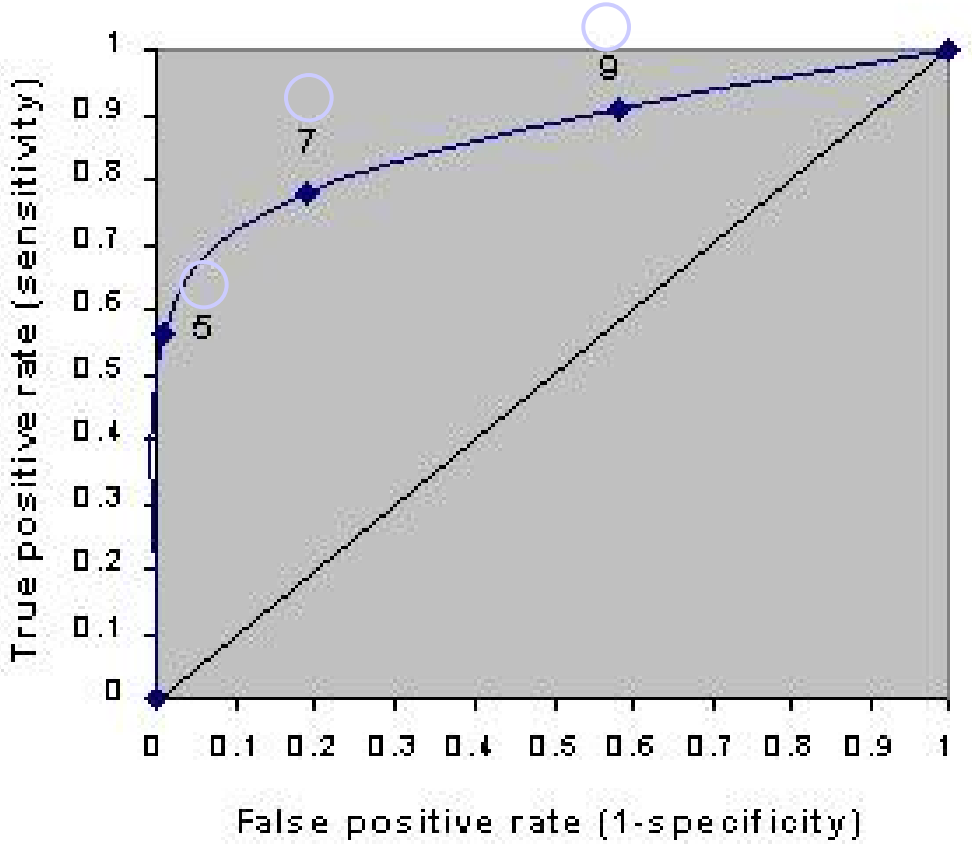
When you select a lower criterion value, then the true positive fraction and sensitivity will increase. On the other hand the false positive fraction will also increase, and therefore the true negative fraction and specificity will decrease.

Plotting ROC Curve
Receiver Operating Characteristics Curve

- **Y軸**：Sensitivity (true positive)
- **X軸**（1-specificity）(false positive)
  (normal, but wrong diagnosis)

Cutpoint	True Positives	False Positives
5	0.56	0.01
7	0.78	0.19
9	0.91	0.58

不同判定標準



Uses of ROC curve to Determine Diagnosis Threshold

- **Area under Curve (AUC)**
  - 0.9 ~ 1.0: excellent
  - 0.8 ~ 0.9: good
  - 0.7 ~ 0.8: fair
  - 0.6 ~ 0.7: poor