

Development of methods and algorithms for diet design based on gut microbiota analysis

Statistical analysis

Number of clinical trial: 1631

Project number: NSP67

Approved by Ethical Committe of National Institute for Health Development

25.10.2016

Statistical analysis

Based on sample size calculations, we estimated that with 17 participants, the study would have more than 80% power to detect a significant difference among weight loss study group, assuming a mean BMI reduction by 3 kg/m², with a mean BMI and standard deviation of 35 and 3.2 kg/m², respectively, at an alpha level of 5%.

Statistical analysis included bacteria with average colonization frequency > 70% and average abundance > 0.001. Analysis of data was carried out in R statistical programming language, version 3.5.0 (28). The resulting p-values were corrected for multiple comparisons for each phylogenetic level using Benjamini-Hochberg correction (FDR). A corrected p-value < 0.1 was considered statistically significant. Unless stated otherwise, corrected P values are shown in the text.

Pairwise comparisons were evaluated using Wilcoxon signed-rank test, for the comparison of test and reference groups Kruskal-Wallis test was applied.

To control for within-subject variability, we used the subsequent sample-pairs as within-subject controls and compared β -diversity before and after the intervention. This was also applied to reference group samples. The following cutoffs were used: *, p < 0.05; **, p < 0.01; ***, p < 0.001; ****, p < 0.0001.

Agglomerative hierarchical clustering

Ward's agglomerative hierarchical clustering on a distance matrix was generated from a species by sample Bray-Curtis distance matrix. The method produces a dendrogram by treating each sample as a singleton cluster, merging pairs of clusters until all clusters have been merged into one big cluster containing all samples. Ward's agglomeration method minimizes the total within-cluster variance.