Diagnostic tests, Laboratory tests

I. Introduction
II. Informational values of a test
III. Consequences of the prevalence rate
IV. Sequential use of 2 tests
V. Selection of a threshold: the ROC curve
VI. Laboratory tests

Dr Emmanuel Chazard
Assistant Professor Lille 2 University
http://emmanuel.chazard.org

I. Introduction

- Medical decision making:
  - Relies on the observation of the reality
  - This observation *is not* the realty, the observer must keep a critical mind
- Diagnostic test:
  - Every information mean that is useful for medical decision making
  - Binary response (0 or 1)
  - Different kinds: interviewing, clinical exam, paraclinical exam (laboratory, imaging, etc.)
- Gold Standard
  - Test that is considered to be exact
  - Not always available of acceptable...

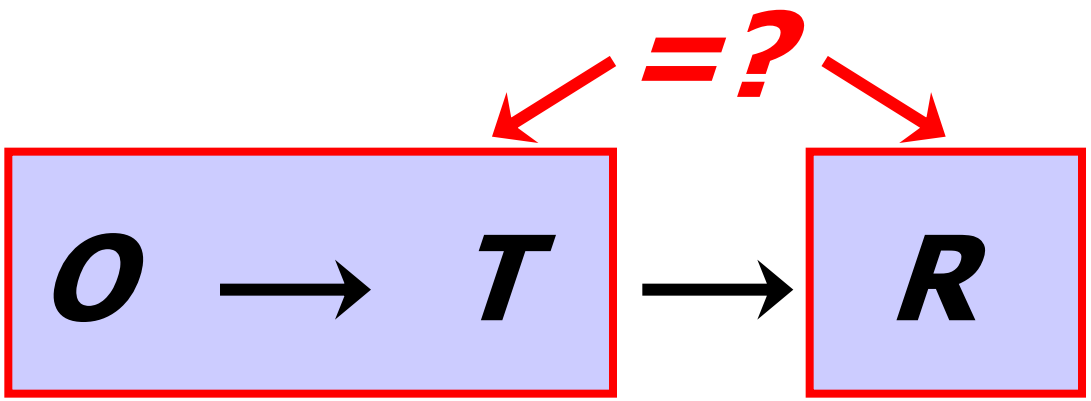
I. Introduction

- Examples:
  - Alzheimer's disease:
    - Gold standard: post-mortem examination of the brain
    - Usual test: clinical tests + brain imaging + laboratory
  - Down Syndrome of the fetus:
    - Gold standard: karyotype (dangerous and expensive)
    - Usual test: triple test (lab) + echography
  - Electrocardiogram interpretation
    - Gold standard: senior cardiologist
    - Test to evaluate: junior cardiologist, automated interpreter, etc.
- Reasons not to use the Gold Standard
  chronology, cost, risk, availability

II. Informationnal value of a test

A. Terminology
B. Intrinsic validity
C. Extrinsic validity
D. Exercise

Informational value of a test
A. Terminology

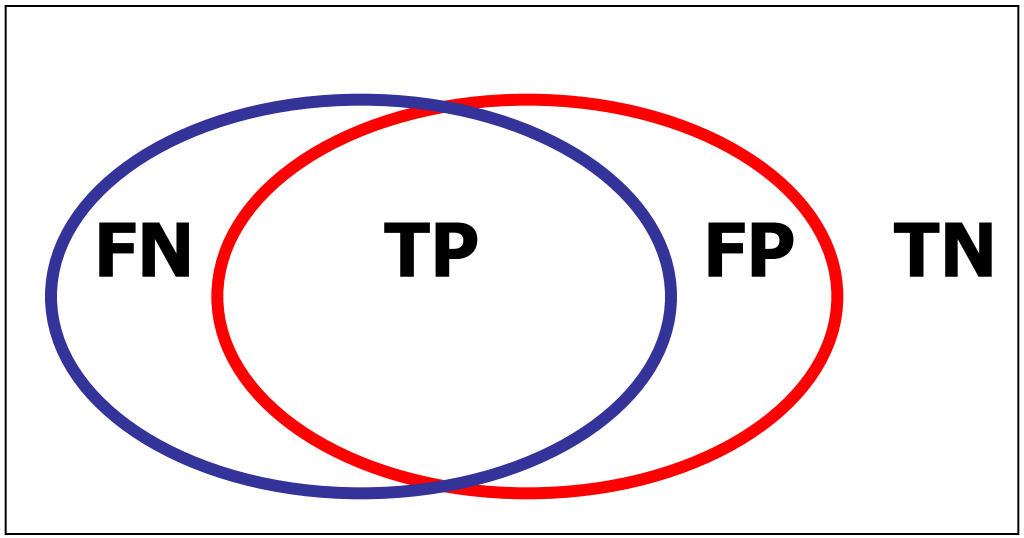


O=observator
T=test
R=realty

Informational value of a test
A. Terminology

Blue set: With/without disease D+/D-

Red set: Positive/negative test T+/T-



$$FN = T- \cap D+$$
$$TP = T+ \cap D+$$
$$FP = T+ \cap D-$$
$$TN = T- \cap D-$$

Dr Emmanuel Chazard
Assistant Professor Lille 2 University
http://emmanuel.chazard.org

Informational value of a test
A. Terminology

	D+	D-
T+	# TP	# FP
T-	# FN	# TN

Dr Emmanuel Chazard
Assistant Professor Lille 2 University
http://emmanuel.chazard.org

Informational value of a test
B. Intrinsic validity

- Experimental conditions:
  - We already know which patients are D+ or D-, we want to observe the result of the test
  - “Pre-test probabilities”
- Sensitivity = Se
  = P(T+ | D+)
  = TP / (TP+FN)
- Specificity = Sp
  = P(T- | D-)
  = TN / (TN+FP)

Informational value of a test
C. Extrinsic validity

- Practical use of the test:
  - We can observe the result of the test (T+ or T-) and we want to predict the status of the patients (D+ or D-)
  - “Post-test probabilities”
- Positive predictive value = PPV
  = P(D+ | T+)
  = TP / (TP+FP)
- Negative predictive value = NPV
  = P(D- | T-)
  = TN / (TN+FN)

Exercise

- We interest on 150 patients from the Urology unit:
  - test = PSA dosage (positive over 4ng/ml)
  - disease = confirmed prostate cancer
- Compute the following numbers:
  - Prevalence rate P=
  - Se=
  - Sp=
  - PPV=
  - NPV=

	D+	D-
T+	20	3
T-	80	47

Dr Emmanuel Chazard
Assistant Professor Lille 2 University
http://emmanuel.chazard.org

III. Consequences of the prevalence rate of a disease

A. Bayes' Theorem
B. Intuitive presentation
C. Exercise
D. How to proceed for screening?

Dr Emmanuel Chazard
Assistant Professor Lille 2 University
http://emmanuel.chazard.org

Consequences of the prevalence rate
A. Bayes' Theorem

- Let P be the prevalence rate, P=P(D+)

- PPV= $\dfrac{Se*P}{Se*P + (1-Sp)*(1-P)}$

- NPV = $\dfrac{Sp*(1-P)}{Sp*(1-P) + (1-Se)*P}$

Consequences of the prevalence rate
B. Intuitive presentation

	D+	D-	
T+	TP	FP	*PPV*
T-	FN	TN	*NPV*
	Se	*Sp*	

- Intrinsic validity: both Se & Sp are computed using only one column, separately from the other
- Extrinsic validity: PPV & NPV are computed using one line, including both columns
- The prevalence rate is in relation with the ratio between both columns
  => modifies the extrinsic validity, not the intrinsic one

Exercise

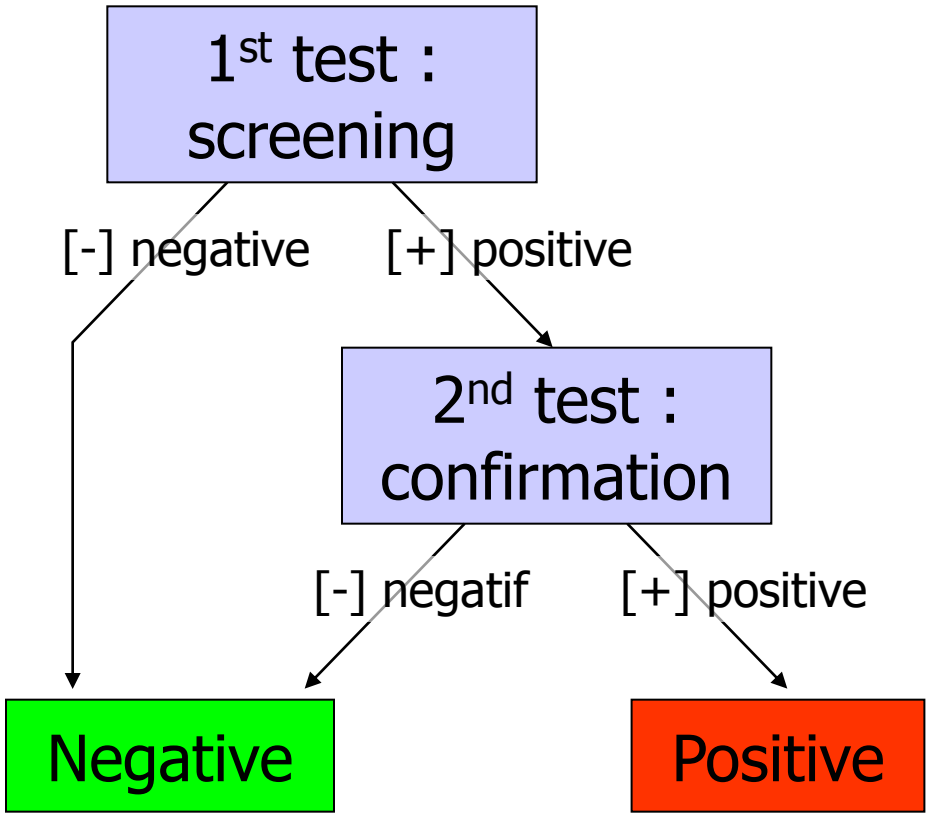
- In last exercise, we found:
  - P= 0.67
  - Se= 0.20
  - Sp= 0.94
  - PPV= 0.87
  - NPV= 0.37
- Compute the PPV in both next populations:
  - General population of men having age>74 years (P=2.66%)
  - General population of men having 45<age<55 years (P=0.06%)

Dr Emmanuel Chazard
Assistant Professor Lille 2 University
http://emmanuel.chazard.org

Consequences of the prevalence rate
D. How to proceed for screening?

- When a disease is rare (all the diseases are rare!):
  - Intrinsic validity is not affected
    …but useless for medical decision making
  - Increase of the NPV
    A negative test is comforting
  - Strong decrease of the PPV
    Strong risk to wrongly announce a diagnosis!
- How to proceed:
  - Only use the test only in a *very* meaningful context (increased prevalence rate): compatible clinical picture, exposure to an infectious disease, risk factor, etc.
  - Only few tests can be used for mass screening. Most often, they are coupled with confirmation tests (with higher PPV)

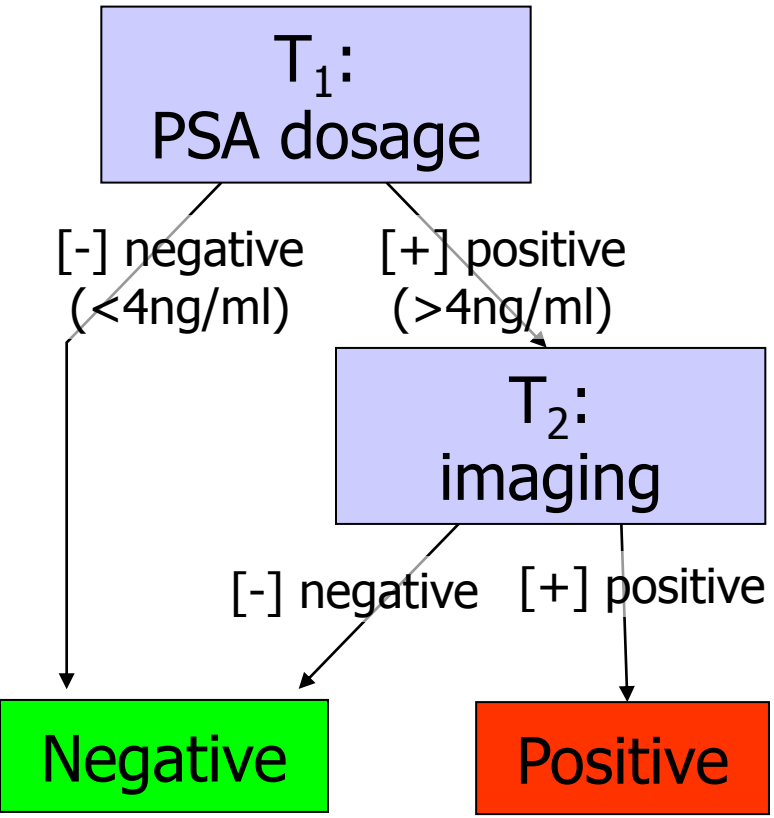
IV. Sequential use of 2 tests



Example of screening schema:

- Use a 1st test with high sensitivity
  and then excellent NPV, because in addition the disease is rare
- Confirm with a 2nd test having a high PPV
  notably because the population has been selected before!
  /! *Ideally, the misclassification of both tests is not due to the same factors, T1 and T2 are of different natures (e.g. lab & imaging)*
- The result is positive if and only if both tests are positive

Exercise



- We study 100,000 people from the general population having age>80 (P=2.66%).
- $T+ = T_1+ \cap T_2+$
- T_1: Se=0.2 & Sp=0.94
- T_2: Se=0.5 & Sp=0.95
- What is the PPV of T? use

$$PPV = \frac{Se*P}{Se*P + (1-Sp)*(1-P)}$$

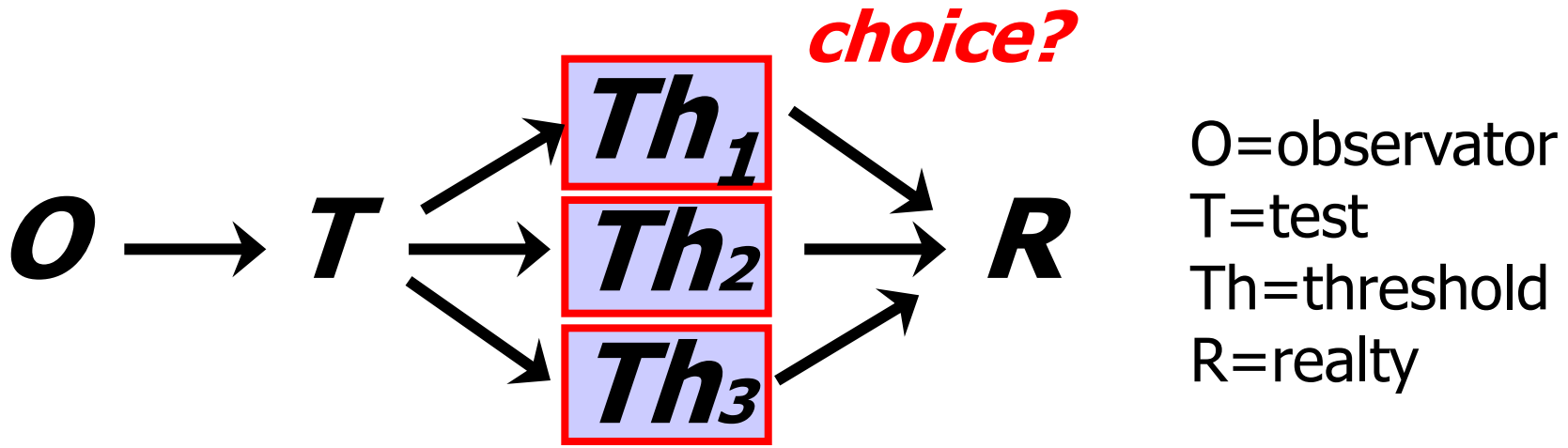
Dr Emmanuel Chazard
Assistant Professor Lille 2 University
http://emmanuel.chazard.org

V. Selection of a threshold: the ROC curve

A. Introduction
B. Construction
C. Interpretation
D. Selection of the best threshold
E. Exercise

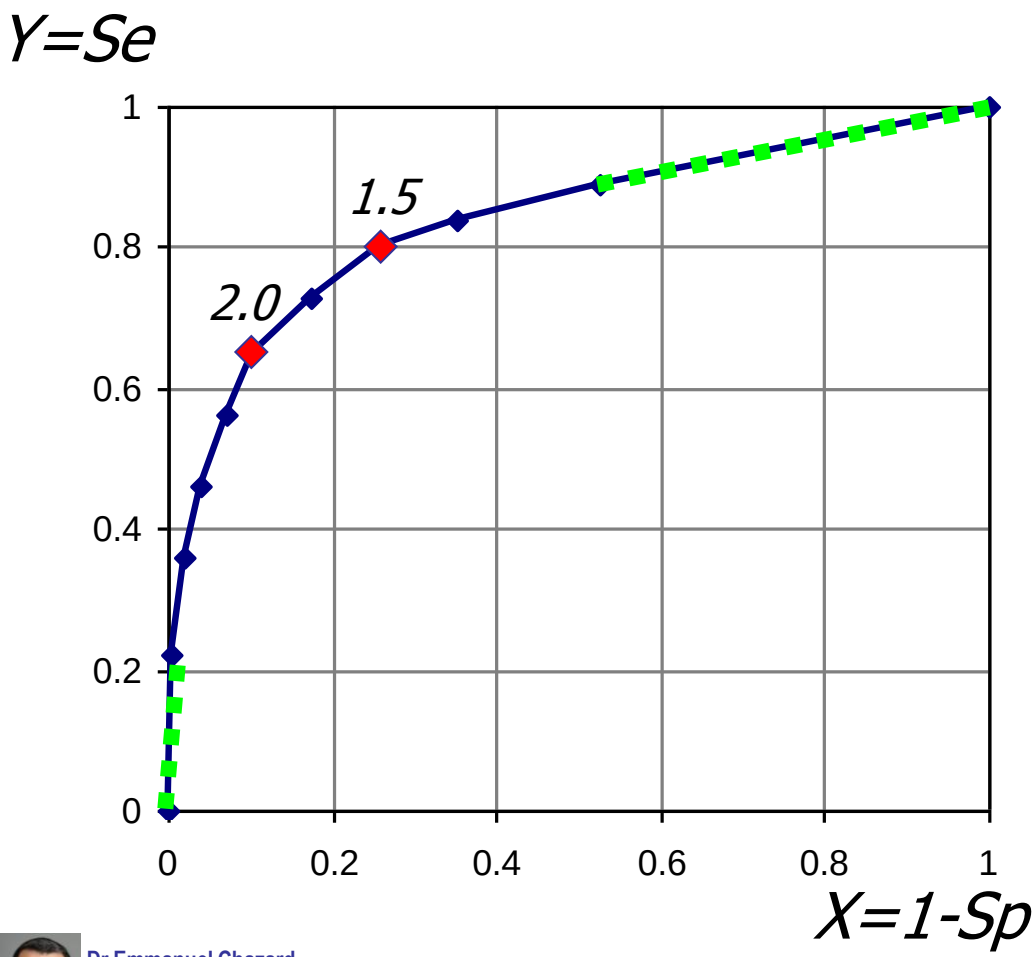
Dr Emmanuel Chazard
Assistant Professor Lille 2 University
http://emmanuel.chazard.org

Selection of a threshold: ROC curve
A. Introduction



- Problem:
  - A test provides with a quantitative response
  - We wish to *binarize* the output (yes/no)
  - Depending on the chosen threshold, the prediction differs…

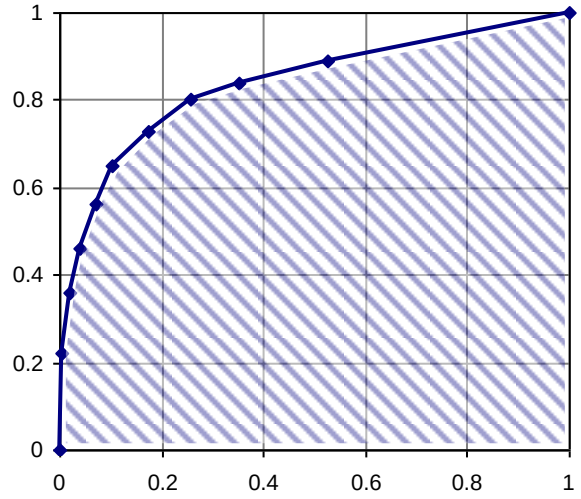
Selection of a threshold: ROC curve
B. Construction



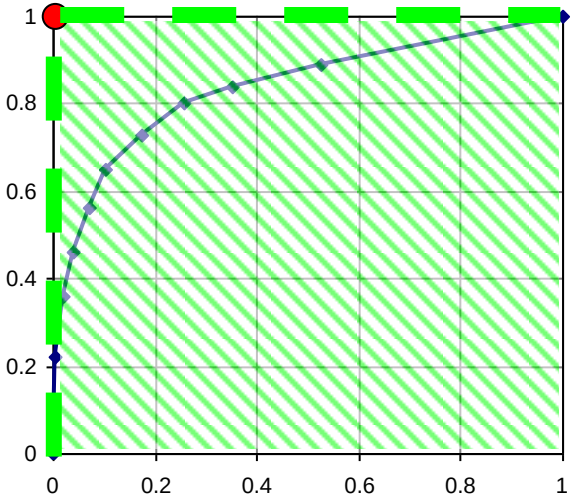
- ROC curve: simply displays the {Se;Sp} couples that are found using different thresholds
- Abscissa : X=1-Sp
- Ordinate : Y=Se

Dr Emmanuel Chazard
Assistant Professor Lille 2 University
http://emmanuel.chazard.org

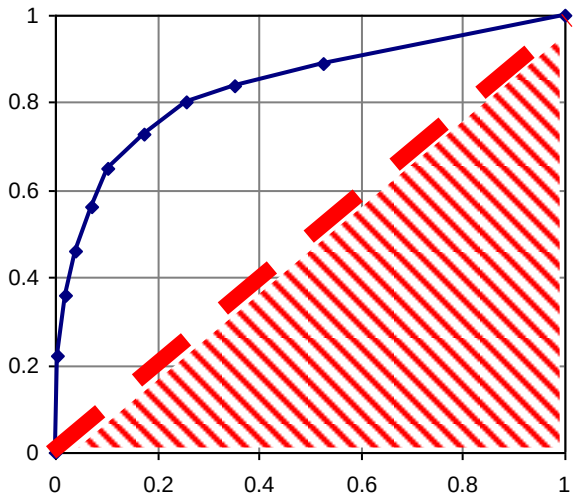
Selection of a threshold: ROC curve
C. Interpretation



- AUC=Area Under the Curve
- $0.5 \leq AUC \leq 1$



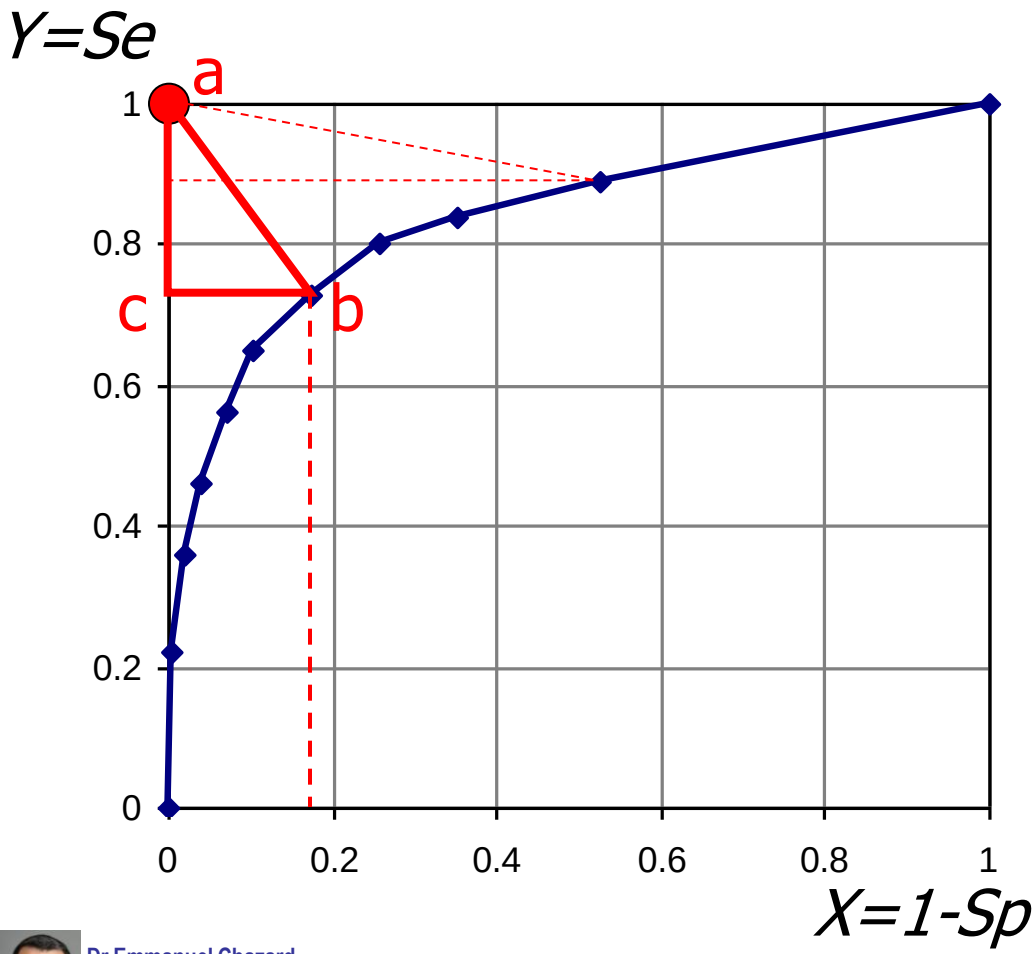
- Perfect point with Se=Sp=1
- AUC =1



- "hasard diagonal" (useless test)
- AUC=0.5

Dr Emmanuel Chazard
Assistant Professor Lille 2 University
http://emmanuel.chazard.org

Selection of a threshold: ROC curve
D. Selection of the best threshold



- Selection of the best threshold: point that is closer to the perfect point
- Minimizing the [a;b] distance means minimizing $(1\text{-Se})^2 + (1\text{-Sp})^2$

Dr Emmanuel Chazard
Assistant Professor Lille 2 University
http://emmanuel.chazard.org

Exercise

- As in exercise 1, we compute Se and Sp using several thresholds for the PSA dosage.
- We get the present table:

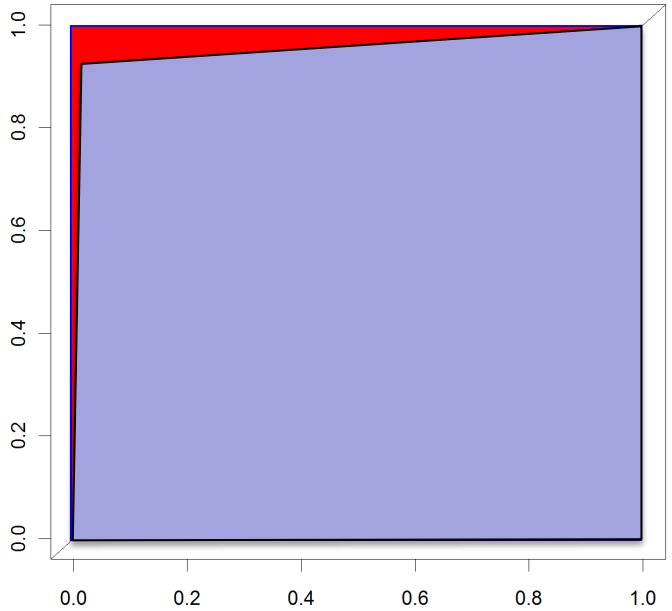
Threshold	Se	Sp
1	0.83	0.39
2	0.52	0.72
4	0.20	0.94

- Trace the ROC curve.
- Which threshold would you choose using the geometric criterion?

Dr Emmanuel Chazard
Assistant Professor Lille 2 University
http://emmanuel.chazard.org

Exercise

- We analyze the ability of the Glasgow University Interpreter to automatically detect “atrial fibrillation” by analyzing ECGs from a database.
- We find the following (there are 2 thresholds in the software):
  - Se=0.923 Sp=0.984 PPV=0.462 NPV=0.999
  - Se=0.923 Sp=0.987 PPV=0.522 NPV=0.999
  - Roc curve: AUC=0.96
- What do you think about it?
  Can it be used to replace the Cardiologists?
  Why is the PPV quite low?



Dr Emmanuel Chazard
Assistant Professor Lille 2 University
http://emmanuel.chazard.org

VI. Laboratory tests

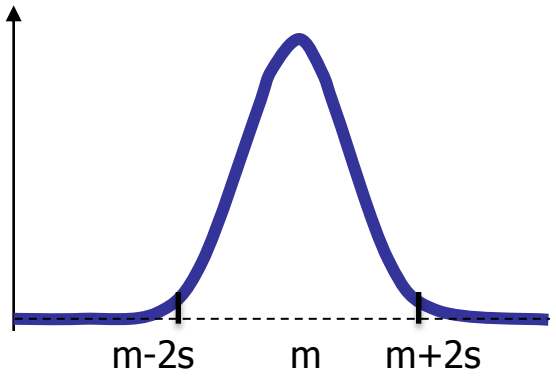
I. Distributions
II. Interval of normal values, alpha and beta risks
III. Effect of the population on the normal values
IV. Multiple testing

Dr Emmanuel Chazard
Assistant Professor Lille 2 University
http://emmanuel.chazard.org

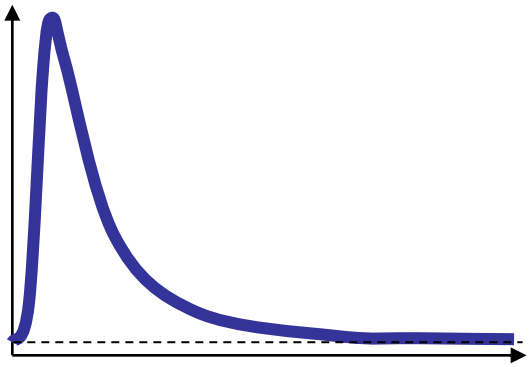
Biological parameters

- Dosages from various liquids or tissues (blood, urine, cerebrospinal fluid…)
- We interest on exams that output a quantitative response (most of them)
- Distribution of the values is known in healthy people:

Normal distribution (often)



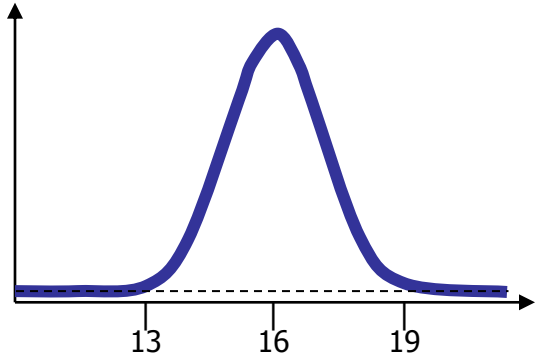
Sometimes lognormal distribution, "Galton's distribution"



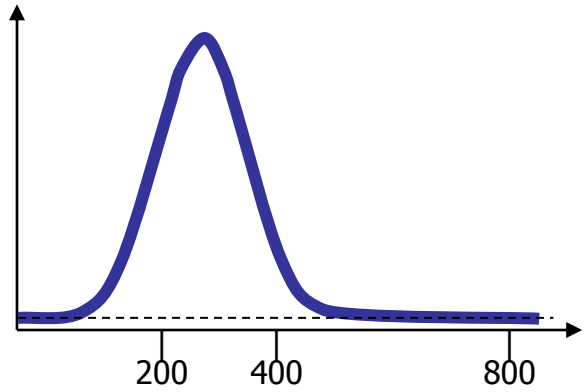
Examples of distribution in healthy people

Normal distributions:

Hemoglo-binemia in g/dl

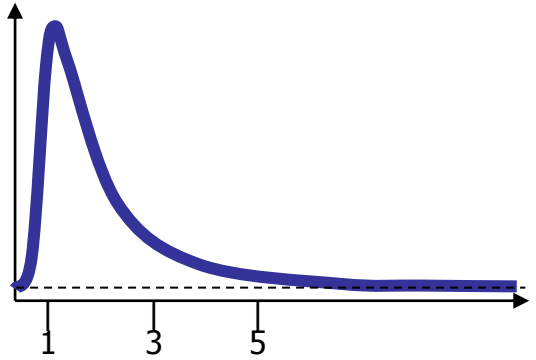


Platelet rate in 10^9/l

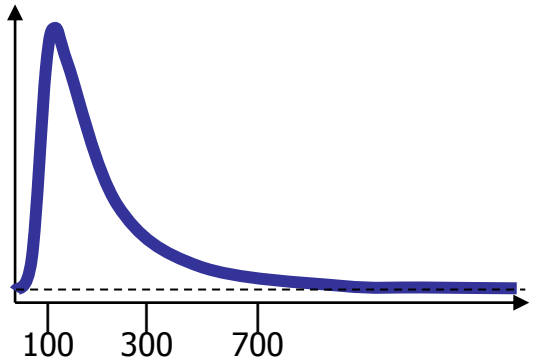


Lognormal distributions:

TSH plasmatic concentration in mUI/l



Ferritin plasmatic concentration in ng/ml



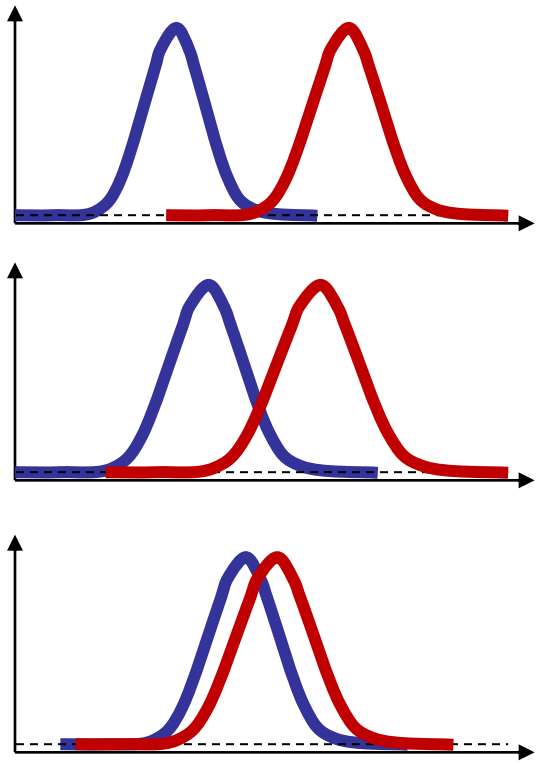
Dr Emmanuel Chazard
Assistant Professor Lille 2 University
http://emmanuel.chazard.org

How is the normality range defined?

- The normality range is defined by the interval containing 95% of the values of healthy people
  - Normal distribution: [μ-2ds ; μ+2ds]
  - Comparable for lognormal distribution
  - Other distributions: defined by the quantiles [$F^{-1}(0.025)$; $F^{-1}(0.975)$]
- Immediate consequence: 5% of healthy people have “abnormal” values!!
- In the next slides, we will assume that only one of the thresholds is used, to simplify things.

Usage of a biological parameter to detect ill people

Distribution of parameter in **healthy people** and **ill people**.



- Examples of situations:
- Ideal situation:
  - It is easy to find a threshold that discriminates healthy and ill people
- Moderate overlapping:
  - Acceptable misclassification
- Important overlapping:
  - Useless diagnostic test


Dr Emmanuel Chazard
Assistant Professor Lille 2 University
http://emmanuel.chazard.org

Alpha and beta risks of a diagnostic test

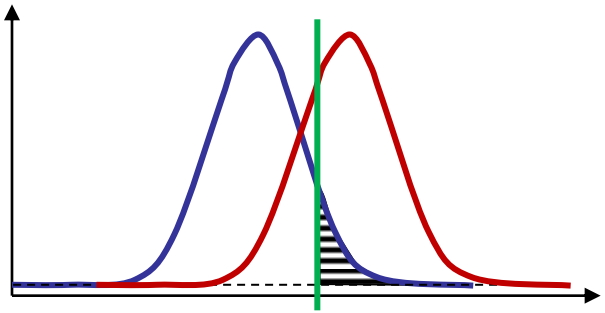
- Alpha risk (type 1 error): declaring that a patient is ill despite he is healthy = 1-Sp
- Beta risk (type 2 error): declaring that a patient is healthy despite he is ill = 1-Se
- Power = 1-β = Se : probability that a ill patient is declared ill

		Unknown realty	
		D+	D-
Decision with the test	T+ (value ∉ range)	No error	**α risk**
	T- (value ∈ range)	**β risk**	No error

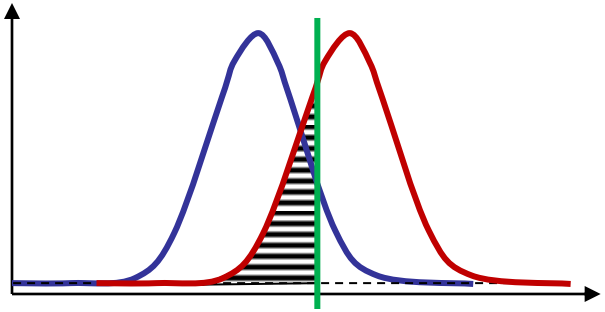
Alpha risk, beta risk, power

Distribution of parameter in **healthy people** and **ill people**. Representation of the **chosen threshold**.

Zone of the Alpha risk

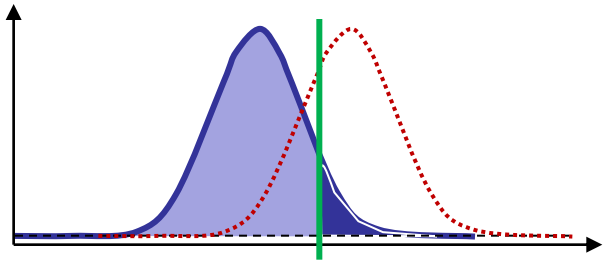


Zone of the beta risk



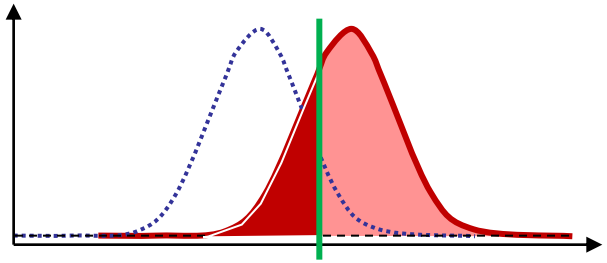
Beware, in this example ill people have higher values than healthy people.

Alpha risk (probability)



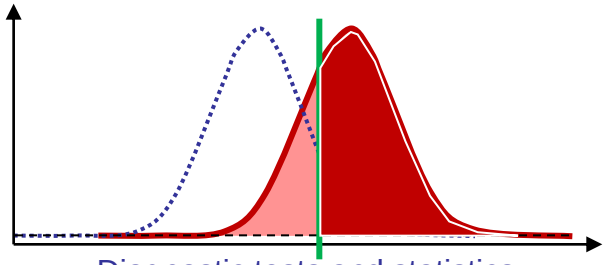
Probability that healthy people are declared ill =1-Sp

Beta risk (probability)



Probability that ill people are declared healthy =1-Se

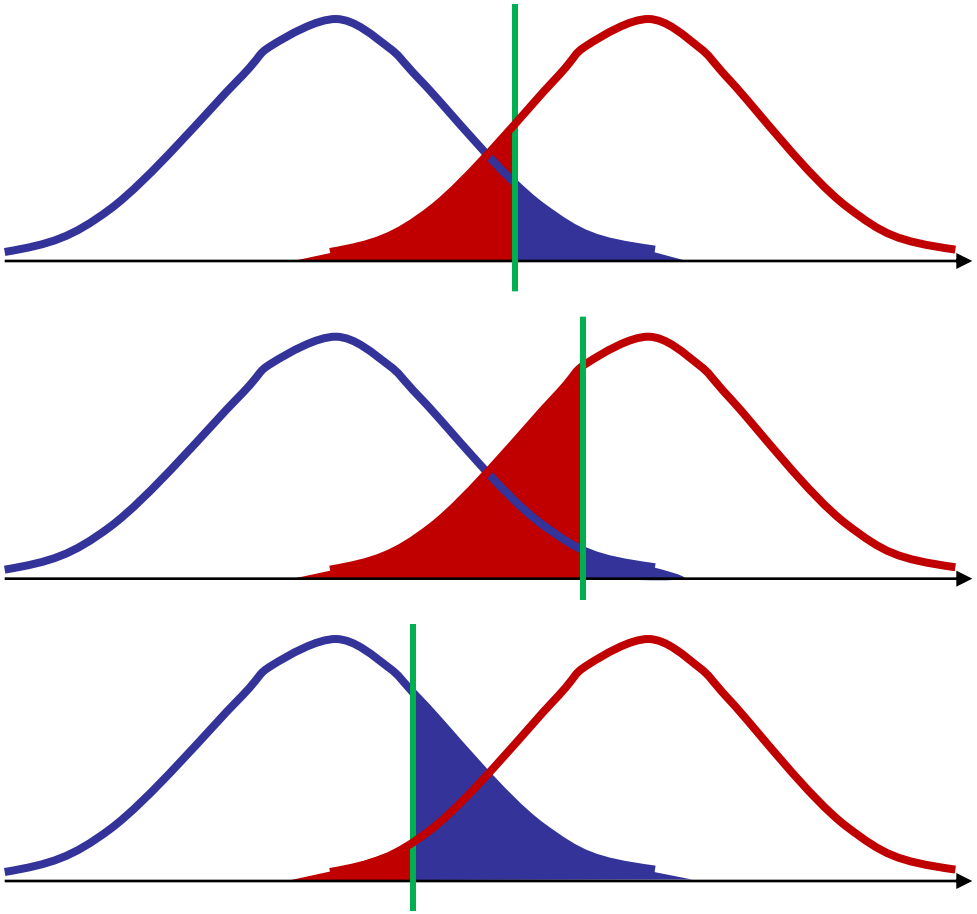
Power (probability)



Probability that ill patients are declared ill =Se


Dr Emmanuel Chazard
Assistant Professor Lille 2 University
http://emmanuel.chazard.org

Consequences of the thresholds on alpha and beta risks



In this example (because ill people have higher values) :
Increase of the threshold =>

- Decrease of alpha risk
- Increase of beta risk
- Decrease of power

Decrease of the threshold =>

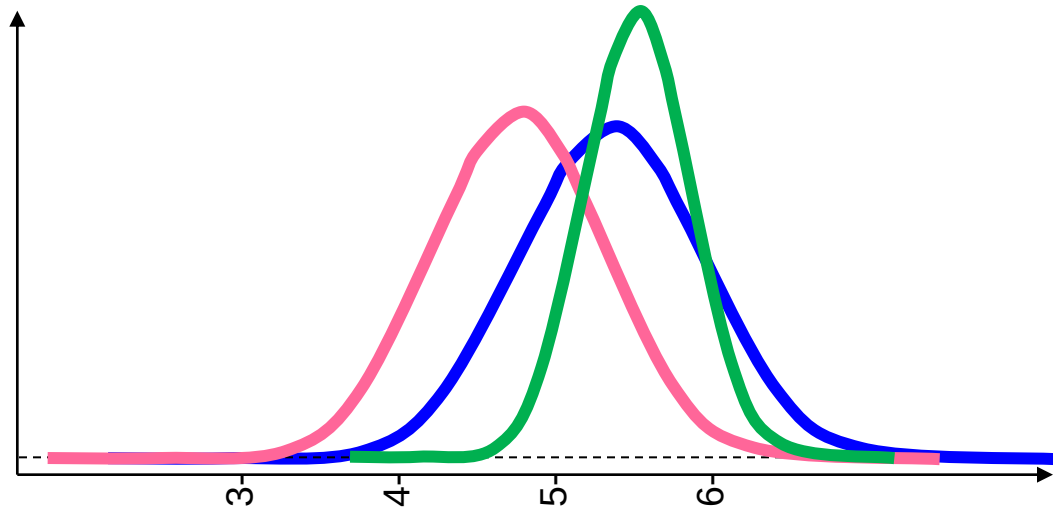
- Increase of alpha risk
- Decrease of beta risk
- Increase of power

Dr Emmanuel Chazard
Assistant Professor Lille 2 University
http://emmanuel.chazard.org

Some distributions of biological parameters may vary depending on the subpopulation

- Example: red cells blood concentration (en 10^6/µl)
- Variation of normal values among 3 populations
- The range used for diagnosis have to be adapted

Population	Lower bound	Upper bound
Male adult	4.5	6.2
Female adult	4	5.4
Newborn	5	6

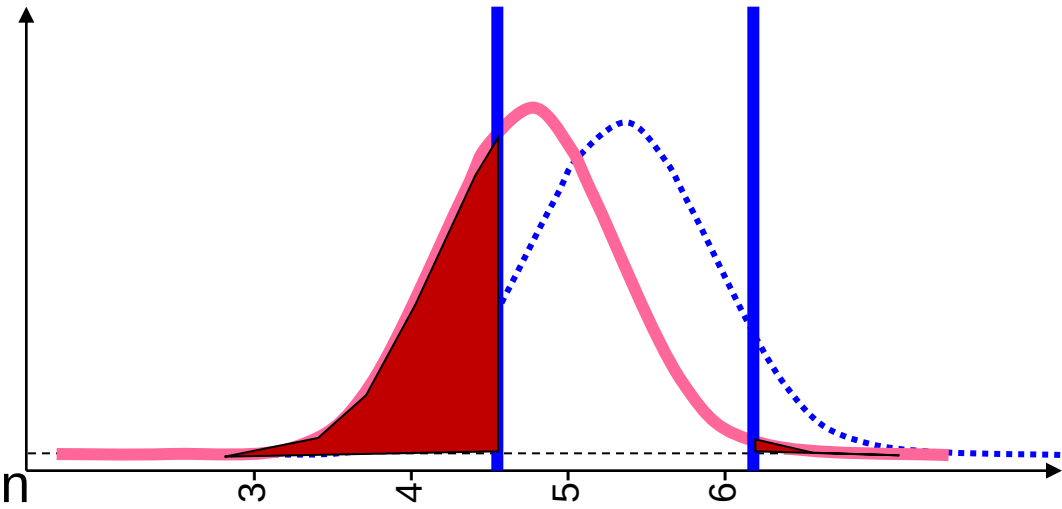


Dr Emmanuel Chazard
Assistant Professor Lille 2 University
http://emmanuel.chazard.org

Some distributions of biological parameters may vary depending on the subpopulation

- For example, if we use the boundaries of adult males for adult females, we obtain:
- For too high values detection: beta risk ↑, alpha risk ↓
- For too low values detection: alpha risk ↑, beta risk ↓

Population	Lower bound	Upper bound
Male adult	4.5	6.2
Female adult	4	5.4



- For some parameters, the boundaries are adapted depending on the subpopulation the patient belongs to.

Effect of multiple testing in healthy people

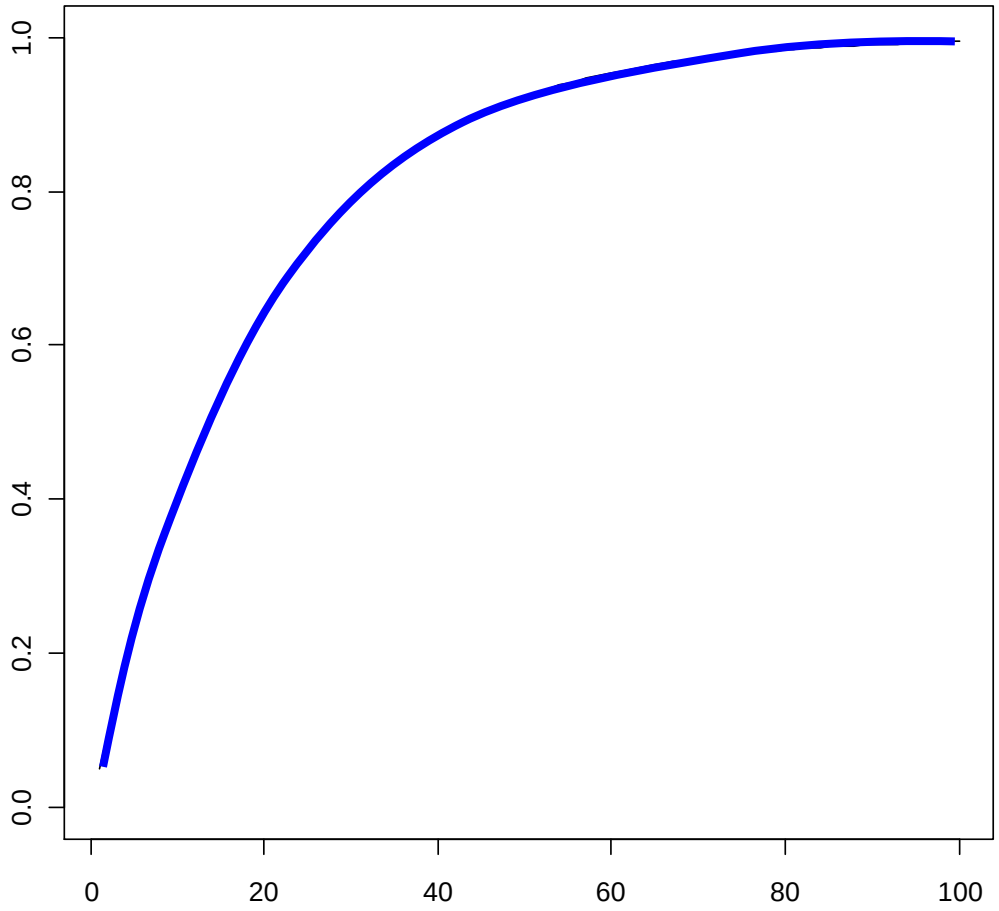
- With 1 test:
  - probability that healthy people are declared ill = alpha risk = 5%
- Scenario with several tests:
  - k independent laboratory tests are realized
  - The patient is declared ill if <u>at least one</u> of the k tests is positive (for each test, α_{indiv}=5%)
  - If the patient is healthy, what is the probability to declare him ill?
  - $\alpha_{total} = 1-(1- \alpha_{indiv})^k$
  - => <u>inflation of the α risk</u>.

Dr Emmanuel Chazard
Assistant Professor Lille 2 University
http://emmanuel.chazard.org

Effect of multiple testing in healthy people

- => inflation of the α risk.

Total alpha risk if every test is computed with a 5% individual alpha risk



Number of tests (0-100)

Dr Emmanuel Chazard
Assistant Professor Lille 2 University
http://emmanuel.chazard.org