No data were missing, but one observation of a diastolic blood pressure of 66 was considered unlikely and this patient was removed from the dataset. To ensure that all investigated patients were sufficiently exposed to the treatment, only participants with an adherence of at least 80% were selected for analysis.[[1]](#footnote-1)1 This led to a further removal of 29 patients from the dataset.

Two variables, the diastolic blood pressure measured at the last visit (denoted *dbp\_end*) and the change in blood pressure from baseline (denoted *dbp\_dif)*, were investigated for the first treatment arm and the placebo arm, which consisted of 39 and 44 patients, respectively. Although the QQ-plots of the variables revealed minor deviations from normality, the sample sizes were considered large enough for the Central Limit Theorem to apply and to assume that the sample means are normally distributed. The two-sample T-test was therefore considered a suitable test. The confidence interval for the F-ratio test of the variances was wide and only just contained one for both variables. As random simulations based on these data might result in different variances, the Welch two-sample T-test was chosen for the simulation study. The test was applied one-sided, with alternative hypothesis , where is the mean of treatment group and that of the placebo group.

The power of the T-test was estimated via 10 000 Monte Carlo simulations for each of the two variables. The treatment and placebo groups were simulated by random samples from a normal distribution with the mean and standard deviation of the respective group and were then compared via the Welch T-test. The power was determined as the percentage of p-values below 0.05. This resulted in a power of 100% for each variable. For more informative power estimates, the procedure was repeated 40 times, with a decrease in the difference between the means of the treatment arms of 2.5% each time. The powers of the T-test for the two variables is plotted in Figure 1.

*Figure 1: power of the Welch T-test when applied to the variables* dbp\_dif *(blue solid line) and* dbp\_end *(red dashed line) for multiple differences in means between the two treatment arms. These differences in means are expressed as a percentage of the observed difference in the data.*

This figure shows that the variables start at a similar power of 5%, corresponding to the significance level of 0.05, when there is no difference in means and converge to a power of 100% when the difference in means is 75% or more of the observed differences. However, when the means differ 20% - 75% of their observed differences, the test using the variable *dbp\_end* has more power than the test applied to *dbp\_dif*.

To explain the discrepancy in power of the Welch T-test the following power function was used:

Where , and the non-centrality parameter is defined as:

This function was applied to a difference in means () decreased to 42.5% of the observed difference, as simulating with this led to the largest difference in power. The power is 70% for *dbp\_dif* and 74% for *dbb\_end*.

As , and are fixed, the difference in power can either be due to a different Δ or different values of and for the variables. As a higher increases the power, the variable with the largest and smallest variances should give the highest power. In this case, *dbp\_dif* has smaller values for and , but *dpb\_end* has a larger . The larger here has outweighed the smaller variances, leading to a higher power for *dbp\_end*.

1. 1 Valgimigli M, Garcia Garcia H, Vrijens B *et al*. Standardized classification and framework for reporting, interpreting, and analysing medication non-adherence in cardiovascular clinical trials: a consensus report from the Non-adherence Academic Research Consortium (NARC). *European Heart Journal* 2018; **00**: 1-16. [↑](#footnote-ref-1)