Statistical analysis plan

for the "Clinical study on Bifidobacterium lactis BL-99 assisting to improve functional dyspepsia"

This study includes 2 sub-studies, which correspond to different research purposes. Therefore, the following statistical analysis plans are presented separately according to sub-studies.

Sub-study 1

1. Research purpose

(1) Main purpose

To evaluate the effect of *Bifidobacterium lactis* BL-99 （BL-99） probiotics intervention on symptoms of functional dyspepsia (FD) in FD population.

(2) Secondary purpose

To study the effect of BL-99 probiotics intervention on gut microbiota and metabolites of FD population.

2. Research design

2.1 Overall design

This study is a randomized, open, parallel controlled study involving 4 intervention groups.

Randomization method: A computer-generated list of random numbers is used to randomly assign the participants to the 4 groups.

2.2 Group and Intervention

Intervention groups	**Intervention substances**	**Intervention time**
(1) Mild placebo group	2 g/day maltodextrin	8 weeks
(2) Mild low dose group	2 g solid beverages containing 1×10^{10} CFU/day BL-99	8 weeks

(3) Mild high dose group	2 g solid beverages containing 5×10^{10} CFU/day BL-99	8 weeks
(4) Severe placebo group	10 mg/day proton pump inhibitors (PPI)	8 weeks

2.3 Time Taken

After a 2-week run-in period, the participants underwent an 8-week intervention period; and an 8-week follow-up period. There will be two intervention visits, one at 4 weeks of intervention and one at 8 weeks of intervention. After the intervention, in order to evaluate the long-term effect of probiotic intervention on the improvement of dyspepsia symptoms, follow-up will be continued for 8 weeks, and 2 visits will be set at 2 weeks and 8 weeks post the intervention, respectively.

3. Outcome indicators

3.1 Main outcome indicators

Clinical response rate of PDS+EPS score at 8 weeks of intervention. Clinical response rate was defined as a score reduction of ≥0.5 from pre-intervention. (PDS: postprandial distress syndrome; EPS: epigastric pain syndrome)

3.2 Secondary outcome indicators

- Clinical response rate of PDS+EPS at 4 weeks of intervention, 2 weeks after the intervention, and 8 weeks after the intervention.
- Clinical response rate of PDS and EPS scores at 4 weeks of intervention, at 8 weeks of intervention, 2 weeks after the intervention, and 8 weeks after the intervention.
- Changes of serum index values reflecting gastric digestion ability at 8 weeks of intervention, and 2 weeks after the intervention, mainly including serum pepsinogen I (PG I), pepsinogen II (PG II), pepsinogen ratio (PGR) = PG I/PG II, and gastrin 17 (G17).
- Gut microbiota at 8 weeks of intervention.
- Faecal metabolites at 8 weeks of intervention.
- Fecal and Serum Short chain Fatty Acids (SCFA) at 8 weeks of intervention.

3.3 Safety indicators

Adverse event rate: Participants are asked to report any adverse effects during the treatment and follow-up periods, such as bloating, nausea, diarrhea, itchy skin, etc. Safety is assessed by classifying adverse events using the Common Terminology Criteria for Adverse Events (CTCAE) version 5.0 at each study period or in the case of early termination.

4. Sample size calculation

Based on previous studies, the power for the primary outcome reached 80% (two-sided test, $\alpha = 0.05$) with a sample size of 42 participants per group. This study assumes the overall dropout rate to be 10-12% during the entire study period. To account for probable follow-up loss, at least 200 participants will be recruited.

5. Data sets of statistical analysis

(1) Per-Protocol Set (PPS)

The main data set for efficacy analysis in this study is PPS. PPS refers to all participants that have completed the planned intervention and visits according to the protocol and have no obvious effect on the therapeutic effect. Violations that significantly affect efficacy are determined at the time of data review and may include (but are not limited to) the following: ① failure to meet inclusion criteria; ② there was interference therapy after inclusion; ③ poor compliance; ④ follow-up beyond the window period.

(2) Safety Set (SS)

The main data set for safety analysis in this study is SS. Safety evaluation data of participants in 8-week intervention and 8-week follow-up period constituted the SS of this study.

6. Statistical analysis

6.1 Software

Statistical analysis of clinical indicators will be performed using SPSS Statistics 24 (SPSS Institute, Chicago, IL, USA). Figures other than those related to microbial analysis will be created using GraphPad Prism 9.0.0. Fecal metabolites will be processed for peak pick and deconvolution with Unknowns Analysis tool of the

MassHunter Quantitative Analysis software package (B.10.1, Agilent Technologies). Mass Profiler Professional Software (MPP) (version 14.5, Agilent Technologies) will be used for alignment, normalization and annotation. Metagenomics analysis will be performed using the online platform of Majorbio Cloud Platform (www.majorbio.com). And fastp v0.20.0 (https://github.com/OpenGene/fastp), MEGAHIT v1.1.2 (https://github.com/voutcn/megahit) and MetaGene (http://metagene.cb.k.u-tokyo.ac.jp/) software will be used for statistics and quality control of raw sequencing data, assembly of sequencing data and gene prediction. NR (nr_20200604, https://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/), eggnog (v4.5.1, http://eggnog5.embl.de/#/app/downloads), KEGG (v94.2, https://www.genome.jp/kegg) and CAZy (v8, http://bcb.unl.edu/dbCAN2/download/Data) will be used for species, COG, KEGG, and CAZy annotation.

6.2 General principles

All hypothesis tests are two-sided. In general, $P < 0.05$ is considered significant, except for multiple comparisons, where the threshold of significance is 0.008. In the analysis of intestinal microorganisms and metabolites, the level of significance is further set at $P < 0.05$, *; $P < 0.01$, **; $P < 0.001$, ***; $P < 0.0001$, ****.

6.3 Subject enrollment and completion summary

A flowchart will be used to summarize the enrollment and completion of the study.

6.4 Description and comparison of baseline characteristics

Baseline demographic characteristics (age, sex, body mass index) and clinical characteristics of FD (bloating, early satiety, Epigastric pain, Epigastric burning, PDS, EPS, and Total score) will be described between groups. Continuous variables are described as mean and 95% confidence interval (95%CI). Counting variables are described as frequency and percentage.

For comparison between groups, One-way analysis of variance (ANOVA) or Kruskal-Wallis rank test is used for continuous variables, and chi-square test is used for counting variables.

6.5 Effect and safety analysis after intervention

(1) Clinical response rate of FD

Clinical response rates of PDS+EPS, PDS and EPS scores are calculated for each group at 4 weeks and 8 weeks of intervention, 2 weeks and 8 weeks of post-intervention follow-up. Chi-square test is used to compare the differences in response rates between the groups, and logistic regression is used to calculate the relative risk (RR) and 95% confidence interval (95%CI).

In order to explore whether the effects of probiotic interventions differ between gender groups, analyses of clinical response rates will be conducted separately in the general population, in men, and in women.

(2) Serum indexes reflecting gastric digestibility

The change values of each index from baseline to 8 weeks of intervention (△1), and from 8 weeks of intervention to 2 weeks post the intervention (△2) are described respectively, and least-squares means and 95%CIs are calculated.

Comparisons of △1 and △2 between groups are performed by ANOVA. If the overall difference between the groups is significant, least significant difference (LSD) method was used for multiple comparison. The statistical significance level of P-values for multiple comparisons is set at $P<0.008$ using Bonferroni correction.

(3) Analysis of intestinal microorganisms and metabolites

ANOVA with LSD method is used to analyze the differences of α diversity indexes, the relative abundance of phyla and species, and the microbiome function among the four groups. Non-parametric test with Kruskal-Wallis is applied to detect differences in the un-target metabolites among the four groups. Paired T test is used to analyze the significance of short chain fatty acids before and after intervention. And the correlations between the relative abundance of species and short chain fatty acids are assessed by Spearman’s correlation analysis.

(4) Safety analysis

Frequency and percentage are used to describe the incidence of various adverse events in each group.

Sub-study 2

1. Research purpose

To study the improvement effect of BL-99 probiotic intervention on drug-induced gut microbiota disturbance in functional dyspepsia (FD).

2. Research design

2.1 Overall Design

This is a randomized, double-blind, parallel-controlled study involving 2 intervention groups.

Randomization: A computer-generated list of random numbers is used to randomly assign the participants to the 2 intervention groups.

Blinding: Products blinding will be achieved by preparing probiotics as solid beverages with similar packaging, smell, and taste. Independent statisticians will analyze the data without knowing the specific allocation of interventions.

2.2 Group and intervention

Intervention groups	Intervention substances	Intervention time
(5) Severe low dose group	2 g solid beverages containing 1×10^{10} CFU/day BL-99 and10 mg/day PPI	8 weeks
(6) Severe high dose group	2 g solid beverages containing 5×10^{10} CFU/day BL-99 and10 mg/day PPI	8 weeks

2.3 Time Taken

After a 2-week run-in period, the participants underwent an 8-week intervention period; and an 8-week follow-up period. There will be two intervention visits, one at 4 weeks of intervention and one at 8 weeks of intervention. After the intervention, in order to evaluate the long-term effect of probiotic intervention on the improvement of gut microbiota disturbance, follow-up will be continued for 8 weeks, and 2 visits will be set at 2 weeks and 8 weeks post the intervention, respectively.

3. Outcome indicators

3.1 Main outcome indicators

Gut microbiota at 8 weeks of intervention.

3.2 Secondary outcome indicators

- clinical response rate of PDS+EPS, PDS and EPS scores at 4 weeks of intervention, at 8 weeks of intervention, and 2 weeks after the intervention (PDS: postprandial distress syndrome; EPS: epigastric pain syndrome).
- changes of serum index values reflecting gastric digestion ability at 8 weeks of intervention, and 2 weeks after the intervention, mainly including serum pepsinogen I (PG I), pepsinogen II (PG II), pepsinogen ratio (PGR) = PG I/PG II, and gastrin 17 (G17).
- gut microbiota 2 weeks after the intervention.
- fecal metabolites at 8 weeks of intervention.

4. Sample size calculation

The main outcome of this study was the gut microbiota after intervention. There is no suitable reference for sample size calculation. Considering the consistency of the included population, the sample size of Sub-study 2 was determined in accordance with that of Sub-study 1, that is, 50 subjects were planned to be enrolled in each group.

5. Data set of statistical analysis

Per-Protocol Set (PPS)

The main data set for efficacy analysis in this study is PPS. PPS refers to all participants that have completed the planned intervention and visits according to the protocol and have no obvious effect on the therapeutic effect. Violations that significantly affect efficacy are determined at the time of data review and may include (but are not limited to) the following: ① failure to meet inclusion criteria; ② interference therapy after inclusion; ③ poor compliance; ④ follow-up beyond the window period.

6. Statistical analysis

6.1 Software

Figures other than those related to microbial analysis will be created using GraphPad Prism 9.0.0. Fecal metabolites will be processed for peak pick and deconvolution with Unknowns Analysis tool of the MassHunter Quantitative Analysis software package (B.10.1, Agilent Technologies). Mass Profiler Professional Software (MPP) (version 14.5, Agilent Technologies) will be used for alignment, normalization and annotation. Metagenomics analysis will be performed using the online platform of Majorbio Cloud Platform (www.majorbio.com). And fastp v0.20.0 (https://github.com/OpenGene/fastp), MEGAHIT v1.1.2 (https://github.com/voutcn/megahit) and MetaGene (http://metagene.cb.k.u-tokyo.ac.jp/) software will be used for statistics and quality control of raw sequencing data, assembly of sequencing data and gene prediction. NR (nr_20200604, https://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/), eggnog (v4.5.1, http://eggnog5.embl.de/#/app/downloads), KEGG (v94.2, https://www.genome.jp/kegg) and CAZy (v8, http://bcb.unl.edu/dbCAN2/download/Data) will be used for species, COG, KEGG, and CAZy annotation. Statistical analysis of clinical indicators will be performed using SPSS Statistics 24 (SPSS Institute, Chicago, IL, USA).

6.2 General principles

All hypothesis tests are two-sided. In general, $P < 0.05$ is considered significant. In the analysis of intestinal microorganisms and metabolites, the level of significance is further set at $P < 0.05$, *; $P < 0.01$, **; $P < 0.001$, ***; $P < 0.0001$, ****.

6.3 Subject enrollment and completion summary

A flowchart will be used to summarize the enrollment and completion of the study.

6.4 Description and comparison of baseline characteristics

Baseline demographic characteristics (age, sex, body mass index) and clinical characteristics of FD (bloating, early satiety, Epigastric pain, Epigastric burning, PDS, EPS, and Total score) will be described between groups. Continuous variables are

described as mean and 95% confidence interval (95%CI). Counting variables are described as frequency and percentage.

For comparison between groups, independent t test or Wilcoxon rank test is used for continuous variables, and chi-square test is used for counting variables.

6.5 Effect analysis after intervention

(1) Analysis of intestinal microorganisms and metabolites

Independent t test is used to analyze the differences of α diversity indexes, the relative abundance of phyla and species, and the microbiome function between the 2 groups. Wilcoxon rank test is applied to detect differences in the un-target metabolites between the 2 groups. Paired t test is used to analyze the significance of short chain fatty acids before and after intervention. And the correlations between the relative abundance of species and short chain fatty acids are assessed by Spearman's correlation analysis.

(2) Clinical response rate of FD

Clinical response rates of PDS+EPS, PDS and EPS scores are calculated for each group at 4 weeks and 8 weeks of intervention, 2 weeks and 8 weeks of post-intervention follow-up. Chi-square test is used to compare the differences in response rates between the groups, and logistic regression is used to calculate the relative risk (RR) and 95% confidence interval (95%CI).

(3) Serum indexes reflecting gastric digestibility

The change values of each index from baseline to 8 weeks of intervention (△1), and from 8 weeks of intervention to 2 weeks post the intervention (△2) are described respectively, and least-squares means and 95%CIs are calculated. Comparisons of △1 and △2 between groups are performed by independent t test.