

# Measuring test accuracy

Mark A. Zayzman<sup>1</sup>

<sup>1</sup>Department of Pathology and Immunology, Division of Laboratory and Genomic medicine,  
Washington University in St. Louis, School of Medicine, St. Louis, MO

## Table of contents

- Introduction
- Learning objectives
- Why do we perform diagnostic testing?
- Developing an intuitive sense of test accuracy
  - The case of an ideal test
  - The case of a useless test
  - The case of a realistic test
- Quantitative measures of analytical accuracy
  - Sensitivity, specificity, and ROC curves
  - Positive and negative predictive values
  - Likelihood ratios and diagnostic odds ratio
    - Likelihood ratios relate pre-test and post-test odds given a specific result
    - Diagnostic odds ratio relates post-test odds of the disease for individuals with a negative versus positive test result
- Glossary

## Introduction

A central task of a clinical laboratorian is to critically evaluate laboratory assays in order to curate a menu of tests that will positively impact patient care and that represent a wise investment of limited healthcare resources. So how can we quantify the 'goodness' of a test? This lecture focuses on quantitative metrics of analytical accuracy, a core performance characteristic reflecting how well a test result can discriminate between healthy and diseased individuals. This handout is designed to serve as a companion during the lecture presentation as well as a useful resource to refer back to outside of this course.

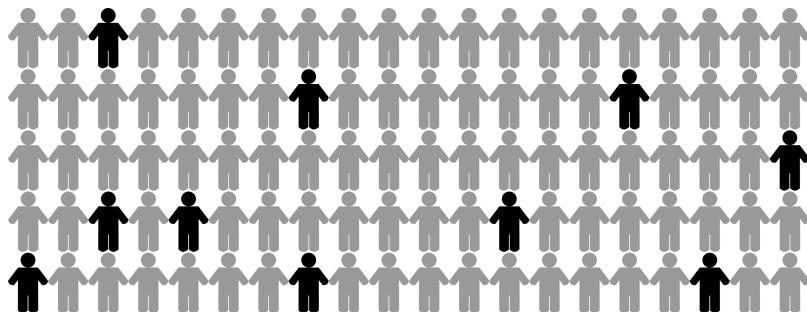
# Learning objectives

- Define common metrics that are used to benchmark test accuracy
- Apply accuracy-based metrics to compare two laboratory tests

[Back to table of contents](#)

## Why do we perform diagnostic testing?

Consider a hypothetical group of patients with clinical attributes that are concerning for a disease. The disease is associated with significant morbidity and/or mortality, but an effective treatment exists that can alter the natural disease course. Unfortunately, the treatment is expensive and causes undesirable side effects. Approximately 10 out of every 100 individuals in the group is afflicted by the disease, that is the disease prevalence is 10% (**Figure 1**). Without additional information, we are faced with a challenging clinical dilemma because we are uncertain as to which specific patients should or should not receive treatment. The aim of diagnostic laboratory testing is to help reduce our uncertainty regarding which patients should be offered treatment.



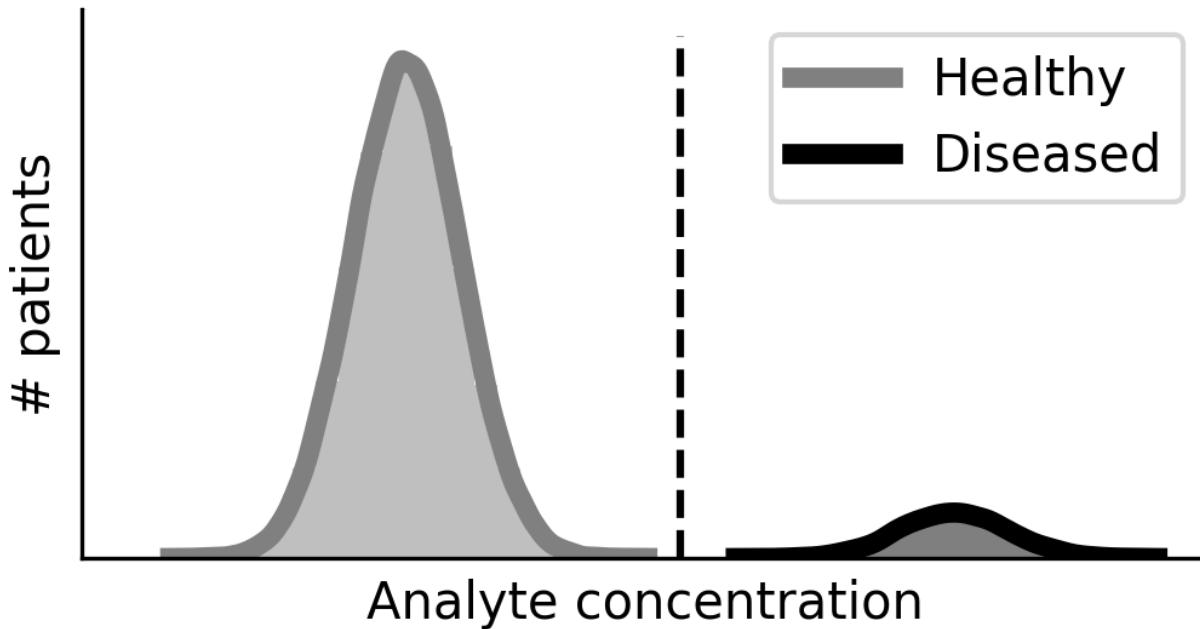
**Figure 1:** A hypothetical patient population with a disease prevalence of 10%. Gray and black colors indicate healthy and diseased individuals respectively.

## Developing an intuitive sense of test accuracy

A typical diagnostic testing process involves the quantification of some bioanalyte in a specimen obtained from a patient. Ideally, the bioanalyte concentration reflects the pathobiology underlying the disease and the analyte concentrations observed in diseased individuals tends to be different from those observed in healthy individuals. As will be discussed below, the degree to which the distributions of analyte concentrations observed amongst healthy and diseased individuals diverge is a critical determinant of test accuracy.

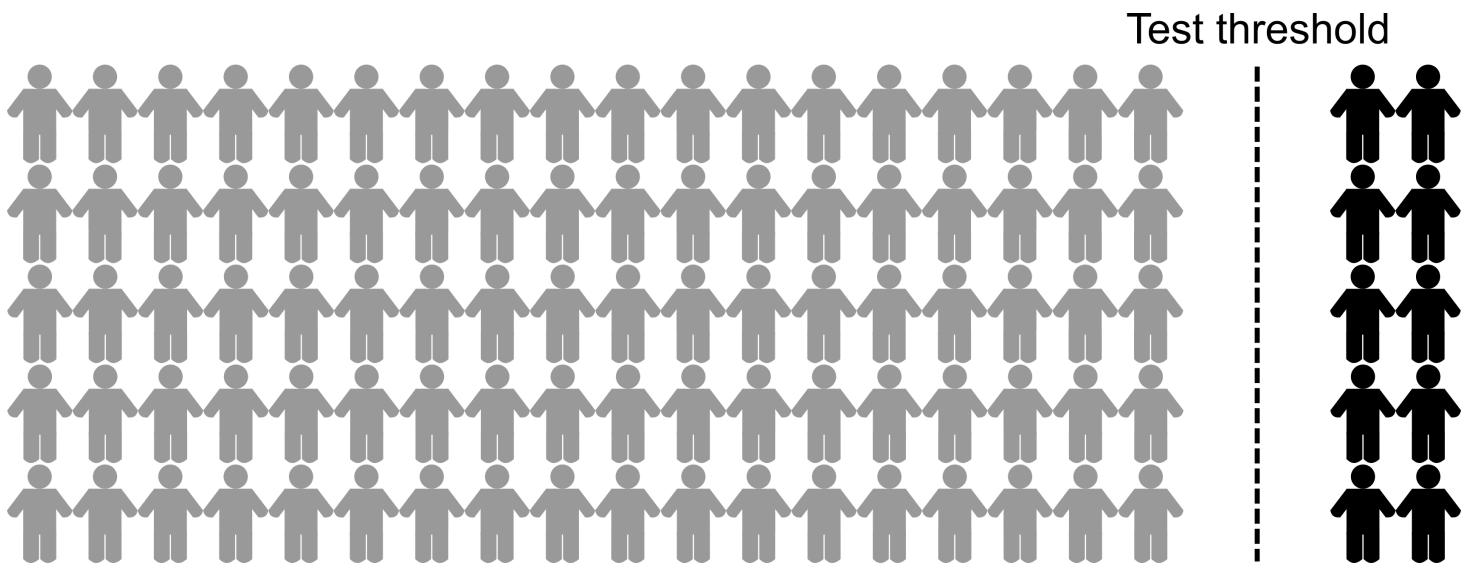
## The case of an ideal test

In the case of an ideal test, the distribution of analyte concentrations would be observed for diseased individuals would be completely distinct (not overlapping) from that observed for healthy individuals (**Figure 2**). If this is the case then we can define a 'decision boundary' (aka test threshold) (vertical dashed line) to divide the space of analyte concentrations into two regions: the negative region to the left of the boundary and the positive region to the right of the boundary. Note that all the healthy individuals in the population will have an analyte concentrations fall into the negative region (i.e. to the left) and all the diseased individuals have analyte concentrations that fall in the positive regions (i.e. to the right).



**Figure 2:** An ideal diagnostic test. The distributions of analyte concentrations for healthy (gray) and diseased (black) individuals are shown. Dashed line indicates a decision boundary (test threshold) that separates all healthy and diseased individuals.

If we apply this ideal test to the hypothetical patient population in Figure 1 all the healthy individuals will have a result less than the test threshold (i.e. test negative) and all the diseased individuals would have a result greater than the test threshold (i.e. test positive) (**Figure 3**). Thus, if we assign patients to treatment based upon their test result we will appropriately treat all of the diseased individuals and avoid inappropriately treating even a single healthy individuals.

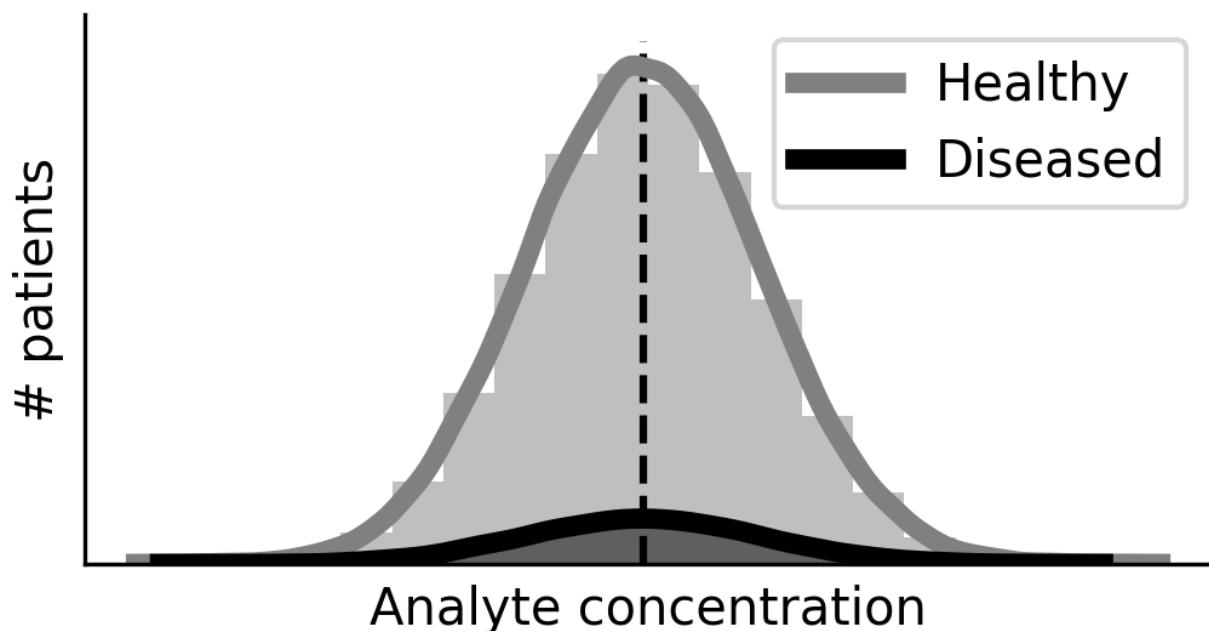


**Figure 3:** Application of the ideal test depicted in Figure 2 to the patient population in depicted in Figure 1 identifies all disease

Note: In real life, the only time we observe an 'ideal test' is when the test itself is used to define the disease. In this case we would say that the test is the 'gold standard'. If the gold standard were to be reassigned to a different diagnostic criteria we would likely find that the test no longer perfectly separates healthy and diseased patients.

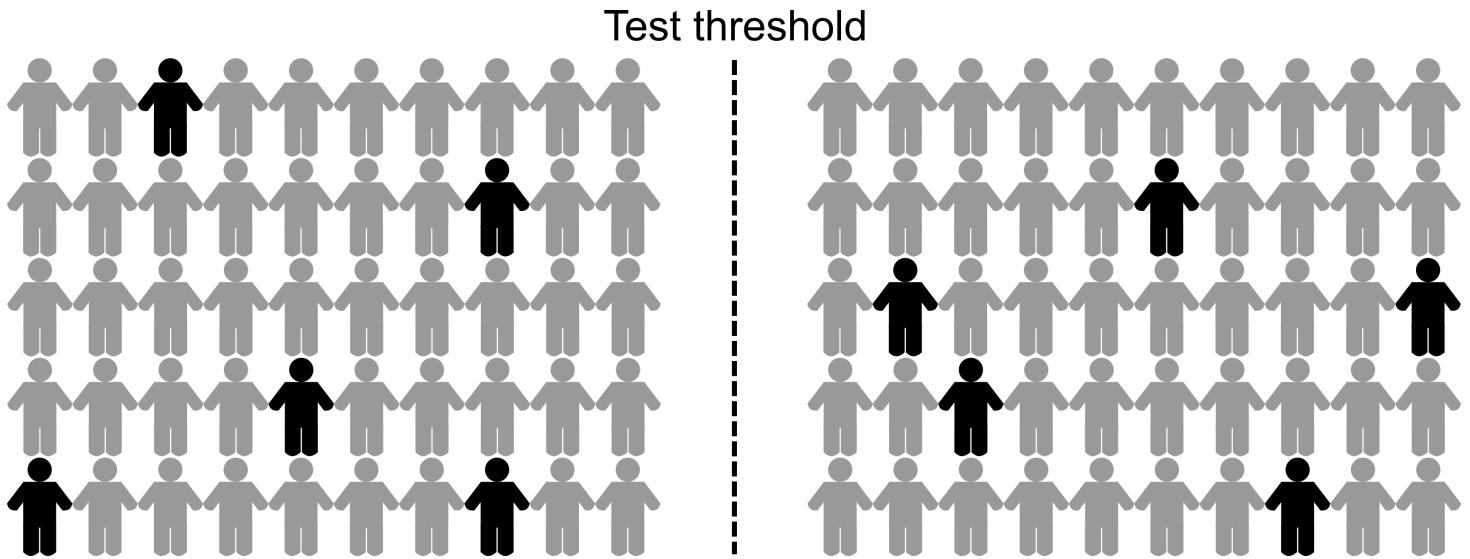
### The case of a useless test

Lets consider the other edge case - a useless test. In this case we have chosen to quantify an analyte that has no association with the disease process and the distributions of analyte concentrations measured for healthy individuals sits right on top of that for diseased individuals (**Figure 4**).



**Figure 4:** For a useless test there is complete overlap in the distributions of analyte values for healthy (gray) and diseased (black) individuals.

For all possible threshold we might consider the fraction of the patients with a test value above or below the threshold that have the disease will be expected to be close to the disease prevalence in the original population (**Figure 5**).

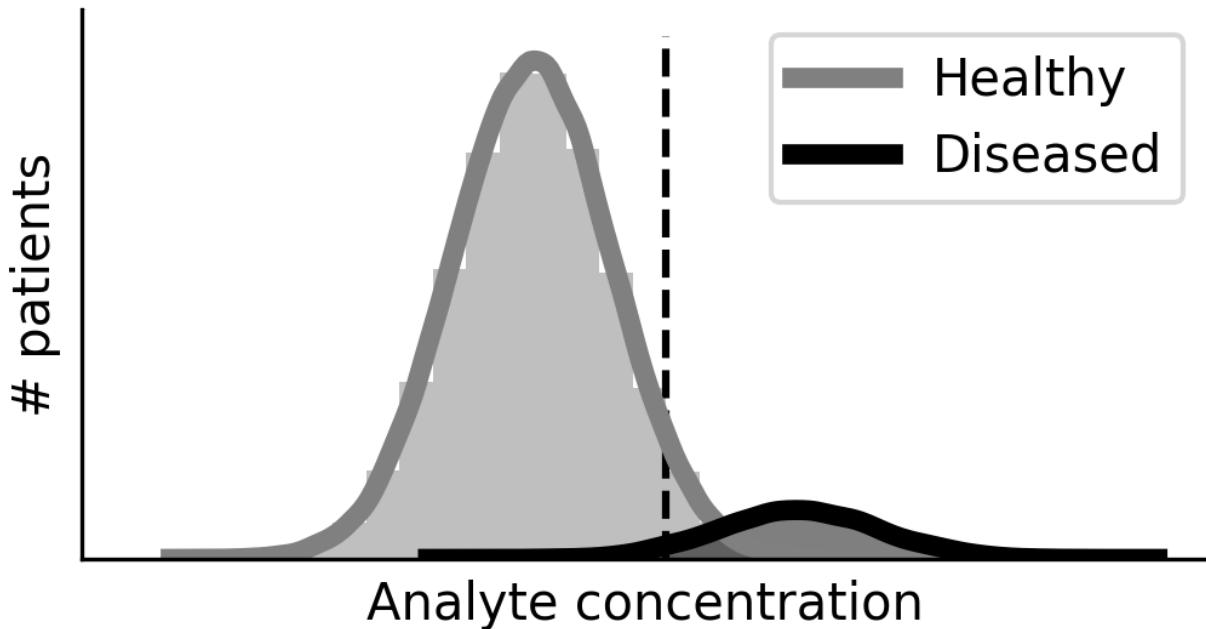


**Figure 5:** Application of the useless test depicted in Figure 4 to the patient population in depicted in Figure 1.

Therefore, if we base our treatment decisions on the results of this useless test we would, on average, do no better than if we had randomly assigned a subset of the patients to treatment.

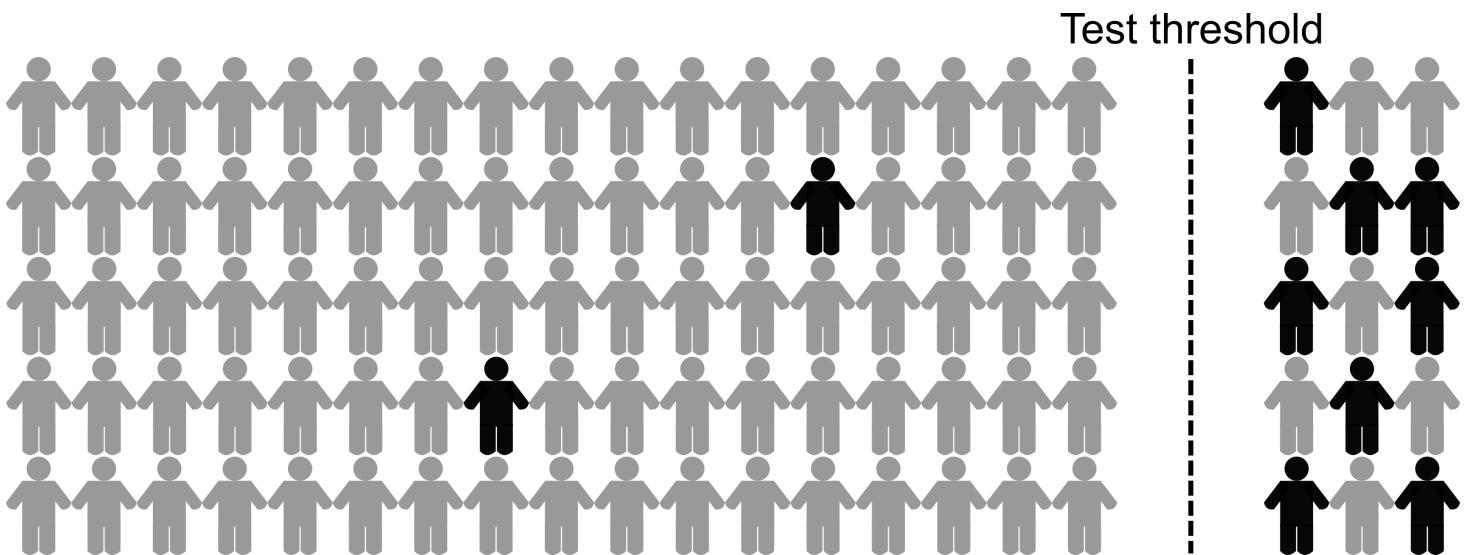
## The case of a realistic test

The case of a realistic test is intermediate to those of the ideal and useless tests with the distributions of analyte concentrations observed for healthy and diseased individuals overlapping partially (**Figure 6**)



**Figure 6:** For a realistic test there is some overlap in the distributions of analyte values for healthy (gray) and diseased (black) individuals.

In this case, there is no decision boundary that will perfectly separate results from healthy and diseased individuals. However, there are many possible boundaries that will define negative and positive regions that, compared to the background prevalence, are enriched for results from healthy and diseased individuals respectively. For example, in the illustration of **Figure 7** we see that 83/85 (97.6%) of the results below the threshold come from healthy individuals and 8/15 (53.3%) of the results above the threshold come from diseased individuals. By random assortment we would have expected these values to be 90% and 10%, respectively. Thus, if we assign patients to treatment based on having a positive result using this realistic test we will end up erroneously treating some healthy individuals and not treating some diseased individuals. Nevertheless, we will on average allocate the treatment to those who need it better than if we had employed a random selection process.



**Figure 7:** Application of the realistic test depicted in Figure 6 to the patient population in depicted in Figure 1.

The rest of this handout will deal with metrics that we use quantify test accuracy. Abstractly, we can think of accuracy as the position of the test along a spectrum ranging from the case of the useless test on one end to the ideal test on the other end. More concretely, metrics of accuracy help to describe how well a decision boundary can be chosen to separate the test results from healthy and diseased individuals.

## Quantitative measures of analytical accuracy

### Sensitivity, specificity, and ROC curves

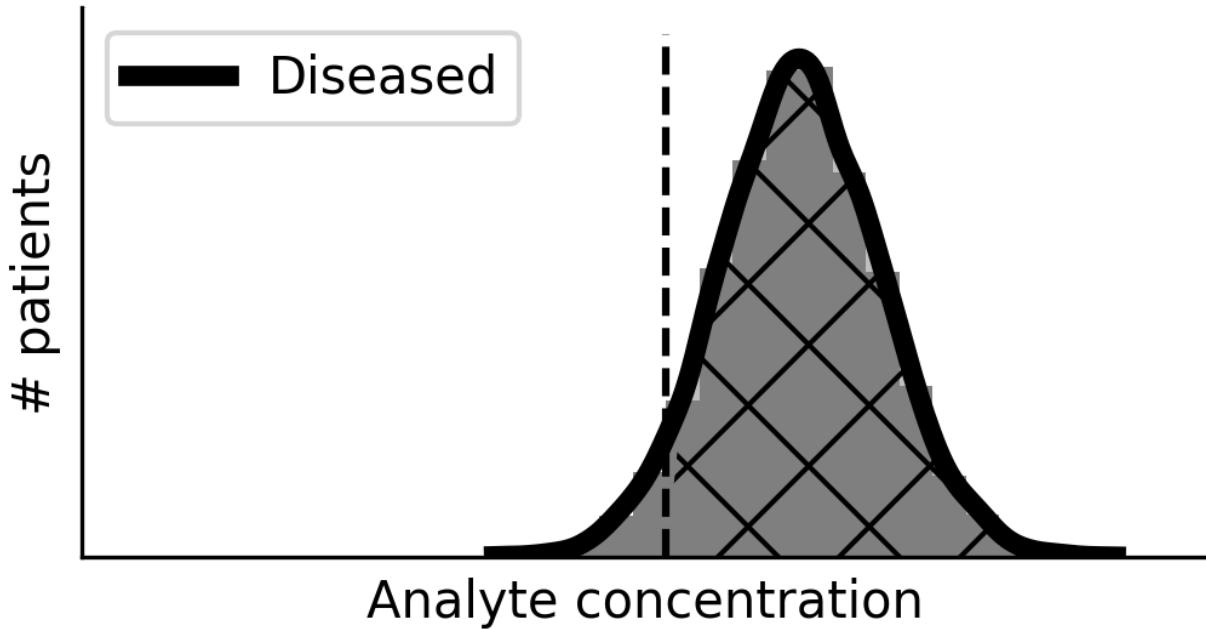
#### Sensitivity

The sensitivity ( $Se$ ) of a test is calculated as the fraction of the diseased individuals that are classified as positive by the test, i.e. that have a result above the threshold value.

$$Se = \frac{TP}{TP + FN}$$

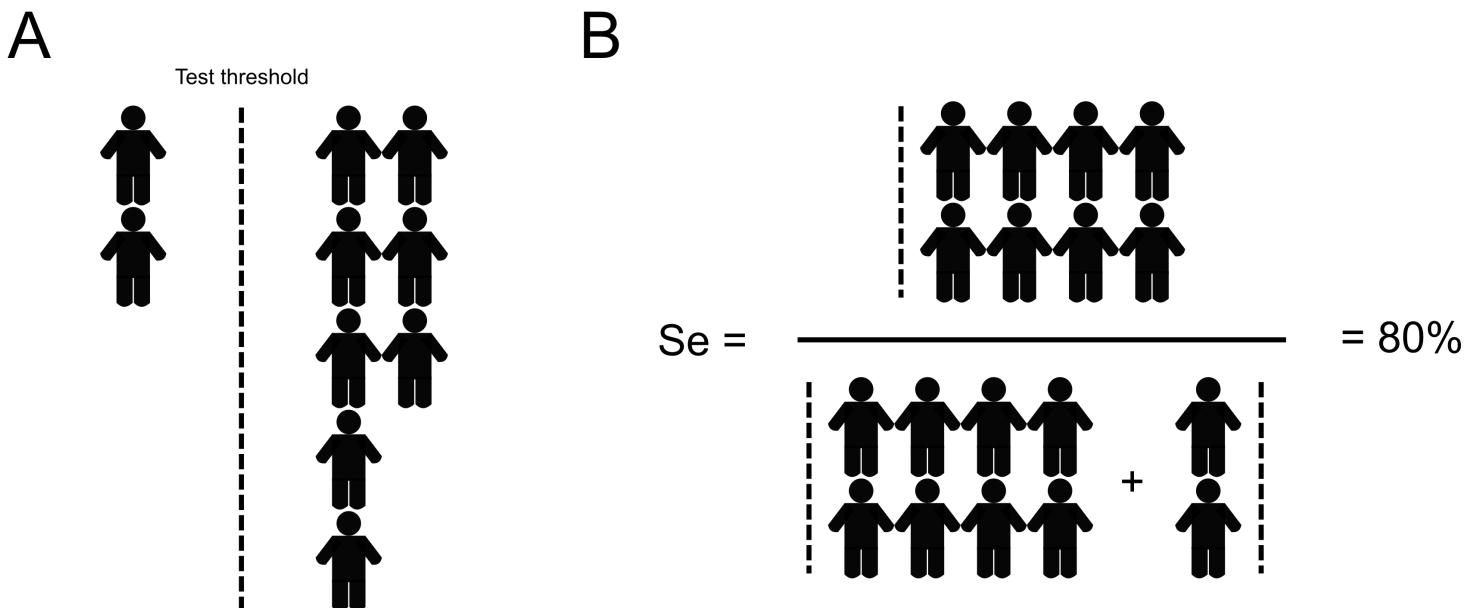
where  $TP$  is the number of diseased individuals classified as positive ('true positives'), and  $FN$  is the number of diseased individuals classified as negative ('false negatives').

Mathematically,  $Se$  is the fraction of the total area under the distribution of analyte concentrations observed in diseased individuals that falls to the right of our threshold value (**Figure 8**).  $Se$  decreases to zero as the threshold is increased and increases to one as the threshold is decreased.



**Figure 8:** Graphical representation of  $Se$ .

Figure 9 illustrates how to calculate  $Se$  for the realistic test depicted in Figures 6 and 7. Note that for the chosen diagnostic threshold, 2 and 8 diseased individuals would be classified as negative and positive, respectively (**Figure 9A**). Plugging into our formula we find that  $Se = \frac{8}{8+2} = \frac{8}{10} = 80\%$  (**Figure 9B**).



**Figure 9:** Illustration of sensitivity ( $Se$ ) calculation for the realistic test depicted in Figures 6 and 7. (A) Test characteristics of diseased patients. (B) Calculation of  $Se$ .

Key points:

1.  $Se$  is the fraction of diseased individuals classified as positive by the test
2.  $Se$  varies from zero to one
3.  $Se$  decreases with increasing threshold value

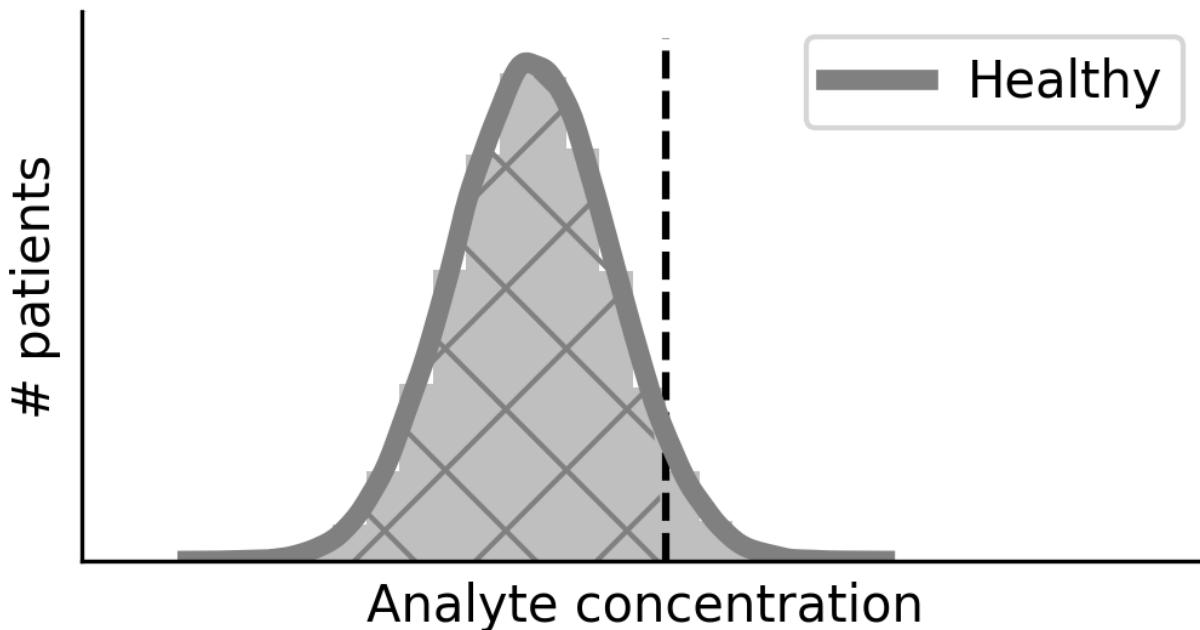
## Specificity

The specificity ( $Sp$ ) of a test is calculated as the fraction of the healthy individuals that are classified as negative by the test, i.e. that have a result below the threshold value.

$$Sp = \frac{TN}{TN + FP}$$

where  $TN$  is the number of healthy individuals classified as negative ('true negative'), and  $FP$  is the number of healthy individuals classified as positive ('false positives').

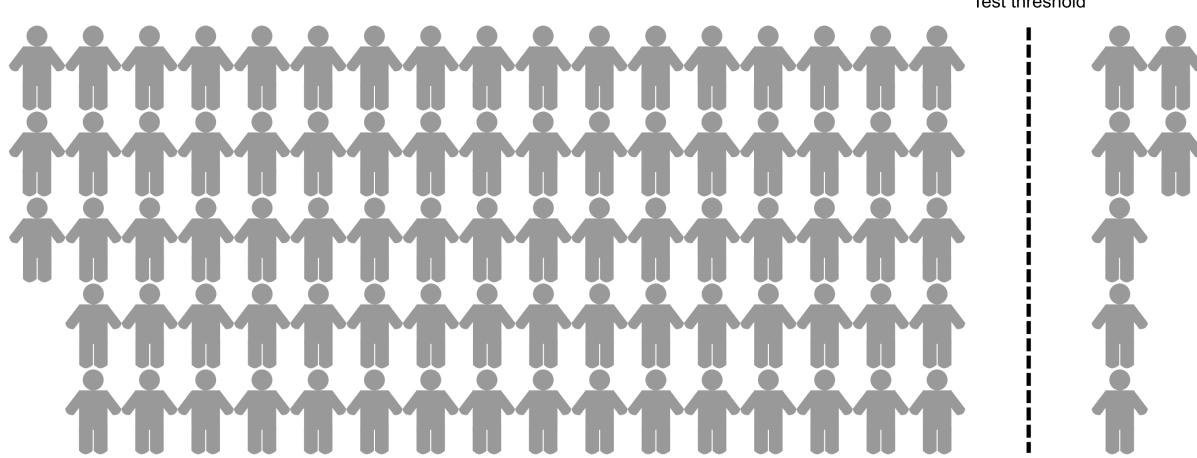
Mathematically,  $Sp$  is the fraction of the total area under the distribution of analyte concentrations observed in healthy individuals that falls to the left of our threshold value (**Figure 10**).  $Sp$  increases to one as the threshold is increased and decreases to zero as the threshold is decreased.



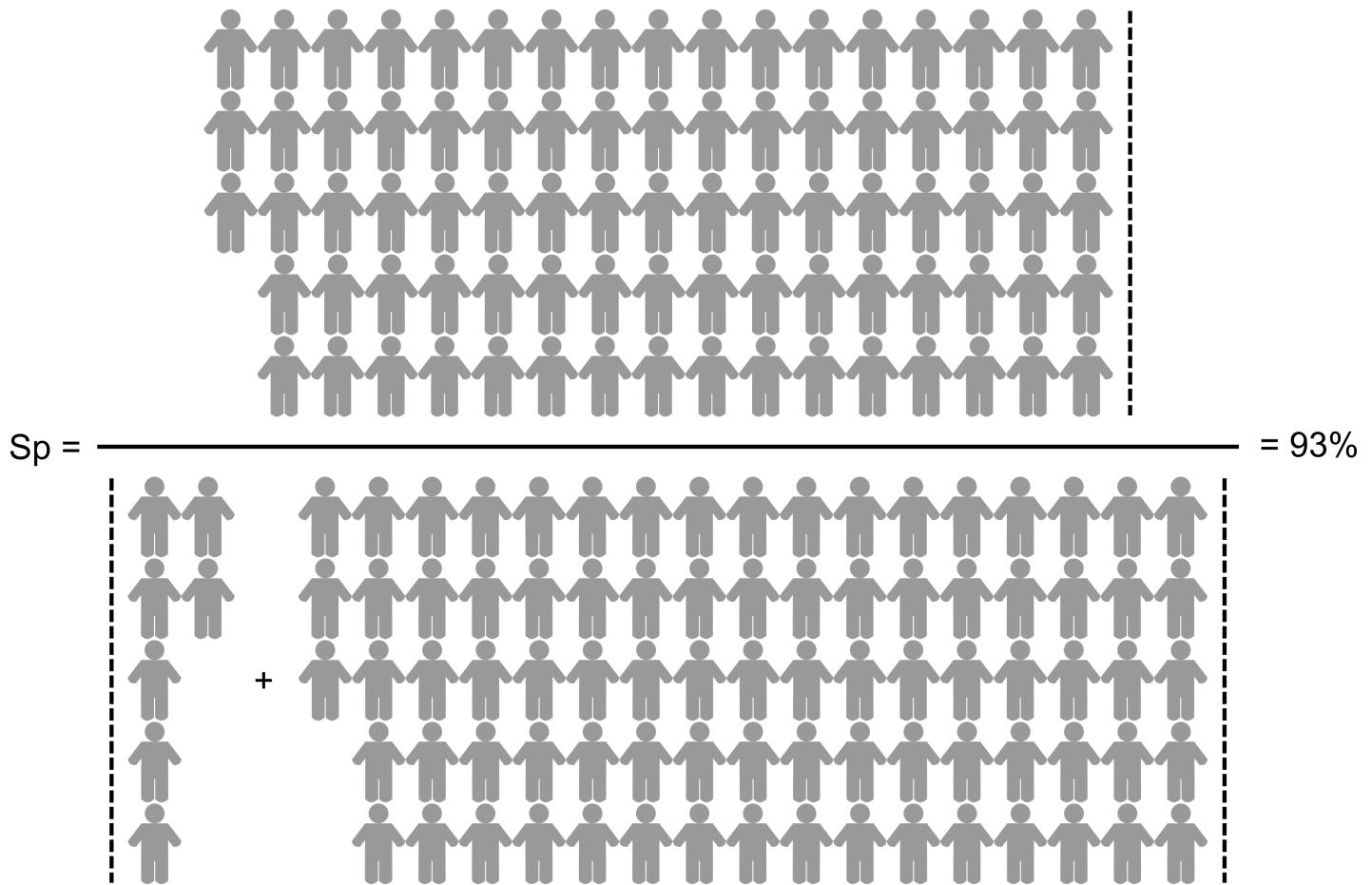
**Figure 10:** Graphical representation of  $Sp$ .

Figure 11 illustrates how to calculate  $Se$  for the realistic test depicted in Figures 6 and 7. Note that for the chosen diagnostic threshold, 83 and 7 healthy individuals would be classified as negative and positive, respectively (**Figure 11A**). Plugging into our formula we find that  $Sp = \frac{83}{83+7} = \frac{83}{90} = 92.2\%$  (**Figure 11B**).

A



B



**Figure 11:** Illustration of Specificity (Sp) calculation for the realistic test depicted in Figures 6 and 7.  
(A) Test characteristics of healthy patients. (B) Calculation of Sp.

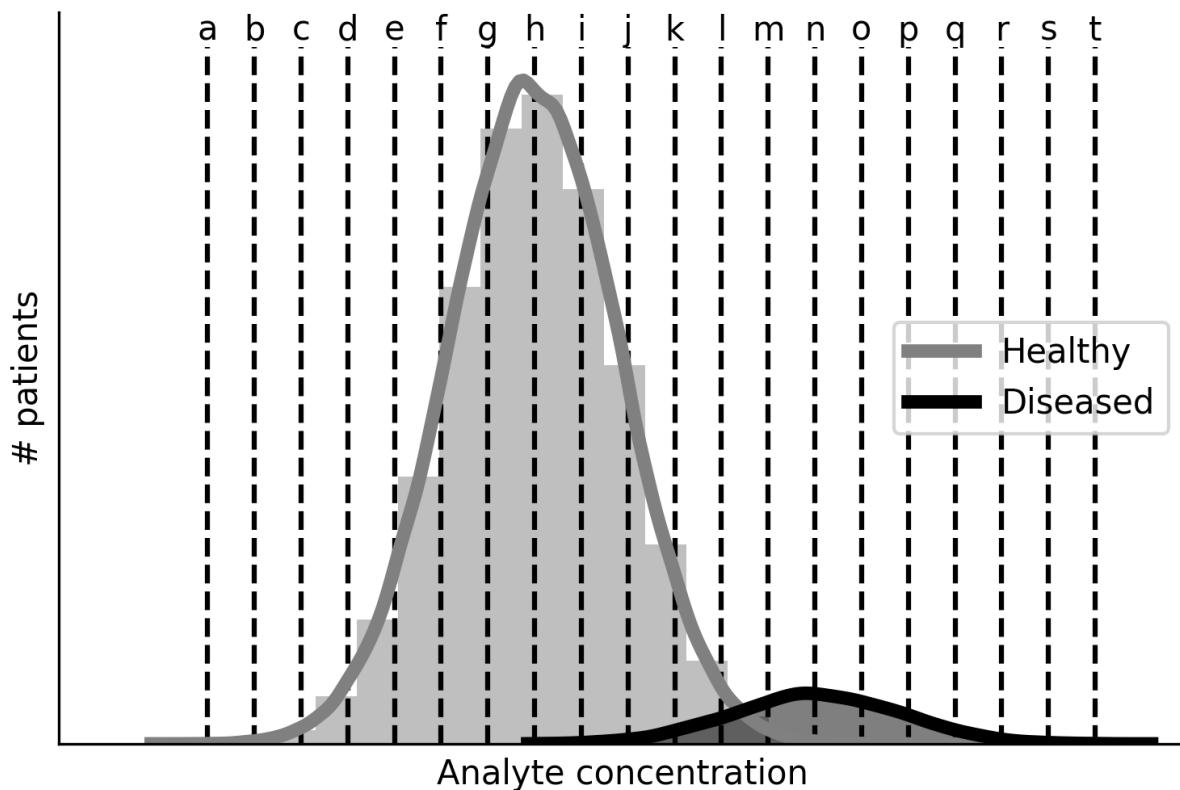
Key points:

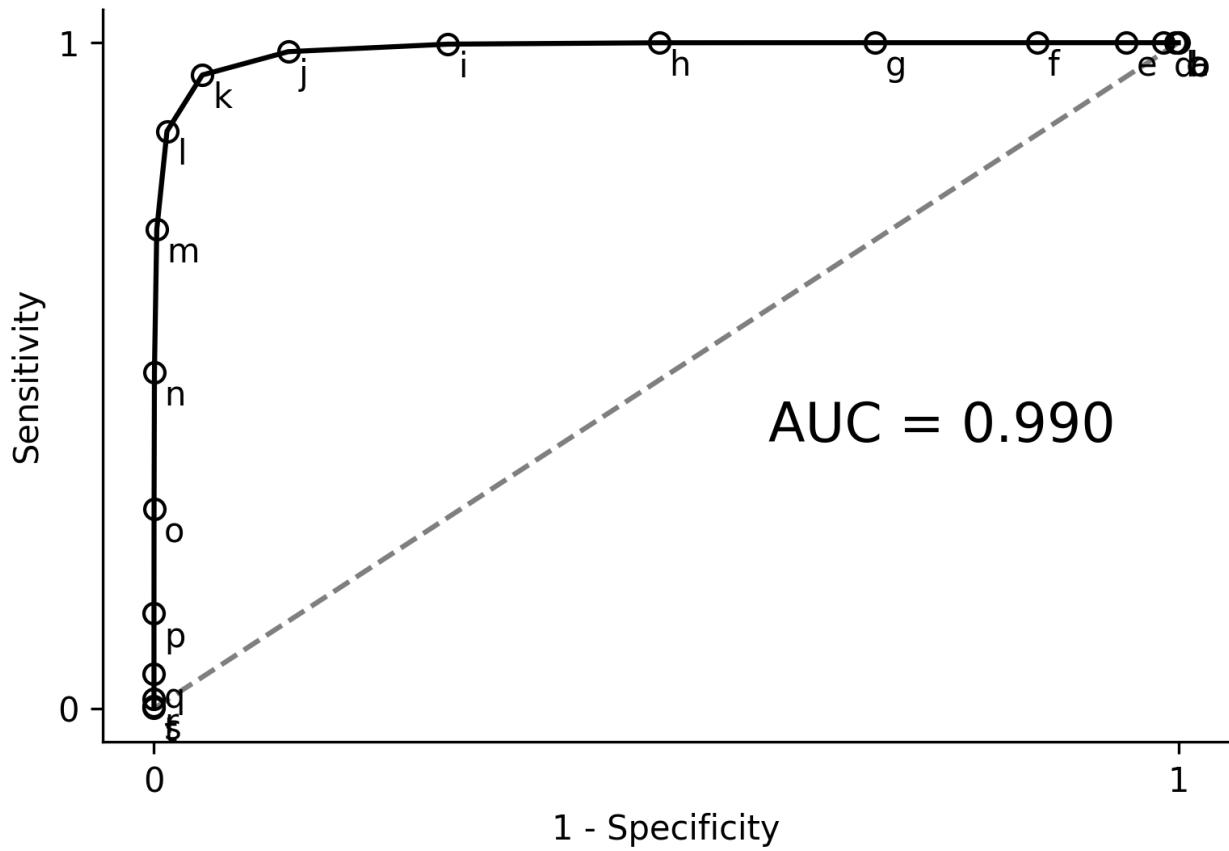
1.  $Sp$  is the fraction of healthy individuals classified as negative by the test
2.  $Sp$  varies from zero to one

### 3. $Sp$ increases with increasing threshold value

#### Receiver operating characteristic (ROC) curves

Reciever operating characteristic (ROC) curves summarize the trade off between sensitivity and the false positive rate ( $1-Sp$ ) that occurs when we vary the threshold value (**Figure 12**). This type of graphical analysis has roots in the tuning of military radar systems where one had to find a balance between optimizing sensitivity to detect potential enemy aircraft and optimizing specificity such that debris or birds did not constantly trigger false alarms.





**Figure 12:** Illustration of the construction of the ROC curve for the test the realistic test in Figure 6. A series of threshold value (a-t) are considered (top). For each potential threshold  $Se$  is plotted versus  $1-Sp$  (bottom).

The area under the ROC curve (AUROCC) is an important single statistic that is often used to summarize test accuracy. As the name would suggest, AUROCC is computed by integrating the area under the ROC curve. AUROCC varies from 0.5 in the case of a useless test to 1 in the case of an ideal test. AUROCC is not sensitive to disease prevalence as neither  $Se$  nor  $Sp$  vary with prevalence. AUROCC can be a confusing statistic because in practice we typically choose a single diagnostic threshold, so what does AUROCC actually mean? It turns out that AUROCC has a very interesting statistical meaning. It is the probability that a randomly selected diseased individual will have a higher analyte concentration than a randomly selected healthy individual. In this light it becomes clear why the useless test will have an AUROCC of 0.5 - since the probability distributions functions of analyte concentrations for healthy and diseased individuals are the same in the case of a useless test there is a 50% chance that a randomly selected diseased individual will have a higher analyte concentration than a randomly selected healthy individual.

## Positive and negative predictive values

The positive and negative predictive values of a test are often of the most immediate interest to the clinician who needs to estimate the probability of disease or health for a specific patient given the patient's test result. The posttest probability of the disease can be estimated from the predictive values as long as the patient's pretest probability of the disease is equal to the prevalence of the disease in the population. Beware that the predictive values for the population should not be used to estimate the post-test disease probability for a patient that has characteristics that identify them as low or high risk relative to the background population. In this case the post-test probability should be estimated using the likelihood ratio and Bayes theorem as will be described later in the section on likelihood ratios.

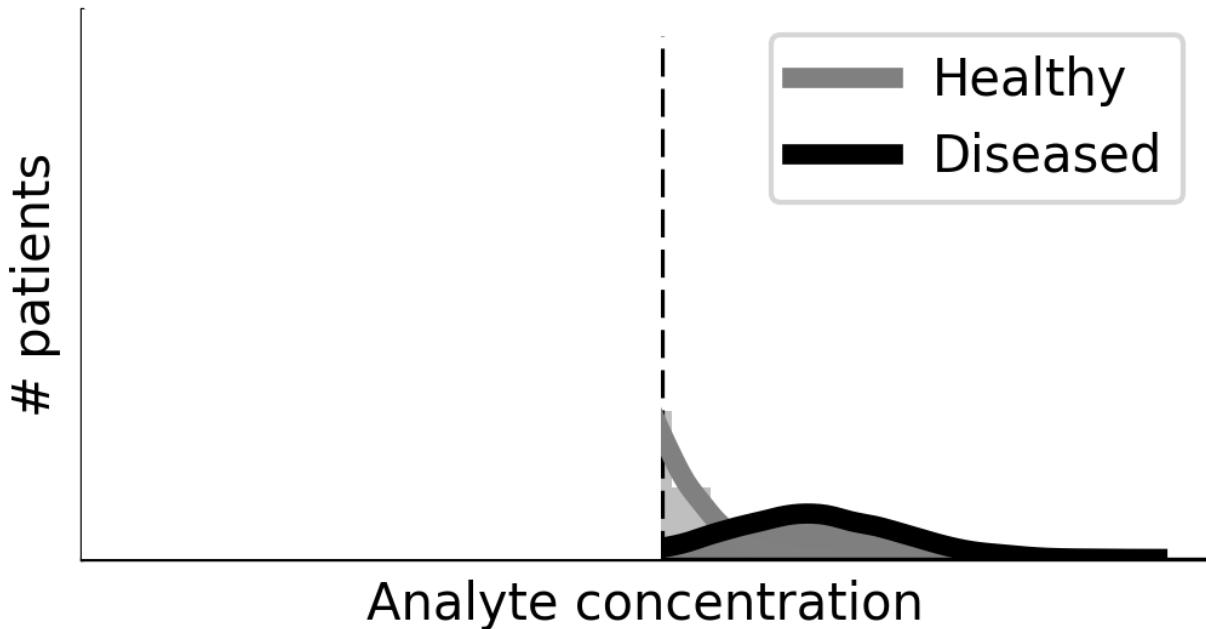
### **Positive predictive value (PPV)**

The positive predictive value (PPV) of a test is the fraction of individuals with a positive test result who carry the disease. Hence a high PPV is important if we are trying to establish (i.e. rule-in) a diagnosis.

$$PPV = \frac{TP}{TP + FP}$$

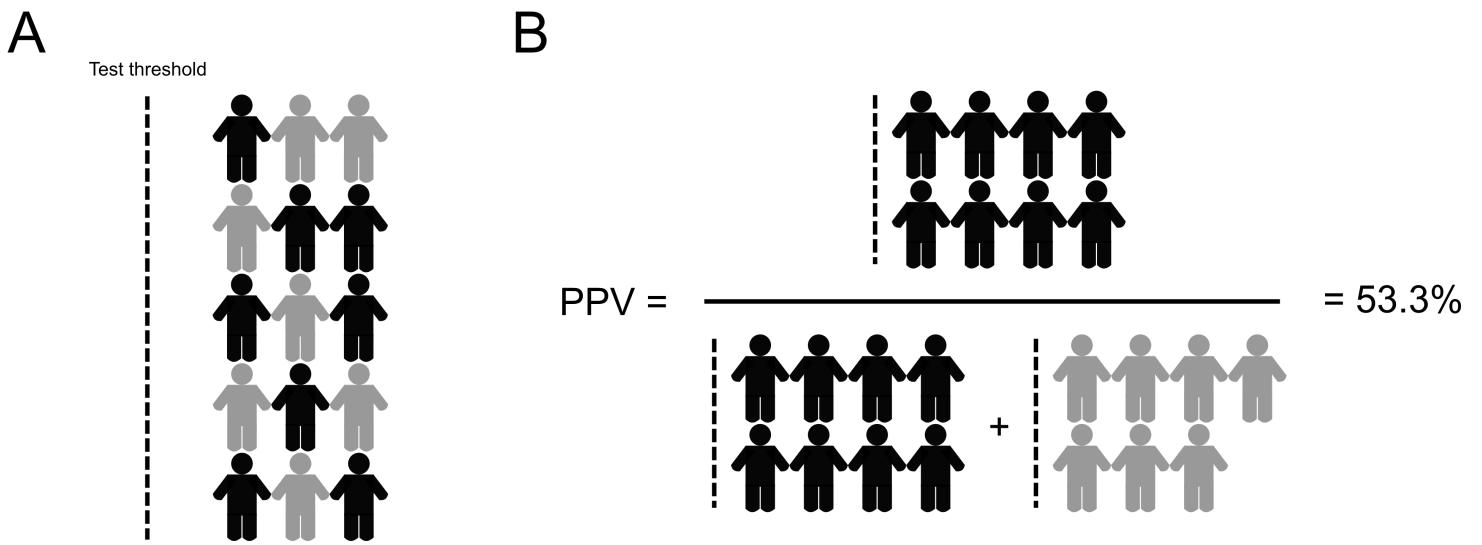
where  $TP$  is the number of diseased individuals classified as positive ('true positives'), and  $FP$  is the number of healthy individuals classified as positive ('false positives').

Mathematically,  $PPV$  is the ratio of the number of diseased and healthy individuals that fall above the test threshold in the positive region (**Figure 13**).  $PPV$  increases to one as the threshold is increased and decreases to zero as the threshold is decreased. Note that PPV will change with diseased prevalence because it depends on the relative numbers of healthy and diseased individuals. Specifically, we expect PPV to increase to one as the prevalence is increased and decrease to zero as prevalence is decreased.



**Figure 13:** Graphical representation of *PPV*.

Figure 14 illustrates how to calculate *PPV* for the realistic test depicted in Figures 6 and 7. Note that for the chosen diagnostic threshold, 7 and 8 healthy and diseased individuals would be classified as positive, respectively (**Figure 14A**). Plugging into our formula we find that  $PPV = \frac{8}{8+7} = \frac{8}{15} = 53.3\%$  (**Figure 14B**).



**Figure 14:** Illustration of positive predictive value (PPV) calculation for the realistic test depicted in Figure 7. (A) Disease status of patients with a result above the test threshold. (B) Calculation of PPV.

#### Key points:

1. *PPV* is the fraction individuals with a positive test who carry the disease
2. *PPV* varies from zero to one

3.  $PPV$  increases with increasing threshold value
4.  $PPV$  increases with increasing disease prevalence

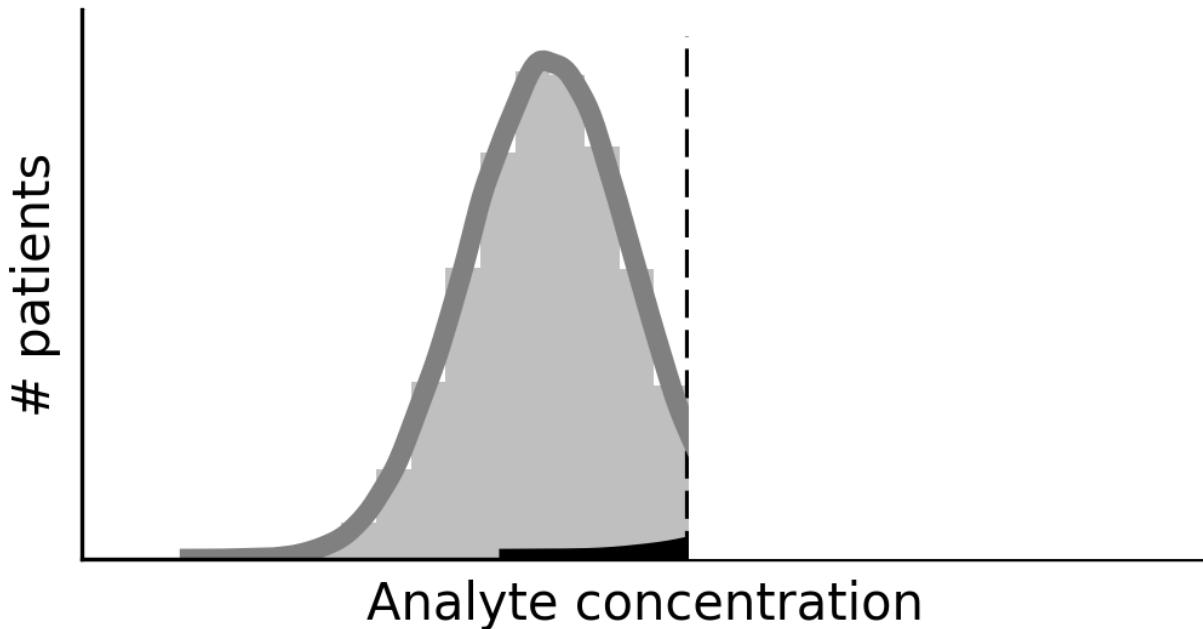
### Negative predictive value

The negative predictive value (NPV) of a test is the fraction of individuals with a negative test result who are healthy. A high NPV is important if we are trying to rule-out a diagnosis.

$$NPV = \frac{TN}{TN + FN}$$

where  $TN$  is the number of healthy individuals classified as negative ('true negatives'), and  $FN$  is the number of diseased individuals classified as positive ('false negatives').

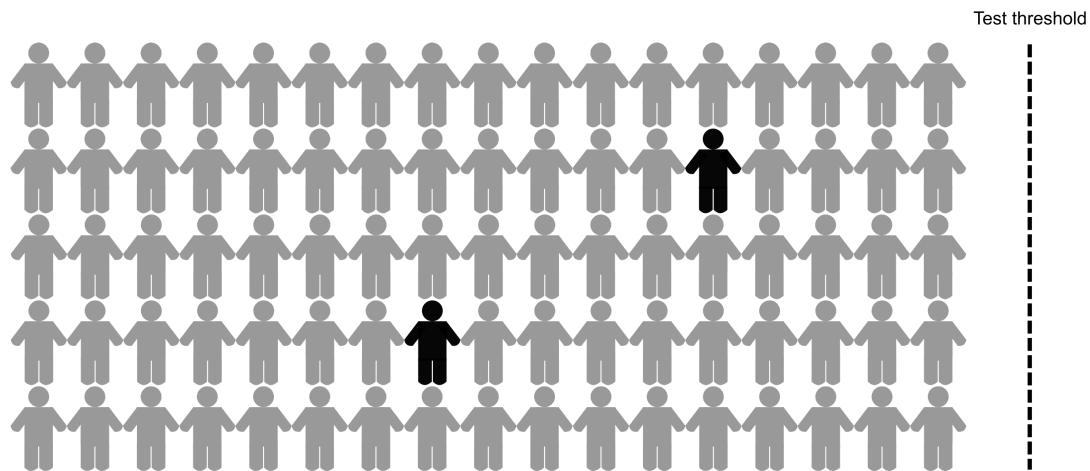
Mathematically,  $NPV$  is the ratio of the number of healthy and diseased individuals that fall below the test threshold into the negative region (**Figure 15**).  $NPV$  increases to one as the threshold is decreased and decreases to zero as the threshold is increased. Note that NPV will change with diseased prevalence because it depends on the relative numbers of healthy and diseased individuals. Specifically, we expect NPV to increase to one as the prevalence is decreased and increase to zero as prevalence is decreased.



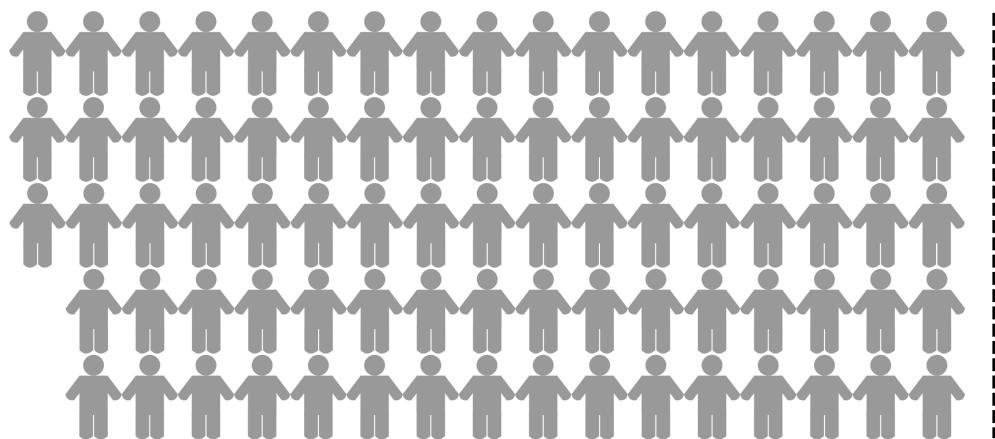
**Figure 15:** Graphical representation of  $NPV$ .

Figure 16 illustrates how to calculate  $NPV$  for the realistic test depicted in Figures 6 and 7. Note that for the chosen diagnostic threshold, 98 healthy individuals (gray) and 2 diseased individuals (black) had a test result below threshold, so  $NPV = \frac{98}{2+98} = \frac{98}{100} = 0.98$ .

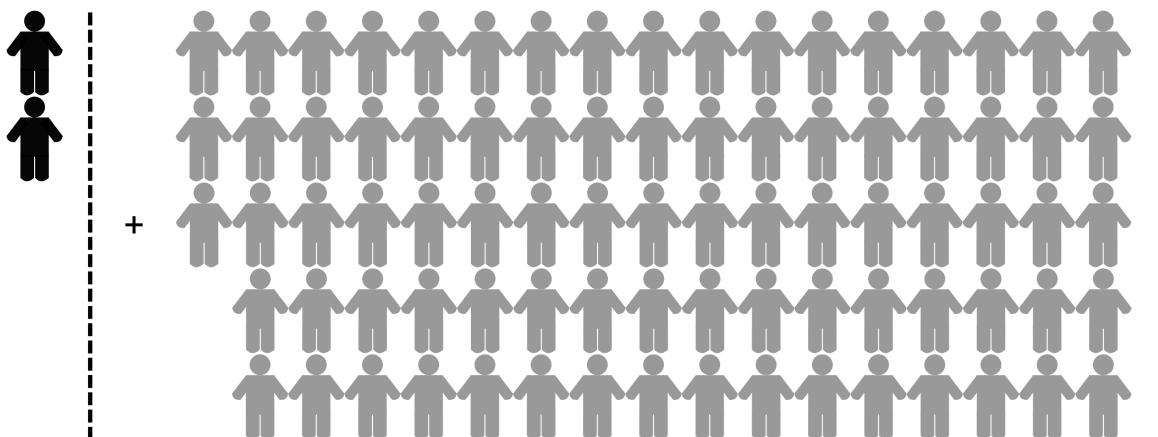
A



B



$$NPV = \frac{\text{Number of healthy patients}}{\text{Total number of patients}} = 98\%$$



**Figure 16:** Illustration of Negative predictive value (NPV) calculation for the realistic test depicted in Figure 7. (A) Disease status of patients with a result below the test threshold. (B) Calculation of NPV.

Key points:

1.  $NPV$  is the fraction individuals with a negative test who are healthy
2.  $NPV$  varies from zero to one
3.  $NPV$  decreases with increasing threshold value
4.  $NPV$  decreases with increasing disease prevalence

## Likelihood ratios and diagnostic odds ratio

Likelihood ratios are extremely useful because they provide a direct relationship between the pre-test and post-test odds of a disease. In other words, we can use the likelihood ratio to estimate post-test probability of the disease for a specific patient even if that patient's pre-test probability of the disease differs from the prevalence of the disease in the background population. Likelihood ratios are specific to a particular test result (i.e. negative or positive) and give a sense of how the odds will adjust when that particular result is observed.

Diagnostic odds ratio ( $DOR$ ) relates the post-test odds of the disease among individuals who test positive versus negative. Like AUROCC,  $DOR$  is a single statistic that provides a sense of the overall accuracy of the test.

### Positive likelihood ratio ( $LR^+$ )

The positive likelihood ratio ( $LR^+$ ) is the ratio of the likelihood of a positive test among diseased individuals to that among healthy individuals. As will be described below,  $LR^+$  provides a direct relationship between the pre-test and post-test odds of the disease in patients with a positive test result.

$$LR^+ = \frac{p(T^+|D^+)}{p(T^+|D^-)} = \frac{\frac{TP}{TP+FN}}{\frac{FP}{FP+TN}} = \frac{Se}{1 - Sp}$$

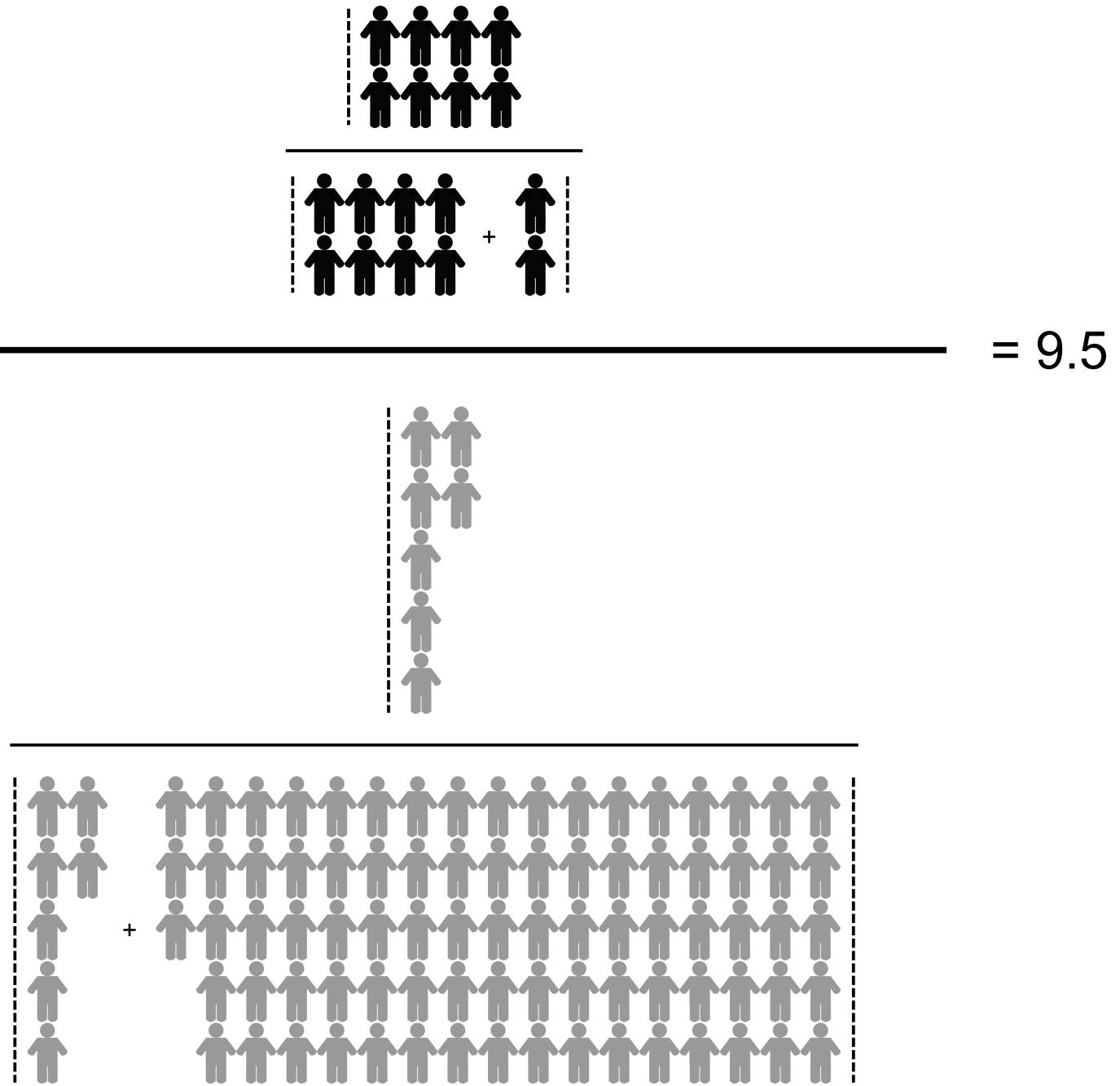
where  $p(T^+|D^+)$  and  $p(T^+|D^-)$  are the probabilities of a positive test in a diseased and healthy individual respectively.

$LR^+$  varies from zero to positive infinity and can be interpreted as follows:

Case	LR <sup>+</sup> Value	Interpretation
1	$0 < LR^+ < 1$	Backwards test (positive = healthy) Accuracy increases as $LR^+ \rightarrow 0$
2	$LR^+ = 1$	Positive result is useless
3	$1 < LR^+ < +\infty$	Forwards test (positive = diseased) Accuracy increases as $LR^+ \rightarrow +\infty$

Figure 17 illustrates how to calculate  $LR^+$  for the realistic test depicted in Figures 6 and 7. Plugging the sensitivity and specificity that we calculated above into our formula we find that  $LR^+ =$

$$\frac{0.8}{1-0.922} = \frac{0.8}{0.078} = 10.3$$



**Figure 17:** Illustration of the positive likelihood ratio ( $LR^+$ ) calculation for the realistic test depicted in Figure 7.

#### Key points:

1.  $LR^+$  is ratio of the likelihood of a positive test in diseased versus healthy individuals
2.  $LR^+$  varies from zero to infinity with accuracy increasing bidirectionally towards zero and infinity.
3.  $LR^+$  does not change with prevalence

## Negative likelihood ratio ( $LR^-$ )

The negative likelihood ratio ( $LR^-$ ) is the ratio of the likelihood of a negative test among diseased individuals to that among healthy individuals. As will be described below,  $LR^-$  provides a direct relationship between the pre-test and post-test odds of the disease in patients with a negative test result.

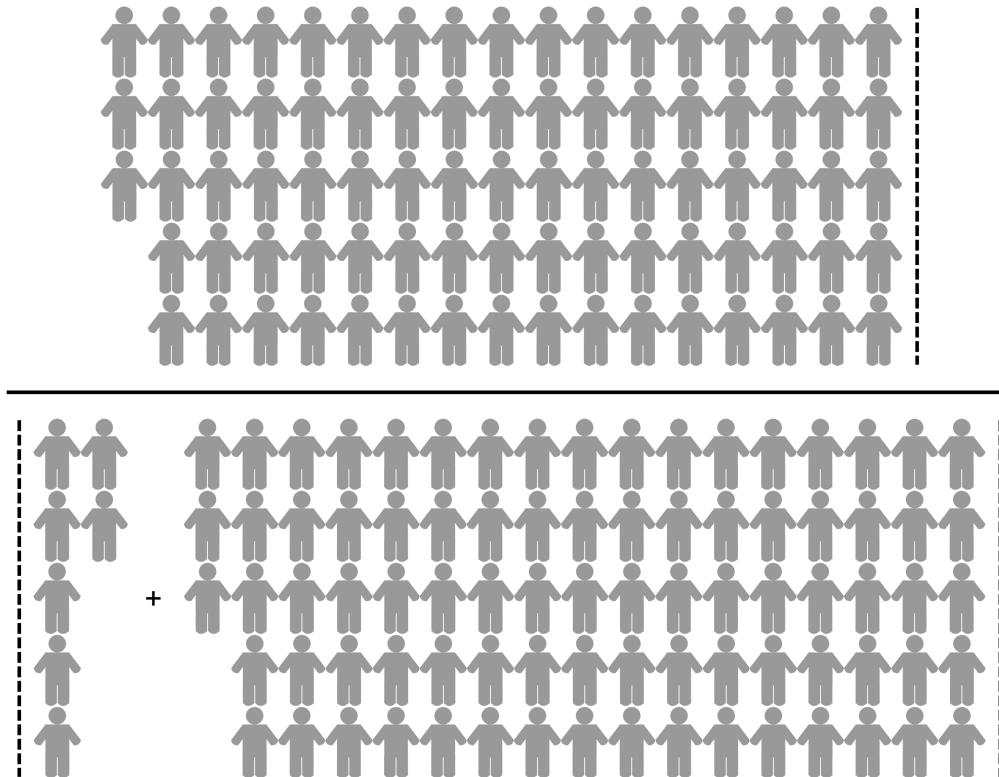
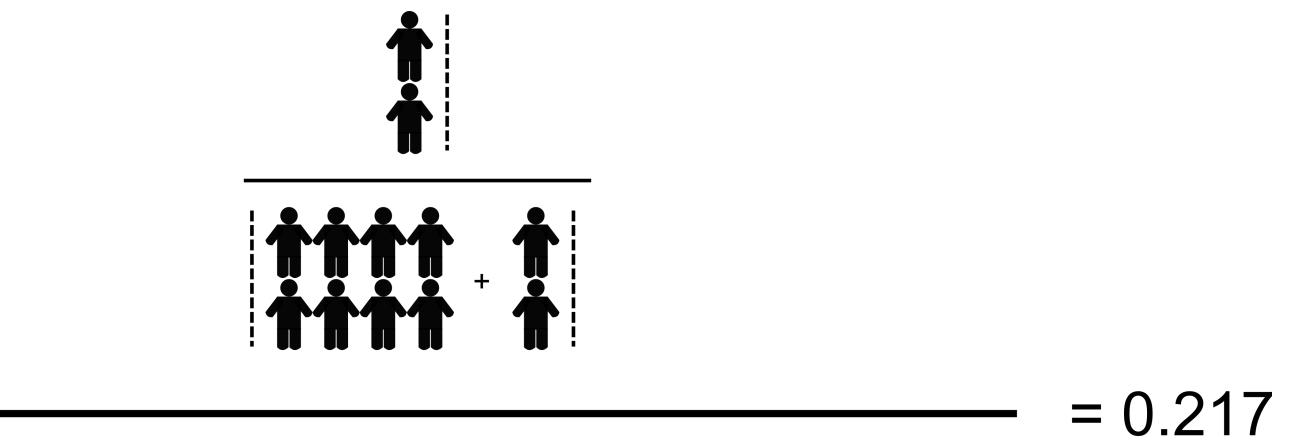
$$LR^- = \frac{p(T^-|D^+)}{p(T^-|D^-)} = \frac{\frac{FN}{TP+FN}}{\frac{TN}{FP+TN}} = \frac{1 - Se}{Sp}$$

where  $p(T^-|D^+)$  and  $p(T^-|D^-)$  are the probabilities of a negative test in a diseased and healthy individual respectively.

$LR^-$  varies from zero to positive infinity and can be interpreted as follows:

Case	$LR^-$ Value	Interpretation
1	$0 < LR^- < 1$	Forwards test (positive = diseased) Accuracy increases as $LR^- \rightarrow 0$
2	$LR^- = 1$	Positive result is useless
3	$1 < LR^- < +\infty$	Backwards test (positive = healthy) Accuracy increases as $LR^- \rightarrow +\infty$

Figure 18 illustrates how to calculate  $LR^-$  for the realistic test depicted in Figures 6 and 7. Plugging the sensitivity and specificity that we calculated above into our formula we find that  $LR^- = \frac{1-0.8}{0.922} = \frac{0.2}{0.922} = 0.22$



**Figure 18:** Illustration of the negative likelihood ratio ( $LR^-$ ) calculation for the realistic test depicted in Figure 7.

Key points:

1.  $LR^-$  is ratio of the likelihood of a negative test in diseased versus healthy individuals
2.  $LR^-$  varies from zero to infinity with accuracy increasing bidirectionally towards zero and infinity.
3.  $LR^-$  does not change with prevalence

### Diagnostic Odds Ratio (DOR)

Diagnostic Odds Ratio (DOR) is the ratio of  $LR^+$  and  $LR^-$ .

$$DOR = \frac{LR^+}{LR^-} = \frac{TP * TN}{FP * FN}$$

As will be described below,  $DOR$  provides a direct relationship between the post-test odds of the disease among patients with a positive versus negative test result. Like AUROCC, DOR provides a single statistic reflecting the overall accuracy of the test.  $DOR$  varies from zero to positive infinity and can be interpreted as follows:

Case	DOR Value	Interpretation
1	$0 < DOR < 1$	Backwards test (positive = healthy) Accuracy increases as $LR^- \rightarrow 0$
2	$DOR = 1$	Useless test
3	$1 < DOR < +\infty$	Forwards test (positive = diseased) Accuracy increases as $DOR \rightarrow +\infty$

We can calculate the  $DOR$  for the realistic test depicted in Figure 7 by plugging the  $LR^+$  and  $LR^-$  that we calculated above into our formula.

$$DOR = \frac{10.3}{0.22} = 46.8$$

#### Key points:

1.  $LR^-$  is ratio of the likelihood of a negative test in diseased versus healthy individuals
2.  $LR^-$  varies from zero to infinity with accuracy increasing bidirectionally towards zero and infinity.
3.  $LR^-$  does not change with prevalence

#### Likelihood ratios relate pre-test and post-test odds given a specific result

This section will discuss how the  $LR^+$  and  $LR^-$  can be used to relate pre-test and post-test odds of the diseased. Before we start, it is important review the difference between the probability and odds of disease.

#### **Odds versus Probabilities**

Probability refers to the ratio of the number of individuals with the disease to the total number of individuals.

$$P(\text{disease}) = \frac{N_{\text{diseased}}}{N_{\text{diseased}} + N_{\text{healthy}}}$$

where  $P(\text{disease})$  is the probability of the disease,  $N_{\text{diseased}}$  is the number of diseased individuals and  $N_{\text{healthy}}$  is the number of healthy individuals.

Odds refers to the ratio of the number of individuals with the disease to the number of individuals who do not have the disease, i.e. healthy.

$$O(\text{disease}) = \frac{N_{\text{diseased}}}{N_{\text{healthy}}}$$

where  $O(\text{disease})$  is the odds of the disease.

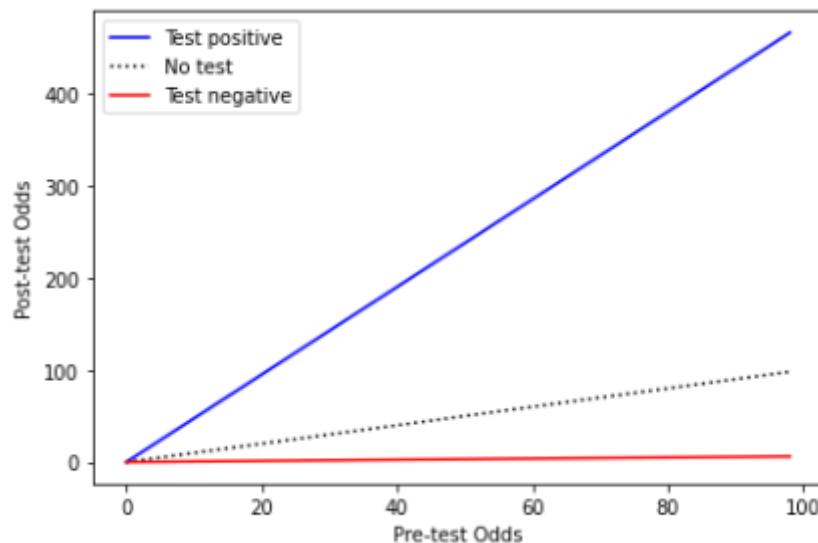
Odds can be computed from probability using the following relationship:

$$O(\text{disease}) = \frac{P(\text{disease})}{1 - P(\text{disease})}$$

Probability can be computed from odds as follows:

$$P(\text{disease}) = \frac{O(\text{disease})}{1 + O(\text{disease})}$$

Figure 19 illustrates the relationship between the post-test and pre-test odds of the diseased for a patient with a positive (blue) or negative (red) test result using the realistic test illustrated in Figure 7.



**Figure 19:** Illustration of the relationship between post-test odds and pre-test odds given a positive (blue) or negative (red) test result using the realistic test depicted in Figure 7. The dashed black line

indicates equivalency between post-test and pre-test odds as if no test had been performed.

Note that there is a linear relationship in each case, but the slope is steep for the case of a positive case such that the post-test odds are greater than the pre-test odds and shallow for the case of the negative case such that the post-test odds are less than the pre-test odds. Given these two curves we can easily estimate the post-test odds (and in turn the post-test probability) for any individual patient given their test result and an estimate of their pre-test probability based on their specific clinical characteristics. In the following, we prove that slopes of these blue and red lines in figure 19 are equal to the  $LR^+$  and  $LR^-$  of the test respectively.

### Mathematical proof #1

We start with the definition of  $LR^+$  :

$$LR^+ = \frac{P(T^+|D^+)}{P(T^+|D^-)}$$

where  $P(T^+|D^+)$  and  $P(T^+|D^-)$  is the probability of a positive test result for patients with and without the disease respectively.

Bayes' theorem states that:

$$P(T^+|D^+) = \frac{P(D^+|T^+)P(T^+)}{P(D^+)}$$

and

$$P(T^+|D^-) = \frac{P(D^-|T^+)P(T^+)}{P(D^-)}$$

where  $P(D^+|T^+)$  and  $P(D^-|T^+)$  is the probability of disease and health among individuals with a positive test result respectively,  $P(T^+)$  is the probability of a positive test among all individuals, and  $P(D^+)$  and  $P(D^-)$  is the probability of disease and health among all individuals.

Plugging in to the equation for  $LR^+$ :

$$LR^+ = \frac{P(T^+|D^+)}{P(T^+|D^-)} = \frac{\frac{P(D^+|T^+)P(T^+)}{P(D^+)}}{\frac{P(D^-|T^+)P(T^+)}{P(D^-)}} = \frac{\frac{P(D^+|T^+)}{P(D^+)}}{\frac{P(D^-|T^+)}{P(D^-)}}$$

Next we consider the following based on the definition of  $PPV$  and prevalence:

$$P(D^+|T^+) = PPV$$

$$P(D^-|T^+) = 1 - PPV$$

and

$$P(D^+) = \text{Prevalence}$$

$$P(D^-) = 1 - \text{Prevalence}$$

Plugging in we get:

$$LR^+ = \frac{\frac{PPV}{\text{Prevalence}}}{\frac{1-PPV}{1-\text{Prevalence}}}$$

Rearranging:

$$\frac{PPV}{1-PPV} = LR^+ * \frac{\text{Prevalence}}{1-\text{Prevalence}}$$

Considering the definition of Odds we find that:

$$\frac{PPV}{1-PPV} = O^{T^+}$$

and

$$\frac{\text{Prevalence}}{1-\text{Prevalence}} = O^{\text{Pretest}}$$

Where  $O^{T^+}$  are the post test odds of the disease for a patient with a positive test result and  $O^{\text{Pretest}}$  are the pretest odds of the disease.

Thus we arrive at the relationship:

$$O^{T^+} = LR^+ * O^{\text{Pretest}}$$

This relationship indicates that the  $LR^+$  represents a scalar factor by which the odds of the disease are adjusted upon receipt of a positive test result.

An analogous derivation would show that the following relationship is true:

$$O^{T^-} = LR^- * O^{\text{Pretest}}$$

Where  $O^{T^-}$  is the odds of the disease in an individual with a negative test result. This relationship indicates that the  $LR^-$  represents a scalar factor by which the odds of the disease are adjusted upon receipt of a negative test result.

## Diagnostic odds ratio relates post-test odds of the disease for individuals with a negative versus positive test result

In Figure 19, any point on the blue curve (the post-test probability for individuals with a positive test) is X times greater than the equivalent point on the red line (the post-test probability for individuals with a negative test). It turns out that the value of the scalar X is the diagnostic odds ratio. In the following we prove this point.

### Mathematical proof #2

We start by plugging the relationships from the conclusion of mathematical proof #1 above into the formula for diagnostic odds ratio:

$$DOR = \frac{LR+}{LR-} = \frac{\frac{O^{T^-}}{O^{Pretest}}}{\frac{O^{T^+}}{O^{Pretest}}} = \frac{O^{T^+}}{O^{T^-}}$$

Hence, as the name would suggest, DOR represents the ratio of the odds of the disease for patients with a positive versus negative test result.

## Glossary

**True positives (TP)** : The number of positive test cases that occur when the disease is present.  
Synonyms: "hit"

**False positives (FP)** : The number of positive test cases that occur when the disease is absent.  
Synonyms: "Type-I error,  $\alpha$  error, false alarm"

**True negatives (TN)** : The number of negative test cases that occur when the disease is absent.

**False negatives (FN)** : The number of positive test cases that occur when the disease is present.  
Synonyms: "Type-II error,  $\beta$  error, miss"

**Prevalence** : The fraction of the individuals in the tested population afflicted by the disease.

$$Prevalence = \frac{TP + FN}{TP + FP + FN + TN}$$

**Sensitivity (Se)** : The fraction of diseased individuals that will have a positive test result. Synonyms: "Recall, True Positive Rate, Hit Rate, Probability of Detection, Power"

$$Se = \frac{TP}{TP + FN}$$

**Specificity (Sp)** : The fraction of healthy individuals that will have a negative test result. Synonyms: "True Negative Rate, Selectivity"

$$Sp = \frac{TN}{TN + FP}$$

**Accuracy (Acc)** : Fraction of individuals who were correctly classified by the test. In other words the fraction of the total cases that falls along the diagonal elements of the two by two contingency table.

$$ACC = \frac{TP + TN}{TP + FP + FN + TN}$$

\*\*\*False Positive Rate (FPR)\*\*\*: Fraction of the healthy individuals who incorrectly test as positive. Synonyms: "fall-out, probability of false alarm"

$$ACC = 1 - Sp = \frac{FP}{TN + FP}$$

\*\*\*False Negative Rate (FNR)\*\*\*: Fraction of the healthy individuals who incorrectly test as positive. Synonyms: "fall-out, probability of false alarm"

$$ACC = \frac{FP}{TN + FP}$$

\*\*\*Positive Predictive Value (PPV)\*\*\*: The fraction of individuals with a positive test that do not have the disease. Synonyms: "Precision"

$$PPV = \frac{TP}{TP + FP}$$

**False Discovery Rate (FDR)** : The fraction of individuals with a positive test that do not have the disease (i.e. they are healthy).

$$FDR = \frac{FP}{TP + FP}$$

\*\*\*Negative Predictive Value (NPV)\*\*\*: The fraction of individuals with a negative test that do not have the disease (i.e. they are healthy).

$$NPV = \frac{TN}{TN + FN}$$

**False Omission Rate (FOR)** : The fraction of individuals with a negative test that have the disease.

$$FOR = \frac{FN}{FN + TN}$$

**F<sub>1</sub> Score** : Harmonic mean of precision (i.e. PPV) and recall (i.e. Se). Indicates the tests ability to both accurately and completely classify diseased patients. Will change to reflect changes in the disease prevalence.

$$F_1 = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * PPV * Se}{PPV + Se}$$

**Positive Likelihood Ratio (LR<sup>+</sup>)** : The ratio of the likelihood of a positive test result in diseased versus healthy individuals.

$$LR^+ = \frac{TPR}{FPR} = \frac{\frac{TP}{TP+FN}}{\frac{FP}{FP+TN}}$$

**Negative Likelihood Ratio (LR<sup>-</sup>)** : The ratio of the likelihood of a negative test result in diseased versus healthy individuals.

$$LR^- = \frac{FNR}{TNR} = \frac{\frac{FN}{TP+FN}}{\frac{TN}{FP+TN}}$$

**Diagnostic Odds Ratio (DOR)** : The ratio of the LR<sup>+</sup> to LR<sup>-</sup>

$$DOR = \frac{LR^+}{LR^-} = \frac{TP * TN}{FP * FN}$$

**Prevalence Threshold (PT)** :

$$PT = \frac{\sqrt{TPR * FPR} - FPR}{TPR - FPR}$$

**Jaccard Index** :

**Contingency table (aka 2x2 Table)** : A 2x2 matrix describing the joint distribution of the test and disease results. Synonyms: "confusion matrix"

<u>Disease status</u>	<u>Test pos</u>	<u>Test neg</u>
Diseased	TP	FN
Healthy	FN	TN

**Table 1:** A generic contingency table.

