Glossary of Statistical Terms (2017)

# Study Designs- Key Questions

1. Was there a **control group**? (Case reports and case series have no control so are less reliable; we do not know what would have happened without the treatment?)
2. Is the design **experimental** or **observational**? (In other words did the patients choose their treatment and the results were **observational,** or was their treatment chosen for them by the investigator in an **experimental design (**such as a **randomised trial)**?
3. **Observational designs** include Cross-sectional surveys, Case-control studies and Cohort studies. The key difference between them relates to how they use time. **Cross-sectional surveys** are a snap shot at a single moment in time, **Case-control studies look back** from the final event (such as lung cancer) and compare risk factors (such as smoking) between the cases who had lung cancer and the controls who did not. **Cohorts march forwards**. So they start with two groups that had differences in their risk factors (those who are taking HRT and those who are not for example) and then follow up the cohorts looking for events to occur (such as PE, MI, breast cancer). Looking forwards is called **prospective** study design, whilst looking back is called **retrospective**.
4. **Experimental designs (randomised trials) are the only type of study that can assess causation.** They also move forwards in time from treatments to events (like cohort studies), but the randomisation tries to ensure that there are no differences between the two groups apart from the randomised treatment. This means that there is less likelihood of different results in each group due to other **confounding** factors.
5. I would recommend reading the wonderfully well described example of a **case-control study on smoking and lung cancer [**Doll R, Hill AB. Smoking and Carcinoma of the Lung. BMJ 1950;2(4682):739-48 **]**, followed by one of the reports on his subsequent **cohort study on UK doctors.**

# The problem of Bias

Bias can mislead us into thinking that a treatment is associated with a particular benefit or harm, when actually the difference is due to something else. The early cohort studies on HRT are a good example of this. They showed that women taking HRT had lower risks of suffering cardiovascular disease than those control women who did not take HRT. However when randomised trials were carried out later they showed the opposite effect, if anything taking HRT was associated with a small increase in risks of heart attacks. So why is there the difference? Which would you trust?

Careful analysis of all the outcomes in the HRT cohort studies gave some clues, and there was serious bias present in the cohort studies (Davey Smith G, Ebrahim S. Data dredging, bias, or confounding. BMJ. 2002;325(7378):1437-8). It turned out that the women who chose to take HRT were less likely to smoke, had a more healthy diet and took more exercise. This was the reason that less of them had heart attacks. So there was **confounding** of the results by another factor. If you spot that such differences are present between the groups it is possible to make statistical adjustments, but you can only adjust for the differences that you spot and cannot adjust for **unknown** confounders.

## Types of Bias that can occur in Randomised Trials

**Selection Bias** occurs when there is a difference between the groups in prognostic factors that influence the outcomes (such as HRT above). Randomisation with secure **allocation concealment** is the best protection against such bias. **Allocation concealment** is hiding the treatment that the patient will receive until they have been recruited into the trial, so the investigator cannot tell which treatment they will receive in advance, and put healthier patients on their favourite treatment.

**Performance Bias** occurs when the patient tries harder on one treatment than another, and **Ascertainment Bias** is where the investigator looks harder for an outcome on one treatment too. The best protection against these biases is through **blinding** of the patient and investigator respectively. Remember that although you cannot always blind a trial (such as for surgical procedures)**, but you can always conceal allocation**.

**Reporting Bias** occurs when an outcome is measured over many time points and the investigator highlights the one point at which there is a statistically significant difference. Alternatively an important outcome may have been measured but not reported at all. Published protocols with a specified plan of outcomes and how they will be analysed helps to reduce this bias.

**Attrition Bias** occurs when the people who do not complete that study have worse outcomes, but because they have dropped out, their results are not included in the analysis.

**Publication Bias** is a similar problem on a larger scale when a whole trial remains unpublished, especially if the results are not significant. For this a trial registry is helpful.

## Additional Types of Bias

**Recall Bias** can be a problem in retrospective case-control studies if the outcome status means that one group is more likely to recall a risk factor accurately than the other group.

When there is no randomisation (such as in a cohort study), **selection bias** occurs when the reason that people choose the treatment is related to the outcomes being measured. So in the HRT example above, those less at risk of heart disease were more likely to choose to take HRT.

Finally, remember that the possible effects of bias are not reflected in P values or 95% confidence intervals, as these only relate to uncertainty in the results due to the play of chance. However, assessment of bias is now routinely used in Cochrane **systematic reviews** and is reflected in the Summary of Findings tables in the reviews, which consider both the play of chance and the likely impact of bias and how together these may alter our confidence in the results of the review. Trials at **low risk of bias** have good **Internal Validity** (we can trust their results more), whereas **External Validity** (how well the results apply to our patients), depends on trial patients being similar to ours.

# Descriptive Statistics

### Samples from a population are described by where the middle of the sample is and what its spread is.

### The middle can be defined by mean, median or mode.

## The Mean is:

The average of all the readings (the total divided by the number of readings). It is the most useful measure **for Parametric data** (in other words data that follows a normal distribution).

## The Median is:

The reading which has an equal number of other readings that are larger and smaller (regardless of how much higher or lower they may be). This will be the same as the mean if the data are normally distributed, but it is less influenced by outliers if the distribution is skewed. Is the measure of choice for skewed data.

## The Mode is:

The result found in the largest number of participants in the sample. It is not usually used in statistical analysis. For parametric data it will be the same as the mean and median.

### The spread of a normal distribution is described by its standard deviation (SD), which is the square root of the sum of the squares of the difference between each reading and the mean.

Why do we have to use the squares of the distance of each reading from the mean? The answer is that by definition the sum of the differences between each reading and the mean is always zero, but by squaring the differences, all the minus signs disappear so the distances can be added together without making zero!

### In a normal distribution you are expected to know that half of the sample will have results smaller than the mean and half larger; 97.5% of the sample will have a result smaller than the mean plus 1.96 standard deviations.

The spread of a non-parametric sample (such as a skewed sample) is described using the inter-quartile range (IQR), which divides the distribution into quarters. This means that 75% of people in the sample will have a measurement that is smaller than the upper inter-quartile limit, and 25% will be smaller than the lower limit with 50% of the sample inside the IQR. Again the IQR is less influenced by outliers. The median, IQR and range may be shown in a **Box and Whisker** **Plot**.

There is an example of a **Box and Whisker Plot** on the next page. The range is shown by the ends of the vertical line, the top and bottom of the box are the upper and lower IQR and the Median is the central horizontal line.



Median

IQR

Range

The asterisks and solid lines at the top of the graph show that there is a statistically significant difference between the BODE index Pre and Post intervention in all the settings (Out-Patient, At Home and Control Group). The long line across the top (with NS above it, standing for Not Significant), shows that there is not a significant difference between the changes for people treated as outpatients or at-home.

# Inductive Statistics

When carrying out experiments to find out how a treatment works, or to describe the average (mean) FEV1 in a group of patients with asthma, we cannot measure every possible person who exists with the condition. We have to use a sample of a limited size to estimate what the true mean of the population really is.

But how accurate is our estimated mean from the sample in relation to the mean of the whole population? You will not be surprised that the bigger the sample size, the closer the sample mean will be to the true mean in the whole population.

The “error” that might be present in our sample mean (in comparison to the mean of the whole population), can be calculated by taking the standard deviation of the sample and dividing it by the number of participants. This is called the **standard error of the mean (SE)**. We would expect the true population mean to be within 1.96 standard errors of our sample mean, with a 95% degree of confidence.

In other words if we took 100 random samples from the same population, the standard error is calculated such that the sample mean will be within 1.96 standard errors of the true population mean in 95 of those 100 samples.

This is the conceptual basis of a **95% confidence interval**, which is calculated such that in 95 experiments out of 100, the true population treatment effect will lie within the 95% confidence interval of that experiment.

In practice this is usually simplified to mean that the **95% confidence interval is where we are 95% sure that the true treatment effect lies**.

Please notice that this uncertainty relates to the **population mean only**, and does not describe what will happen to 95% of the patients in any experiment. Consider this experiment in which you throw 625 red dice, each thrown once. You record the scores on each of the red dice and the average score for all the rolls of the red dice will be very close to 3.5. To see why this is the average, add 1+2+3+4+5+6, making a total of 21 for every six throws, giving an average of 21/6 which is 3.5 per throw.

However if you were to pick any throw of the red die at random, the score can be anything between one and six. If you look at ten throws of dice, you would not be too surprised if the average was two or four, but for all 625 throws together, an average of two or four would make you very suspicious that the dice were weighted!

So how do you work out whether the average of the throws is further from 3.5 than you would expect due to the play of chance? This is the same problem that we have to address when analysing the results of clinical trials, cohort studies or even case-control studies in medicine. Are any differences between the groups of patients more than we might expect from the play of chance alone?

In this case the SD for the mean score on a single die is 2.5, so the standard error is 2.5/25 = 0.1 (as the square root of 625 is conveniently 25)! So the 95% CI of the mean of 625 throws or a die should be 3.5 plus or minus 1.96\*0.1(i.e. 0.196), which is about 3.3 to 3.7. We would be surprised if the average of 625 throws was outside this range. Even though the scores are not normally distributed the central limit theorem says that the means will follow a normal distribution.

# Types of Data

Before we consider how to analyse data from clinical trials and other studies, we need to check some definitions in relation to ways of handling data. Let’s think again about the dice. When you collect the results from each throw of the die what type of data is it?

The answer is that it is **categorical data (also known as ordinal or nominal data)**. There are only six possible scores for each dice, you cannot throw 2.75! Therefore each throw falls into one of six categories. Categorical data can be analysed by tests such as the **Chi-squared test**, which we will look at later.

There is a special type of categorical data called **dichotomous data (or binary data)** in which the results are considered as falling into one of only two categories. Mortality would be a good example, and gender would be another.

However, notice that you could treat the dice throws as dichotomous if you arbitrarily divided the results into low (1 to 3) or high (4 to 6). Dichotomous data has the advantage of being easier to analyse (using **Risk Ratios, Risk Differences, Odds Ratios or Numbers Needed to Treat**), but by combining the categories you are losing some of the available information.

A third type of data is called **continuous data**, because you can divide it up on a continuous scale that is only limited by the ability to measure very small differences. So some good examples of continuous data would be height, weight, age, FEV1 and so on.

Again, any continuous data set can be analysed as categories. Take the St George’s Respiratory Questionnaire that measures quality of life in people with COPD. You can score between zero and 100, but it has been found that a difference of 4 units on this scale makes a clinically important difference to patients. For this reason it can be helpful to divide patients into those that change by less than 4 units, those that fall by 4 units or more, and those that increase by 4 units or more; in other words you can divide the data into three categories. This will enable you to compare how many on each treatment show a clinically important increase or decrease in their score.

Exacerbations are often measured in COPD and asthma. Can you think of different possible ways of measuring data on exacerbations?

1. Continuous?
2. Dichotomous?
3. Categorical?

## Parametric and Non-parametric data

Non-parametric data (which does not have the pattern of a normal distribution), has to be analysed using non-parametric tests, which are less powerful. Examples are the Mann-Witney U test and Wilcoxon test (which both rank the data points and compare the ranks) and Spearman’s rho (for correlation) and Chi-squared test (for categorical data). Parametric Tests (described in more detail below) are T tests, ANOVA or ANCOVA and Pearson’s correlation.

# Null Hypothesis, P values, 95% Confidence Intervals (95%CI)

Let’s return to the example earlier of throwing dice, but this time compare the average score of 625 red dice and 625 black dice, each thrown once. In this experiment there is a **null hypothesis, which is that there is no difference between the average score on red and black dice.** When we carry out a comparison of the dice, our analysis is designed to see if we can **disprove the null hypothesis.**

The statistical tests that we can use are often reported with a **“P value”.** This is defined as the probability that the results of our experiment (or any result that is more extreme), might have occurred by chance **if the null hypothesis is true (i.e. that there is actually no difference between the two groups).** This can also be described as the chance of a **“Type One Error”** in coming to a conclusion about the groups. A **“Type One Error” occurs if we conclude that there is a real difference between the groups, when actually there is no difference.** Traditionally we are prepared to accept making this error 5% of the time, which is why a **“P value”** of 0.05 is considered to be significant.

However, **“P values”** do not tell you anything about the **direction** of the difference between the groups or about the **size** of this difference**.** For this reason papers now usually report **“95% Confidence Intervals”** in preference to P values. The **“95% Confidence Interval”** describes the **uncertainty** around the treatment effect that we have found, due to the play of chance**.**

The other type of error that can occur when coming to a conclusion about the groups is called a **“Type Two Error”.** This is when we wrongly conclude that there is not a difference between the groups, when in fact there is. This is most likely to happen when the numbers in a trial are small, and therefore the **“95% Confidence Intervals”** are wide, and include the possibility of no difference between the groups.

To try to avoid a **“Type Two Error”,** investigators usually carry out a **“Power Calculation”,** but you have to guess how much difference there will actually be between the groups beforehand, so these are not always very reliable. Typically a **“Power Calculation”** might ask: how many patients do I need in the trial to have an 80% chance of showing a significant difference between the groups (using 95% CI), if we assume that the difference between groups is “X”?

In this case the **Standard Error (SE)** of the difference between the average of 625 red dice and 625 black dice comes out as 0.14, so we would expect the difference between the average score of the 625 red dice and the 625 black dice to be less than 1.96 times 0.14 either way, which is plus or minus 0.27.

If we compared 25 throws of each dice, the **SE** of the difference is much larger at 0.71, so we should not be surprised if the difference between their averages was as much as plus or minus 1.96 x 0.71 = 1.36 . This 95% CI is five times as small for 625 throws compared to 25 throws, because the number of throws is 25 times larger, and the **SE** is **SD** divided by the square root of the number in the sample.

In general the same applies to clinical trials, if 250 patients are recruited to a large trial and 10 to a small trial, the confidence interval will be 5 times as narrow in the larger trial than the smaller one.

# Systematic Reviews and Meta-analyses

In **narrative reviews** an expert can present any evidence that they wish. There is a therefore a temptation to present the evidence that supports the views of the expert concerned. In contrast a **systematic review** uses pre-defined methods (preferably written in a protocol before the review is started), to identify, assess, summarise and apply all the evidence available to answer a specific clinical question.

The **systematic review** may combine the evidence using a statistical method called **meta-analysis. Meta-analyses** combine the results of the trials together to find the average result from all the trials. In doing so the **power** is increased because the total number of patients goes up, and as a consequence the confidence interval of the average result is narrower than that of the individual trials. The person carrying out the **systematic review** must decide whether the trials are clinically similar enough to combine the results, or whether they are too **heterogeneous** (meaning too different in their results, populations, methods or outcomes) and then no **meta-analysis** is carried out.

The results of any **meta-analyses** are usually shown as a **Forest plot**. **Forest plots** show the results of each trial on a separate horizontal line. The **weight** of each trial is usually shown in a table, and is also displayed as the size of box for that trial on the **Forest plot**. The **95%** **confidence interval** for each trial is shown as a horizontal line (rather like the wings of an aeroplane). Larger trials generally have more weight (with bigger boxes) and narrower confidence intervals. However trials with a smaller number of events or large standard deviations in their outcome measurements will have proportionally less weight and wider confidence intervals in comparison to other trials of a similar size.

The average (or pooled result) of all the trials in a meta-analysis is shown as a diamond under the trial results, and in this case the **95%** **confidence interval** is represented by the whole width of the diamond. If the diamond does not cross the vertical “line of no-difference” then you know that the average result is **statistically significant** (that is the P value is less than 0.05).

# Analysing Continuous Data

If the data is normally distributed, then the mean of two groups of results in a **parallel design** controlled trial can be compared using a **“T Test”.** This makes allowance for small samples, and compares the mean of the two groups. The difference between the two means (or **mean difference**) should be zero if the null hypothesis is true, but even if there is no real difference between the mean of the whole population of the two groups, because samples are taken the mean difference will be expected to differ from zero due to the play of chance. The **standard error** of each mean is combined together to form a standard error of the mean difference, and this can then be used to work out the **95% confidence interval** for the mean difference. As usual, if the **95% confidence interval** includes no difference (zero) then the null hypothesis has not been disproved.

The **“T test”** adjusts for smaller samples and therefore the 95% confidence interval will be at least twice the standard error of the mean difference, and may be much larger for very small samples.

For cross-over trials, since each participant will receive the two treatments in a random order, each participant has a pair of readings. The results are analysed using these pairs and a “**Paired T Test”,** whereas a parallel group trial (where each participant only receives one of the treatments), compares just the mean data from each group and then an “**Unpaired T test”** is used as described above. Cross-over trials have the advantage of comparing each person with themself, thereby usually reducing the variability (**standard deviation**) of the results and smaller numbers of participants are needed. The disadvantage is that the cross-over design needs to be used in conditions that are fairly stable, and a **wash-out period** is needed between the treatments to prevent the treatment given first having a **carry-over effect** on the results in the second arm.

Another way of reducing the variability in the trial results is to adjust for baseline differences in the individual participants using “Analysis of Variance” (ANOVA) or “Analysis of Co-variance” (ANCOVA). The details of how this is done are beyond what you need to know for the AKT, but you may see these acronyms reported in the methods sections of trial reports.

# Analysing Dichotomous Data: Risk Ratios, Odds Ratios, Risk Differences and Numbers Needed to Treat

Below is a 2x2 table showing the results of admission to hospital in a clinical trial in COPD.

|  |  |  |
| --- | --- | --- |
|  | Pulmonary Rehab Group | Usual Care Group |
| Admitted to Hospital | **2** | **10** |
| Not Admitted to Hospital | **28** | **20** |

The risk of being admitted on Pulmonary Rehab was 2/30 and the risk of being admitted on Usual Care was 10/30. These two risks can be compared as a **Risk Ratio** or as a **Risk Difference. We always consider the risk in the active treatment group first** (Pulmonary Rehab in this case). The **Risk Ratio**, also described as the **Relative Risk** (conveniently both abbreviate to **RR), is then calculated by dividing the risk on active treatment by the risk on control.** So in this case 2/30 is divided by 10/30, and as the number in each group is the same, this simplifies to 2/10 or **RR = 0.2**. This can also be described as a **Relative Risk Reduction (RRR) of 80%.** This comes from saying that if the risk on usual care was 100%, this will fall to 20% with Pulmonary Rehab (applying **RR = 0.2 above).** The difference between these two figures is 80% so this is an **80%** **RRR.**

Risk Ratios (or Odds Ratios, which we will consider later) are useful when combining the results to find the average in all trials. The reason for this is that ratios tend to be consistent for treatment effects in high risk and low risk patients; an example is treatment with statins, which have been shown to have a **Risk Ratio** of around 0.73 across a range of populations. However the **Risk Difference** (RD) or Absolute Risk Reduction (ARR which is the same as RD) is dramatically different for high and low risk patients.

So the **Risk Difference** for the trial above is the Risk for Pulmonary Rehab minus that for usual care. This works out as 2/30 minus 10/30 which comes to -8/30 or -0.267. The risk is lower on Pulmonary Rehab, which gives the minus sign. Notice that if you express this as a percentage reduction the **Absolute Risk Reduction** **is -27%**, which is very much smaller than the **80%** **RRR** from the same trial**.** This is because the Relative Risk has to be applied to the risk in the usual care group, so an 80% reduction in the risk of 33% in the usual care group is where -27% comes from.

The **Number Needed to Treat** is the inverse for the Risk Difference or Absolute Risk Reduction. In this case the Risk Difference was -8/30 so the NNT is 30/8. NNT is always rounded up to the next whole number, so in this case 3.75 rounds up to an NNT of 4. Even 3.1 would round up to 4, because you cannot treat less than a whole patient!

The **Odds Ratio** comes from comparing those who were admitted to hospital with those who were not admitted on each treatment. The **Odds** of admission on Pulmonary Rehab are 2/28 (similar to the risk of 2/30) but for usual care the **Odds** of admission are 10/20 which is not the same as the risk of 10/30. The **Odds Ratio** is therefore 2/28 divided by 10/20 which is 4/28 or **0.14.** As a rule of thumb the **Odds Ratio** is similar to the **Risk Ratio** for rare events (under 10% for example) but can be very different for common events.

# Chi-squared Tests for Categorical/Dichotomous Data

Chi-squared tests can be used to test categorical or dichotomous data and test the null hypothesis that there is no difference between two treatments. Consider again the Pulmonary Rehabilitation Data in the 2x2 table, with the totals shown for each row and column.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Pulmonary Rehab Group | Usual Care Group | Row Total |
| Admitted to Hospital | **2** | **10** | **12** |
| Not Admitted to Hospital | **28** | **20** | **48** |
| Column Total | **30** | **30** | **60** |

The Chi-squared test is based on the difference between the observed and expected outcomes for the four data cells. The observed outcomes (O) are the results that were found in the trial. The expected outcomes (E) are calculated for each of the four data cells by multiplying the row total by the column total and dividing by the grand total.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Pulmonary Rehab Group | Usual Care Group | Row Total |
| Admitted to Hospital | **(12x30)/60** **E = 6** | **(12x30)/60** **E = 6** | **12** |
| Not Admitted to Hospital | **(48x30)/60** **E = 24** | **(48x30)/60** **E = 24** | **48** |
| Column Total | **30** | **30** | **60** |

Comparing the two tables we now work out the difference between the observed and expected value for each cell (O – E) and the Chi-squared test is the sum of (O-E)2 / E for each of the four cells.

Total of (O-E)2/E = 42/6 + 42/6 + 42/24 + 42/24 = **6.67**

This test can be done for any number of cells and rows, and the Chi-squared total has to be assessed for a P value using statistical tables against the degrees of freedom (which is calculated as rows minus one multiplied by columns minus one). For a 2x2 table this gives one degree of freedom. This data returns P = 0.01, so the result is statistically significant. The Chi-squared test does not estimate the size or direction of the treatment effect, it merely checks how likely these results would be if the null hypothesis was true (and there was actually no difference between Pulmonary Rehab and Usual Care).

# Appendix 1: Dichotomous Data Analysis Definitions

|  |  |  |
| --- | --- | --- |
|  | Intervention Group | Control Group |
| Admitted to hospital | **a** | **b** |
| Not admitted to hospital | **c** | **d** |
| Total | **a + c** | **b + d** |

Risk of admission on intervention (**EER**): a/(a+c)

Risk of admission on control (**CER**): b/(b+d)

Risk Ratio (= Relative Risk, RR) is: **EER/CER** = a/(a+c) **÷** b/(b+d)

Odds Ratio (OR) is: a/c ÷ b/d = ad/bc

Risk Difference (= Absolute Risk Reduction, RD, ARR) is: **EER - CER** a/(a+c) **minus** b/(b+d)

NNT is: 1/RD (or if RD is expressed as y% points, convert it to a fraction = y/100 then turn the fraction upside down 100/y and round up to next whole number)

Relative Risk Reduction (RRR) is: 1 minus Risk Ratio (or 100% in Relative Risk as a percentage) which is (**CER – EER)/CER**

# Appendix 2: Diagnostic Test Accuracy Definitions

The key to answering questions on Diagnostic Test accuracy is to create a 2x2 table as demonstrated below:

|  |  |  |
| --- | --- | --- |
|  | Disease is Present | Disease is Absent |
| Test is Positive | **True Positive** | **False Positive** |
| Test is Negative | **False Negative** | **True Negative** |

Sensitivity and Specificity are calculated by using the vertical columns, and Positive and Negative Predictive Values by using the horizontal rows.

## Definitions

#### Sensitivity is the proportion of patients with the disease correctly diagnosed.

#### Sensitivity = TP/(TP + FN)

#### Specificity is the proportion of patients without the disease who are correctly diagnosed.

####  Specificity = TN/(TN + FP)

#### Positive Predictive Value is the proportion of all the patients with a positive test result who actually have the disease

####  PPV = TP/ (TP +FP)

#### Negative Predictive Value is the proportion of patients with a negative test who are free from the disease

####  NPV = TN/(TN + FN)

*This section is additional material and is only for those who are interested!*

#### Positive Likelihood Ratio is sensitivity / (1 – specificity). This indicates how much more likely the patient is to have the disease following a positive test result.

####  Positive LR = TP/(TP + FN) divided by FP/(TN + FP)

Technical Note : The Positive LR must be applied to the pre-existing **odds** of the patient having the disease before the test was carried out. So if you think the likelihood that the patient has the disease is 50/50 before the test (so that the pre-test odds are 1:1), and the test with a Positive LR of 9 turns out to be positive, then the post-test odds are 9:1. This means his new likelihood of having the disease following a positive test is 9/(9+1) = 90%.

#### Negative Likelihood Ratio is (1 – sensitivity) / specificity. This indicates how much less likely the patient is to have the disease following a negative test result.

####  Negative LR = FN/(TP + FN) divided by TN/(TN + FP)

The Wikipedia definitions of Likelihood Ratios are as follows:

Two versions of the likelihood ratio exist, one for positive and one for negative test results. Respectively, they are known as the **positive likelihood ratio** (LR+, **likelihood ratio positive**, **likelihood ratio for positive results**) and **negative likelihood ratio** (LR–, **likelihood ratio negative**, **likelihood ratio for negative results**).

The positive likelihood ratio is equivalent to "the probability of a person who has the disease testing positive divided by the probability of a person who does not have the disease testing positive."

The negative likelihood ratio is "the probability of a person who has the disease testing negative divided by the probability of a person who does not have the disease testing negative."

The [pretest odds](https://en.wikipedia.org/wiki/Pre-_and_post-test_probability#Pre-test_probability) of a particular diagnosis, multiplied by the likelihood ratio, determines the [post-test odds](https://en.wikipedia.org/wiki/Pre-_and_post-test_probability). This calculation is based on [Bayes' theorem](https://en.wikipedia.org/wiki/Bayes%27_theorem). (Note that odds can be calculated from, and then converted to, [probability](https://en.wikipedia.org/wiki/Probability).)