Statistical Analysis Plan

Needling Techniques for Tonifying Kidneys and Dredging Meridians for Knee Osteoarthritis: A Randomized Clinical Trial

NCT05014542

Document date: March 11, 2025

Analyses should be performed on 64 participants, randomised for the intention-to-treat (ITT) analysis using the maximum available data set (i.e. excluding the missing values) from repeated measures of variables of interest and their variances. Data will be collected from validated questionnaires, and a DRUG variable was constructed.

Trials could be prone to dropouts. They were expected in this trial due to the high number of planned treatments and long study period. ITT analysis excluding missing values was selected regarding dropouts that emerged in the study completely random without any causal clustering.[154] Dropouts produced missing values. Omitting missing values can compromise the statistics and produce bias. Sensitivity analyses will be provided to compare those who dropped out and those who completed the study to find if data significantly differed.

Sensitivity analysis compared completers and non-completers from baseline (assessment 1) with the assumption that if both do not differ mutually significantly then it could be inferred that non-completers would gain similar scores as completers if they completed the trial. Also, test calculation with the Last observation carried forward technique (LOCF) technique with imputed last observed value in missing one was performed at the time-point of the biggest participant attrition in group A to check if obtained effects still stand. Sensitivity analyses found consistency with data of the primary analysis, which suggests accepting the treatment effect of the primary analysis as valid, therefore the missing values had no bigger impact on the primary analysis conclusions, the primary analysis produces the robust data and the primary results still stand.

The Shapiro-Wilk test will test the normality distribution of all study measures. The normality distribution will be tested to select an appropriate test for further data processing. All statistical tests were two-sided. For gender at baseline, the significant difference between groups will be tested by Chi-square test and the Student t-test for age, BMI, and K-L grade regarding the observed categorical or numeric variables. The non-parametric Mann-Whitney U test will test with 95% power and $\alpha = 0,05$ if a significant difference exist when two groups' data are compared, regardless if parametric or non-parametric.

The first study part will compare interventional group A which will receive acupuncture adjunctive to analgesics with control group (C) which receive analgesics only. 9 assessments will provide observed data from baseline to Week 24 for WOMAC total and subscales, NRS, KDSQ, and DRUG. Variables' values of all assessed time points will be compared between groups to find if data at specified time points significantly differed.

The second study part will last from Week 25 to 39, and will start with the crossover of group C to receive the same acupuncture treatment as A before. Added assessment at Week 39 will compare groups with their baseline by within-group analysis. The Mann-Whitney U test tested if significant differences exist within groups' two various time points, separately for each group.

Repeated measures analyses will present data and calculations for all 10 assessments.

The main efficacy indicators will test the comparability of the groups with the Mann-Whitney U test and quanify effects by mean differences and Cohen's d.

The testing of the existence of significant differences or that is observed by chance by Mann Whitney U test, with α=0,05 and 95% statistical power is a prerequisite to test the significance before considering the meaningness of achieved effects. Also when the relatively small sample size of a study affects significance testing, quantification of effects is suggested.[155,156]

Mean differences will be introduced to compare achieved differences with previous trials[157] with similar study designs using WOMAC as an outcome measure, and to calculate the reduction of effect or its persistence where appropriate. Mean differences will be presented as absolute or relative: to observe absolute differences or percentage of lost or kept effect accordingly, either between or within groups.

Cohen's d will present with calculated standardised mean differences (SMD) effect intensity. Cohen explains that SMD signifies a small therapeutic effect if $\geqslant$0.2, medium $\geqslant$0.5, and large $\geqslant$0.8. It was a suitable absolute outcome measure for continuous variables. Its calculation adjusts two compared sample differences from their pooled SD for the

precision of scale measurement and the sample size used. SMD increases with increased differences between treatments/groups and the measurement precision or lower SD. Its relation to the Number Needed to Treat (NNT) signifies when the SMD=1 then the NNT=2 means that of two treated patients one will have a good outcome, and if the SMD=2 almost all.[155,156]

The DRUG variable had outliers which influence the statistic, they will not be excluded from the analysis because they do not represent the clerical errors, and do not emerge from a different population than the sample.

Compliance of participants to treatment attendance and their withdrawal will be thoroughly presented.