Misuse of a two-sample test of the mean and its effect on the p-value.

Most of the errors arise from two critical areas:

First, the basic assumptions of the sample population are not met
There are 3 basic assumptions:
(1) The variances (and hence the standard deviations) of each sample are nearly equal
(2) The populations from which the samples are drawn are normally distributed
(3) Each sample is independent of the other (ie, samples are drawn from independent populations)

Second, the *t* test is used for multiple comparisons (ie, comparing more than two groups or using the *t* test multiple times for different comparisons)

Each additional test compounds the probability of finding a "significant" result by chance alone, inflating the p-value and making results unreliable. For example, comparing one control group to four others leads to a nearly 19% chance of a false positive rather than the intended 5%.

- Suppose we have one control group and want to compare it to four other groups using the t-test. We'd be running four separate tests (one for each comparison with the control). Even though each individual test has a 5% chance of error, the combined chance of at least one false positive across the four tests is higher than 5%.

To show calculation:

- If each test has a 95% chance of being correct (1 - 0.05), the probability that all four tests are correct is $0.95^4 \approx 0.81$.
- So, the chance of at least one of the four tests being a false positive is $1-0.81=0.19$ or 19%. This means we're more likely to find a false positive when running multiple tests.

Comments on the chart:
The table shows that Hypericum perforatum does not significantly reduce ADHD symptoms compared to a placebo. Both groups experience similar score reductions in hyperactivity, inattentiveness, and total symptoms by week 8, but high p-values (all > 0.05) indicate no statistically significant difference between them. This suggests that the herbal treatment offers no clear advantage over the placebo.