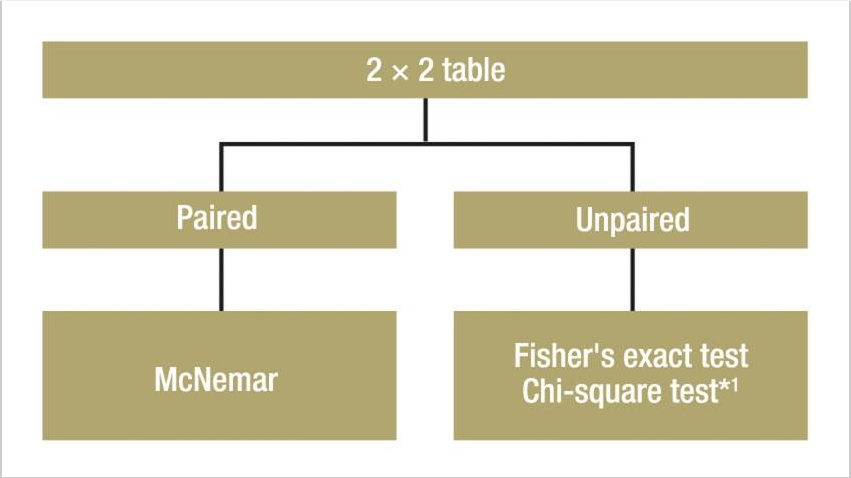
**Frequently used statistical tests**

|  |  |
| --- | --- |
| Statistical Test | Description |
| Fisher’s exact test | Suitable for binary data in unpaired samples:  the 2 x 2 table is used to compare treatment effects or the frequencies of side effects in two treatment groups (unpaired samplle) |
| Chi-square test | Similar to Fisher’s exact test (albeit less precise). Can also compare more than two groups or more than two categories of the outcome variable. Preconditions:   1. Sample size > 60. 2. Expected number in each field ≥5. |
| McNemar test | Preconditions similar to those for Fisher’s exact test, but for “paired samples” |
| Student’s t-test | Test for “continuous data”. Investigates whether the expected values for two groups are the same, assuming that the data are normally distributed. The test can be used for “paired “or “unpaired” groups. |
| Analysis of variance | Test preconditions as for the “unpaired t-test”, for comparison of more than two groups. The methods of analysis of variance are also used to compare more than two paired groups. |
| Wilcoxon’s rank sum test  (also known as the unpaired Wilcoxon rank sum test or the Mann-Whitney U test) | Test for ordinal or continuous data. In contrast to Student’s t-test, does not require the data to be normally distributed. This test too can be used for “paired” or “unpaired” data. |
| Kruskal-Wallis test | Test preconditions as for the “unpaired” Wilcoxon rank sum test for comparing more than two groups |
| Friedman test | Comparison of more than two paired samples, at least ordinally scaled data |
| Log rank test | Test of survival time analysis to compare two or more independent groups |
| Pearson correlation test | Tests whether two continuous normally distributed variables exhibit linear correlation |
| Spearman correlation test | Tests whether there is a monotonous relationship between two continuous, or at least ordinal, variables |

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2881615/#R11>

**Group comparison of two categorical endpoints**



Unpaired samples:

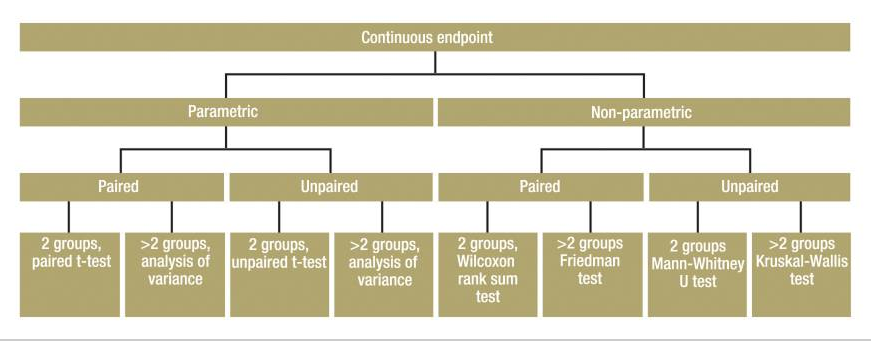
If the frequency of success in two treatment groups is to be compared, Fisher’s exact test is the correct statistical test, particularly with small samples. For large samples (about n >60), the chi-square test can also be used

Paired samples:

One example of the use of this test would be an intervention within a group at two anatomical sites, such as the implantation of two different sorts of IOL lenses in the right and left eyes, with the endpoint “Operation successful: yes or no.” The samples to be compared are paired. In such a case, one has to perform the McNemar test

**Continuous and at least ordinally scaled variables**

*Normally distributed variables* — parametric tests: So-called parametric tests can be used if the endpoint is normally distributed.



*Unpaired samples*:

Where subjects in both groups are independent of each other (persons in first group are different from those in second group), and the parameters are normally distributed and continuous, the unpaired t-test is used. If a comparison is to be made of a normally distributed continuous parameter in more than two independent (unpaired) groups, analysis of variance (ANOVA) can be used. One example would be a study with three or more treatment arms.

ANOVA is a generalization of the unpaired t-test. ANOVA only informs you whether the groups differ, but does not say which groups. This requires methods of multiple testing.

*Paired samples*:

The paired t-test is used for normally distributed continuous parameters in two paired groups. If a normally distributed continuous parameter is compared in more than two paired groups, methods based on analysis of variance are also suitable. The factor describes the paired groups—for example, more than two points of measurement in the use of a therapy.

*Non-normally distributed variables* — non-parametric tests: If the parameter of interest is not normally distributed, but at least ordinally scaled, non-parametric statistical tests are used. One of these tests (the “rank test”) is not directly based on the observed values, but on the resulting rank numbers. This necessitates putting the values in order of size and giving them a running number. The test variable is then calculated from these rank numbers. If the necessary preconditions are fulfilled, parametric tests are more powerful than non-parametric tests. However, the power of parametric tests may sink drastically if the conditions are not fulfilled.

Unpaired samples:

The Mann-Whitney U test (also known as the Wilcoxon rank sum test) can be used for the comparison of a non-normally distributed, but at least ordinally scaled, parameter in two unpaired samples. If more than two unpaired samples are to be compared, the Kruskal-Wallis test can be used as a generalization of the Mann-Whitney U test.

Paired samples:

The Wilcoxon signed rank test can be used for the comparison of two paired samples of non-normally distributed, but at least ordinally scaled, parameters. Alternatively, the sign test should be used when the two values are only distinguished on a binary scale—for example, improvement versus deterioration. If more than two paired samples are being compared, the Friedman test can be used as a generalization of the sign test.