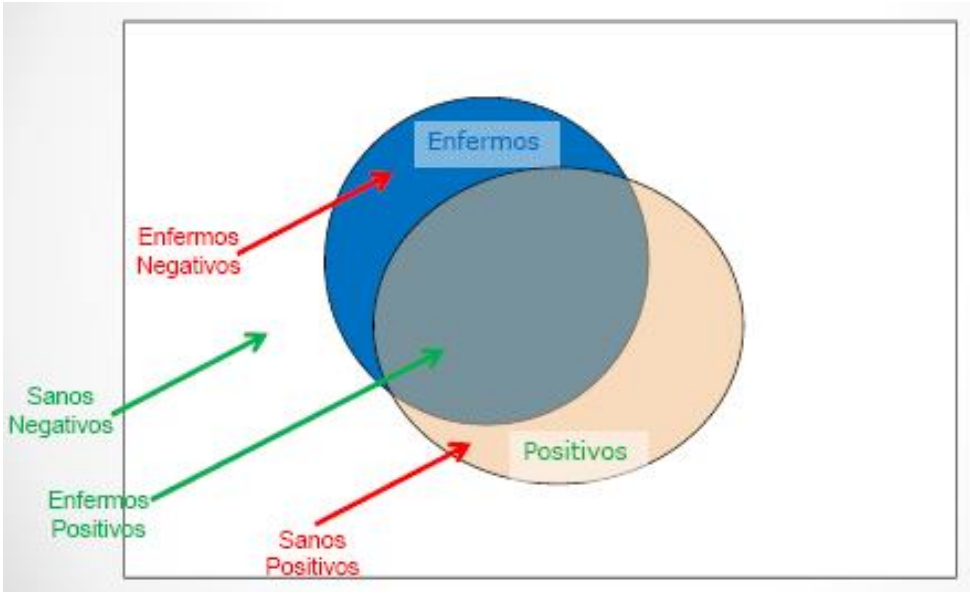
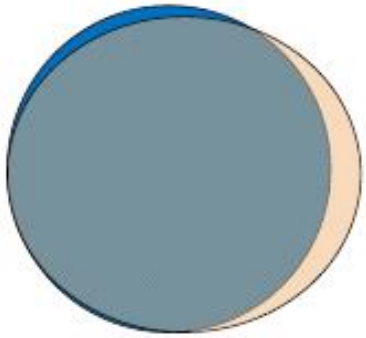
Biosciences: 3.- Health

3.6 Measuring Prediction: diagnosis and prognosis

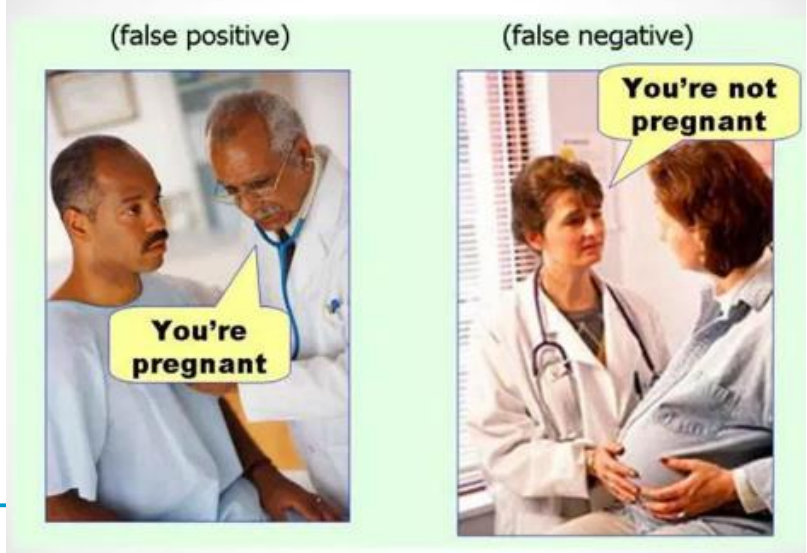
Ejemplos de posibles resultados de un test



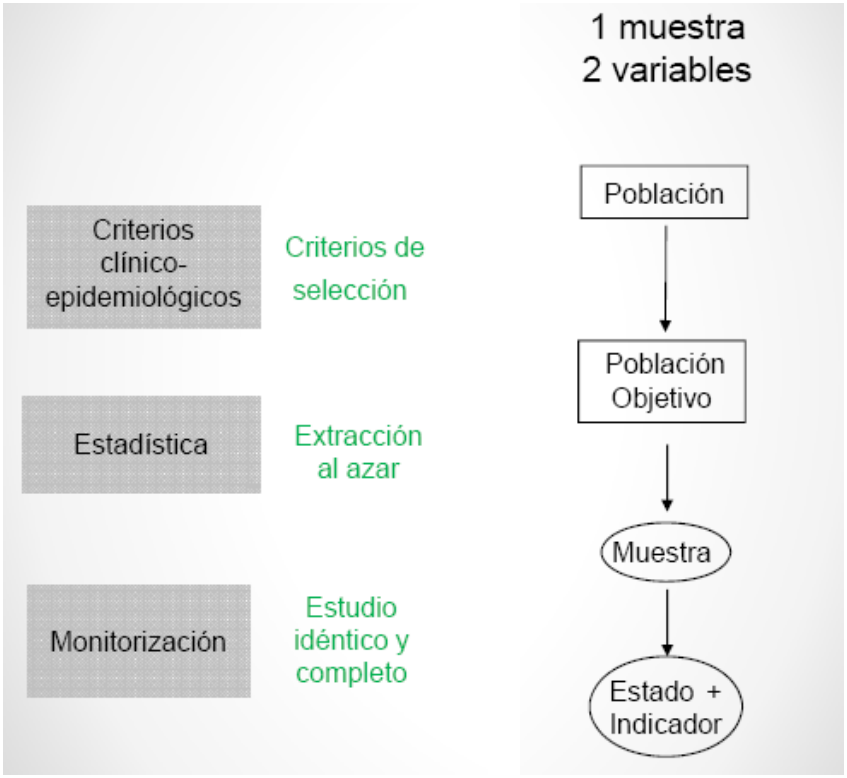
Buena clasificación



No tan buena...



Diseños para recoger la información



	+	-	
Enfermos	9	1	10
Sanos	27	63	90
	36	64	100

	+	-	
Enfermos	90	10	100
Sanos	30	70	100
	120	80	200

Let be: **+** = **Positive** result of a diagnostic test
- = **Negative** result of a diagnostic test
E = Real situation of the patient: Sick
S = Real situation of the patient : Healthy

And the following *non-formal / formal* definitions:

Sensitivity: "*Trend*" from the sick patients to be positive → **P(+|E)**

Specificity: "*Trend*" from the healthy individuals to be negative → **P(-|S)**

Positive predictive value VP+: "*Trust*" *in a positive result* → **P(E|+)**

Negative predictive value VP-: "*Trust*" *in a negative result* → **P(S|-)**

Note that the first 2 are more intuitive, since they go from the cause (E, S) to the consequence (+, -), but the last 2 are more valuable to the clinician.

Ejemplo

1 muestra
2 variables

	+	–	
E=Enfermos	9	1	10
S=Sanos	27	63	90
	36	64	100

$$\text{Sensibilidad} = P(+|E) = \frac{9}{10} = 0.9 = 90\%$$

$$\text{Especificidad} = P(-|S) = \frac{63}{90} = 0.7 = 70\%$$

1 muestra
2 variables

	+	–	
E=Enfermos	9	1	10
S=Sanos	27	63	90
	36	64	100

$$VP+ = P(E|+) = \frac{9}{36} = 0.25 = 25\%$$

$$VP- = P(S|-) = \frac{63}{64} \approx 0.98 = 98\%$$

Exercise

In a certain population, information about the actual state of the patient (sick or healthy), and about the diagnostic test result (+ or -) was collected.

From the results of the table, calculate sensitivity, specificity and predictive values.

	+	-	
Sick	94	38	132
Healthy	215	653	868
	309	691	1000

Sensitivity: **P (+|E)** =
Specificity: **P (- |S)** =

Positive predictive value : **VP+**: **P (E|+)** =
Negative predictive value : **VP-**: **P (S|-)** =

Solution

	+	-	
Sick	94	38	132
Healthy	215	653	868
	309	691	1000

Sensitivity: **P (+|E)** = 94/132 ≈ 0.712 ≈ 71%

Specificity: **P (-|S)** = 653/868 ≈ 0.752 ≈ 75%

Positive predictive value : **VP+**: **P (E|+)** = 94/309 ≈ 0.304 ≈ 30%

Negative predictive value : **VP-**: **P (S|-)** = 653/691≈ 0.945 ≈ 95%

$$0.712 = \mathbf{P(+|E)} \neq \mathbf{P(E|+)} = 0.304$$

Now, assume **that two samples** were obtained, one for sick and one for healthy patients:

	+	-	
Sick	712	288	**1000**
Healthy	248	752	**1000**
	960	1040	2000

Sensitivity: **P (+|E)** = 712/100 = 0.712 ≈ 71%

Specificity: **P (-|S)** = 752/1000 = 0.752 ≈ 75%

Positive predictive value: VP+: **P (E|+)** = 712/960 ≈ 0.742 ≈ 74%

Negative predictive value: VP-: **P (S|-)** = 752/1040≈ 0.723 ≈ 72%

But **be aware**: here, those VPs are meaningless, they depend on an arbitrarily chosen number of sick and healthy recruited people.

Diferencias según diseños

1 muestra y 2 variables

	+	−	
E	9	1	10
S	27	63	90
	36	64	100

2 muestras y 1 variable

	+	−	
E	90	10	100
S	30	70	100
	120	80	200

$\text{Sens} = P(+|E) = \frac{9}{10} = 90\%$ = $\text{Sens} = P(+|E) = \frac{90}{100} = 90\%$

$\text{Esp} = P(-|S) = \frac{63}{90} = 70\%$ = $\text{Esp} = P(-|S) = \frac{70}{100} = 70\%$

$VP+ = P(E|+) = \frac{9}{36} = 25\%$ ≠ $VP+ = P(E|+) = \frac{90}{120} = 75\%$

$VP- = P(S|-) = \frac{63}{64} \approx 98\%$ ≠ $VP- = P(S|-) = \frac{70}{80} \approx 87.5\%$

Obtención de los valores predictivos (*VP*):

Bayes' Theorem

From the definition of conditional probability, we have

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \rightarrow P(A \cap B) = P(A \mid B) \cdot P(B).$$

Hence,

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A \mid B) \cdot P(B)}{P(A)}.$$

$$P(E \mid +) = \frac{P(+\cap E)}{P(+)} = \frac{P(+|E) \cdot P(E)}{P(+)} = \frac{P(+|E) \cdot P(E)}{P[(+\cap E) \cup (+\cap S)]} = \frac{P(+|E) \cdot P(E)}{P(+|E) \cdot P(E) + P(+|S) \cdot P(S)}$$

Analogously, we deduce P (S | -).

Exercise

Calculate the **PV** assuming **two samples** were obtained, one for sick and one for healthy patients, and knowing that in the population **P(E) = 0.132**.

	+	-	
Sick	712	288	1000
Healthy	248	752	1000
	960	1040	2000

We can calculate:

Sensitivity: **P (+|E)** = 712/100 = 0.712 ≈ 71%,
Specificity: **P (-|S)** = 752/1000 = 0.752 ≈ 75%,

and we know the prevalence P(E) = 0.132. We need to find:

Positive predictive value: VP+: **P (E|+)**

Negative predictive value: VP-: **P (S|-)**

Solution

$$P(E \mid +) = \frac{P(+|E) \cdot P(E)}{P(+|E)\,P(E)+P(+|S)\,P(S)} = \frac{\text{Sensitivity} \cdot \text{Prevalence}}{\text{Sens} \cdot \text{Prev} + (1\text{- Specificity}) \cdot (1 - \text{Prev})} \approx$$

$$\approx \frac{0.712 \cdot 0.132}{0.712 \cdot 0.132 + 0.248 \cdot 0.868} \approx 0.304$$

$$P(S \mid -) = \frac{P(-|S) \cdot P(S)}{P(-|S)\,P(S)+P(-|E)\,P(E)} = \frac{\text{Specificity} \cdot (1 - \text{Prevalence})}{\text{Spec} \cdot (1\text{-Prev}) + (1 - \text{Sensitivity}) \cdot \text{Prev}} \approx$$

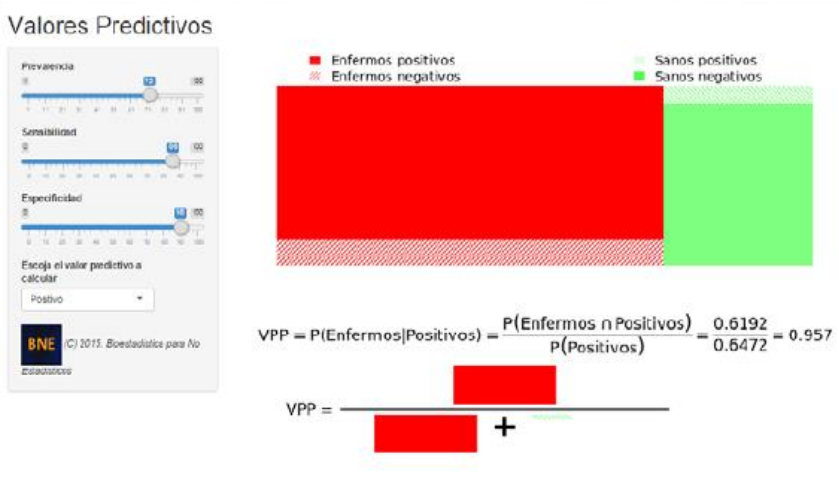
$$\approx \frac{0.752 \cdot 0.868}{0.752 \cdot 0.868 + 0.288 \cdot 0.132} \approx 0.9449$$

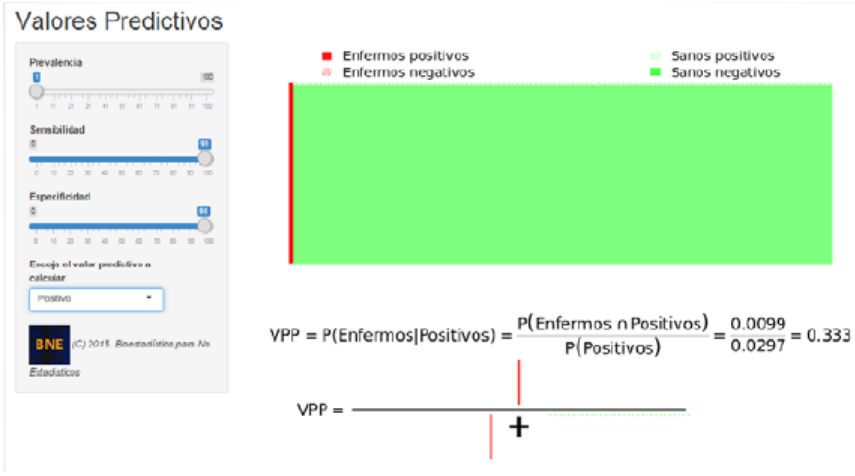
Initial correct values are achieved:

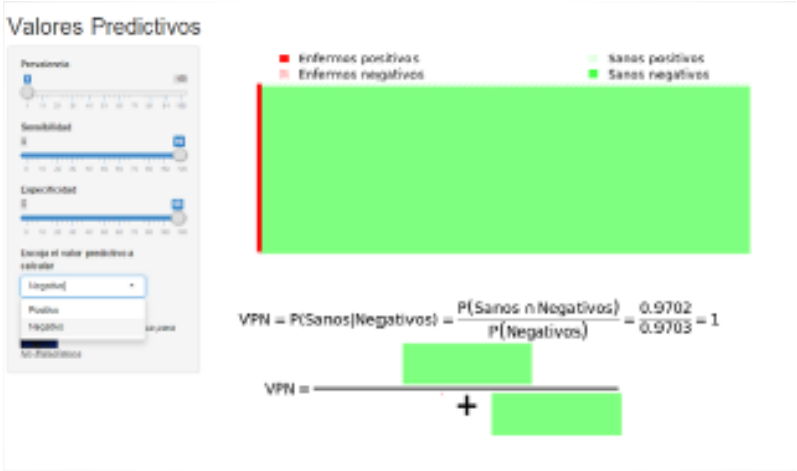
This formula allows to merge correct estimations from the current data, such as P(+|E) or P(-|S), with previous known values, such as P(S) or P(E).

Aplicación shiny

http://shiny-eio.upc.edu/bne/VPs/







Notad que cuando la prevalencia es muy baja, aunque la sensibilidad y la especificidad sean altas, el valor predictivo positivo es muy bajo.

Ej. Para un cáncer poco frecuente, un valor positivo en el test aporta poca información. En cambio, un valor negativo nos dice mucho (el paciente estará realmente sano con una probabilidad muy alta).