Confidential

ICD-01. Randomized, controlled, open-label phase III clinical trial in patients with primary or recurrent Clostridioides difficile (CD) infection to evaluate the efficacy and safety of freeze-dried fecal microbiota capsules vs Fidaxomicin. Statistical Analysis Plan. Version 3.0. (06/10/2023)

ICD-01. Randomized, controlled, open-label phase III clinical trial in patients with primary or recurrent Clostridioides difficile (CD) infection to evaluate the efficacy and safety of freeze-dried fecal microbiota capsules vs Fidaxomicin. Statistical Analysis Plan. Version 3.0. 03/10/2023.

CONTENTS

1. Introduction.

This document describes the statistical analysis plan (SAP) foreseen for the ICD-01 study. The proposed analyses use as a reference what is stated in the study protocol (version 7, dated 04/08/2023), regarding the objectives. The objectives of the sponsor and the PI will also be taken into account in order to arrive at a final version of this SAP.

Once approved by the promoter, the statistical analysis will adhere to the contents of this document. Any additional request not included in this document may entail an additional budget from Sermes CRO.

Optionally, and in order to ensure the repeatability of the results obtained, the compendium of the statistical syntax used will be attached to the final statistical report if the software used is governed by a programming language.

The recommended statistical software for statistical analysis is R (R Core Team, 2023) with the R Studio interface (version 2023.06.2). Apart from the installed functions such as *Rbase,* and among other available data processing and analysis packages, it is suggested to use the ones described in the table below:

Recommended R packages for data processing and analysis.

Package	Version	Target	References
readxl	1.4.3	Reading Excel files	Wichham and Bryan (2019)
psych	2.3.6	Descriptive analysis	Revelle (2023)
ggplot2	3.4.3	Graphic design	Wickham (2016)
coin	1.4-2	Non-parametric tests	Hothorn et al. (2006, 2008)
gmodels	2.18.1	Contingency table analysis	Warnes et al. (2018)
survival	3.5-7	Survival analysis	Therneau (2021)
survminer	0.4.9	Survival charts	Kassambara et al. (2021)
sqldf	0.4-11	SQL Queries	Grothendieck, G. (2017).

For any questions regarding the Statistical Analysis Plan, the SERMES CRO Biometrics Department is at your disposal.

1.1. Objectives of the study.

Note: The objectives are specified in the protocol (page 14).

1.1.1. Main objective.

The main objective is to evaluate the efficacy of fecal microbiota transfer compared to the control (Fidaxomicin) at 8 weeks after initiation of treatment.

1.1.2. Secondary objectives.

Secondary objectives involve assessing the safety of the investigational drug and the quality of life of the patients participating in the study.

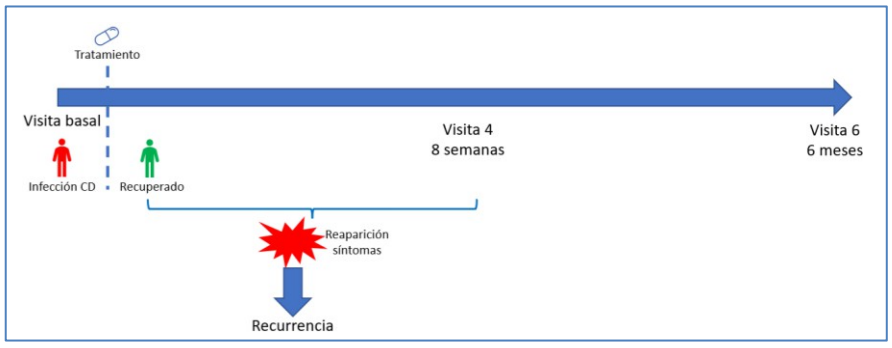
1.2. Study variables.

Note: The study variables are described in the protocol (pages 14-15).

1.2.1. Efficiency variables.

The primary variable is the **absence of diarrhea** absence of CD diarrhea without retreatment (or absence of recurrence as defined in Section 03.07.02 of the protocol), as assessed by subject interview and physical examination. The primary variable is operationalized by the probability of recurrence.

Recurrence is defined as the reappearance of clinical manifestations of a new episode of CDI that recurs within 8 weeks after the onset of symptoms of a previous episode, provided that the symptoms of the previous episode had resolved after completing treatment (Debast et al., 2013).



Important note. A patient who recurs will be withdrawn from the study. Failure of a patient to be effective is defined as the occurrence of a new recurrence of CD infection. Patients with a new recurrence will be withdrawn from the study and treated with rescue medication, which will be antibiotic treatment (vancomycin or fidaxomicin).

Other efficacy variables:

- **Good/bad evolution of the patient:** Bad evolution is understood as the detection 48-72 hours after the start of treatment (MBK-01 or Fidaxomicin) of a worsening of the diarrhea episode (at least one stool more than at baseline*, with baseline being understood as the time of the start of study treatment (MBK-01 or Fidaxomicin). In addition, at least one of the following factors must be detected:
    1. **Elevation of C-reactive protein (CRP) value (**>5% of baseline*).
    2. **Elevation of the absolute value of leukocytes (** >5% of the value a t baseline*).
    3. **Progression to sepsis:** hypotension or organ failure with no other apparent cause.

  ***Note**: baseline is considered to be the time of initiation of study treatment (Fidaxomicin /MBK-01).

- **Time to recurrence according to randomization groups**: Recurrence is defined as the reappearance of clinical manifestations of a new CDI episode in a patient with a treated and cured CDI episode within 8 weeks after the end of treatment. If the reappearance of a new CDI episode is later than 8 weeks, it would be a possible case of reinfection.

- **Overall survival (OS):** percentage of patients who are still alive after a defined period of time since starting treatment. This will take into account the start date of treatment and the end date of the study if the patient has not died and the date of exitus if the patient has died. If the patient leaves the study prematurely, the last available visit date will be considered.

1.2.2. Safety variables.

- Number of Adverse Events per randomization group.
- Number of patients within each Adverse Event per randomization group.
- Type of Adverse Events by randomization group.
- Number of Adverse Events of Special Interest by randomization group.
- Number of patients within each Adverse Event of Special Concern per randomization group.

- Type of Adverse Events of Special Interest by randomization group.
- Number of Serious Adverse Events by randomization group.
- Type of Serious Adverse Events by randomization group.
- Adverse events related to treatment.
- Severity of Adverse Events.
- Adverse Events in relation to ICD.
- Mortality associated with CDI.
- Admissions to the Intensive Care Unit (ICU).
- Adverse Events of Special Interest (AESI) by frequency of occurrence.

1.2.3. Quality of life variables.

Quality of life is measured by the **SF-36 questionnaire** which consists of 36 questions (items) that assess both positive and negative states of quality of life.

These 36 items cover the following areas:

- **Physical function**: Degree to which health limits physical activities such as self-care, walking, climbing stairs, bending, picking up or carrying weights, and moderate to strenuous exertion.
- **Physical Role**: Degree to which physical health interferes with work and other daily activities, including less than desired performance, limitation in the type of activities performed, or difficulty in performing activities.
- **Bodily pain**: The intensity of the pain and its effect on the usual work, both away from home and at home.
- **General health**: Personal assessment of health that includes current health, future health prospects and resistance to illness.
- **Vitality**: Feeling of energy and vitality, as opposed to feeling tired and exhausted.
- **Social function**: Degree to which physical or emotional health problems interfere with usual social life.
- **Emotional role**: Degree to which emotional problems interfere with work or other daily activities, including reduced time spent on those activities, lower than desired performance, and decreased care at work.
- **Mental health**: General mental health, including depression, anxiety, behavioral and emotional control and overall positive affect.

In addition, the SF-36 includes a transition item that asks about the change in general health status with respect to the previous year. This item is not used for the calculation of any of the scales, but provides important information about the perceived change in health status during the year prior to the application of the SF-36 instrument.

The number of items per dimension and the interpretation of the scores are presented in the table below:

Dimension	Number of items	Worst score (0)	Best score (100)
Physical function	10	Very limited to carry out all physical activities including bathing or showering, due to health	Performs all types of physical activities including the most strenuous without any limitation due to health
Physical role	4	Problems with work or other daily activities due to physical health	No problems with work or other daily activities due to health physics
Body pain	2	Very intense pain and extremely limiting	No pain or limitations due to him
General health	5	Evaluates its own health as bad and believes it is likely to get worse	Evaluates one's own health as excellent
Vitality	4	Feeling tired and exhausted all the time	It feels very dynamic and full of energy all the time. time
Social function	2	Extreme and very frequent interference with normal social activities, due to physical or emotional problems	Carries out normal social activities without any interference due to physical problems or emotional
Emotional role	3	Problems with work and other daily activities due to emotional problems	No problem with work and other daily activities due to problems emotional
Mental health	5	Feelings of anguish and depression all the time.	Feeling of happiness, tranquility and calmness all the time
Health transition item	1	He believes his health is much worse now than it was a year ago.	He believes that his overall health is much better now than it is now. one year ago
Note: Source: Vilagut et al.,2005.			

The table below shows schematically the physical and mental health phenomena that are evaluated by the SF-36:

Scale	Physical scale				Mental Scale			
	Functionality	Welfare	Disability	Personal evaluation	Functionality	Welfare	Disability	Personal evaluation
Physical function	X							
Physical role			X					
Body pain		X	X					
General health				X				X
Vitality		X				X		
Social function			X				X	
Emotional role							X	
Mental health					X	X		
Note: Translated from the SF-36 interpretation manual (Ware, 1993).								

The items and dimensions of the SF-36 are scored so that the higher the score the better the health status. The table presented below indicates how the scores for each area should be obtained.

Dimension	Item number	Text	Response option	Pre-coded value	Final value	Score
Physical function	3a	Intense exertion, such as running, lifting heavy objects, or participating in sports exhausting	• Yes, me very limited • Yes, it limits me a little • No, it does not limit me	• 1 • 2 • 3	• 1 • 2 • 3	Calculate the algebraic sum of the final value of items
	3b	Moderate exertion, such as moving a table, vacuuming, bowling, or walking more than one hour				
	3c	Picking up or carrying the shopping bag				
	3d	Climbing several flights of stairs				
	3e	Climbing a single flight of stairs				
	3f	Crouching or kneeling				
	3g	Walking one kilometer or more				
	3h	Walk several blocks (several hundred blocks) to meters)				
	3i	Walk a single block (about 100 meters)				
	3j	Bathing or dressing oneself				
Physical role	4a	Did you have to reduce the amount of time spent at work, or to your daily activities?	• Always • Almost always • Many times • Sometimes • Only sometime • Never	• 1 • 2 • 3 • 4 • 5 • 6	• 1 • 2 • 3 • 4 • 5 • 6	Calculate the algebraic sum of the final value of the items
	4b	Did you do less than you would have liked to do?				
	4c	Did you have to stop doing some tasks at your work or in your daily activities?				
	4d	Did you have difficulty doing your work or daily activities (e.g., did you find it difficult to do your job)? more than normal)?				

Dimension	Item number	Text	Response option	Pre-coded value	Final value	Score
Body pain	7	Did you have pain in any part of your body during the last 4 weeks?	• No, none • Yes, very little • Yes, a little • Yes, moderate • Yes, very much • Yes, very much	• 1 • 2 • 3 • 4 • 5 • 6	• 6 • 5 • 4 • 3 • 2 • 1	Calculate the algebraic sum of the final value of items
	8	During the last 4 weeks, to what extent has the pain made your usual work (including work outside the home and housework) difficult?	• Nothing • A little • Regular • Quite • Much	• 1 • 2 • 3 • 4 • 5	• 5 • 4 • 3 • 2 • 1	
General health	1	In general, you would say that your health is	• Excellent • Very good • Good • Regular • Mala	• 1 • 2 • 3 • 4 • 5	• 5 • 4 • 3 • 2 • 1	Calculate the algebraic sum of the final value of items
	11a	I think I get sick more easily than other people.	• Absolutely true • Quite true • I do not know • Quite false • Totally false	• 1 • 2 • 3 • 4 • 5	• 1 • 2 • 3 • 4 • 5	
	11b	I am as healthy as anyone else	• Absolutely true • Quite true • I do not know • Quite false • Totally false	• 1 • 2 • 3 • 4 • 5	• 5 • 4 • 3 • 2 • 1	

Dimension	Item number	Text	Response option	Pre-coded value	Final value	Score
	11c	I think my health is going to get worse	• Absolutely true • Quite true • I do not know • Quite false • Totally false	• 1 • 2 • 3 • 4 • 5	• 1 • 2 • 3 • 4 • 5	
	11d	My health is excellent	• Absolutely true • Quite true • I do not know • Quite false • Totally false	• 1 • 2 • 3 • 4 • 5	• 5 • 4 • 3 • 2 • 1	
Vitality	9a	Did you feel full of vitality?	• Always • Almost always • Many times • Sometimes • Only sometime • Never	• 1 • 2 • 3 • 4 • 5 • 6	• 6 • 5 • 4 • 3 • 2 • 1	Calculate the algebraic sum of the final value of items
	9e	Did you have a lot of energy?	• Always • Almost always • Many times • Sometimes • Only sometime • Never	• 1 • 2 • 3 • 4 • 5 • 6	• 6 • 5 • 4 • 3 • 2 • 1	
	9g	Did you feel exhausted?	• Always • Almost always • Many times • Sometimes • Only sometime • Never	• 1 • 2 • 3 • 4 • 5 • 6	• 1 • 2 • 3 • 4 • 5 • 6	

Dimension	Item number	Text	Response option	Pre-coded value	Final value	Score
	9i	Did you feel tired?	• Always • Almost always • Many times • Sometimes • Only sometime • Never	• 1 • 2 • 3 • 4 • 5 • 6	• 1 • 2 • 3 • 4 • 5 • 6	
Social function	6	During the past 4 weeks, to what extent have your physical health or emotional problems made your usual social activities with family, friends, neighbors or others difficult?	• Nothing • A little • Regular • Quite • Much	• 1 • 2 • 3 • 4 • 5	• 5 • 4 • 3 • 2 • 1	Calculate the algebraic sum of the final value of items
	10	During the past 4 weeks, how often has physical health or emotional problems made it difficult for you to engage in social activities (such as visiting friends or family)?	• Nothing • A little • Regular • Quite • Much	• 1 • 2 • 3 • 4 • 5	• 1 • 2 • 3 • 4 • 5	
Emotional role	5a	Did you have to reduce the time you spent at work or in your daily activities because of a problem? emotional?	• Always • Almost always • Many times • Sometimes • Only sometime • Never	• 1 • 2 • 3 • 4 • 5 • 6	• 1 • 2 • 3 • 4 • 5 • 6	Calculate the algebraic sum of the final value of items
	5b	Did you do less than you would have liked to do, by any emotional problems?				
	5c	Did you not go about your work or your daily activities as carefully as usual, by any emotional problems?				

Dimension	Item number	Text	Response option	Pre-coded value	Final value	Score
Mental health	9b	were you very nervous?	• Always • Almost always • Many times • Sometimes • Only sometime • Never	• 1 • 2 • 3 • 4 • 5 • 6	• 1 • 2 • 3 • 4 • 5 • 6	Calculate the algebraic sum of the final value of items
	9c	Did he feel so low in morale that nothing could cheer him up?	• Always • Almost always • Many times • Sometimes • Only sometime • Never	• 1 • 2 • 3 • 4 • 5 • 6	• 1 • 2 • 3 • 4 • 5 • 6	
	9d	Did you feel calm and collected?	• Always • Almost always • Many times • Sometimes • Only sometime • Never	• 1 • 2 • 3 • 4 • 5 • 6	• 6 • 5 • 4 • 3 • 2 • 1	
	9f	Did you feel discouraged and sad?	• Always • Almost always • Many times • Sometimes • Only sometime • Never	• 1 • 2 • 3 • 4 • 5 • 6	• 1 • 2 • 3 • 4 • 5 • 6	

Dimension	Item number	Text	Response option	Pre-coded value	Final value	Score
	9h	Did you feel happy?	• Always • Almost always • Many times • Sometimes • Only sometime • Never	• 1 • 2 • 3 • 4 • 5 • 6	• 6 • 5 • 4 • 3 • 2 • 1	
Declared health evolution	2	How would you say your health is today compared to a year ago?	• Much better now than a year ago • Slightly better now than a year ago • About the same as a year ago • Slightly worse now than a year ago • Much worse now than a year ago	• 1 • 2 • 3 • 4 • 5	• 5 • 4 • 3 • 2 • 1	The score for this item is inverted as a final value.

The calculation of the total scores for each SF-36 scale is shown in the table below:

Scale	Final sum after recoding	Minimum and maximum (direct scoring)	Possible range in direct scores
Physical function	3a+3b+3c+3d+3e+3f+3g+3h+3i+3j	10, 30	20
Physical role	4a+4b+4c+4d	4, 20	16
Body pain	7+8	2, 12	10
General health	1+11a+11b+11c+11d	5, 25	20
Vitality	9a+9e+9g+9i	4, 24	20
Social function	6+10	2, 10	8
Emotional role	5a+5b+5c	3, 15	12
Mental health	9b+9c+9d+9f+9h	5, 30	25

After obtaining the direct scores, they should be transformed to a scale from 0 to 100 using the following formula (Alonso et al., 2003):

$$Puntuación\ transformada = \frac{Puntuación\ directa - Puntuación\ mínima}{Rango\ de\ la\ puntuación\ directa} \cdot 100$$

For example, if a patient scores 21 points on the Physical Function scale, his or her score out of 100 would be as follows:

$$Puntuación\ transformada = \frac{21 - 10}{20} \cdot 100 = 55$$

Transformed scale scores are not calculated for the Declared Evolution of Health item. It is recommended to treat responses to this item as ordinal level data and to analyze the percentage of respondents selecting each response option.

1.2.4. Subgroup analysis.

All efficacy variables will be studied in the following subgroups:

- Patients with primary episode of CDI.
- Patients with recurrent episodes of CDI.
- Patients not receiving vancomycin prior to receiving study treatment.

1.3. Analysis populations.

The clinical trial will be conducted in patients presenting with an episode of CD infection (either the first episode or subsequent recurrences).

In the present study 104 patients will be enrolled (52 in the control group treated with Fidaxomicin and 52 patients in the experimental group treated with the investigational product).

Note: The analysis populations have been defined in the protocol (page 92).

1.3.1 Intention-to-treat (ITT) population.

The Intention-to-treat population consists of all patients who have been included in the study, whether or not they meet the selection criteria and have been randomized.

The efficacy analysis will be carried out with the **Intention to Treat (ITT) population**.

1.3.2 Population per Protocol (PP).

The Protocol Population includes all those patients who have been recruited for the study, met all the inclusion criteria, none of the exclusion criteria and have continued in the study without any major deviation from the protocol.

1.3.3 Security Population (SP).

The Safety Population consists of all those patients who have been randomized into either of the two randomization groups and who have taken at least one dose of the corresponding treatment.

The **security analysis** will be carried out with the **Security Population (SP).**

1.4. Study outline.

Visit	Baseline (randomization and treatment)	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6 (end of study)
Weather	(-1 a 0)	72h after the start of treatment	3 weeks	8 weeks	3 months	6 months
Type of visit	On-site	On-site	On-site	On-site	Telefónica	Telefónica
Window period	N/A	±1 day	±4 days	±4 days	±6 days	±6 days
Procedures						
Informed consent	X					
Washing period[1]	X					
Inclusion/exclusion criteria	X					
Randomization	X					
Treatment	X					
Medical history and history of ICD	X					
Physical examination	X	X	X	X		
Pregnancy test	X					
Laboratory tests	X			X		
Stool sample	X2			X1		
Concomitant treatment of patients	X	X	X	X	X	X
Evaluation of therapeutic compliance		X	X			
Assessment of diarrhea (main variable)	X	X	X	X	X	X
Evaluation of length of hospitalization	X	X	X	X		
Evaluation of good / bad evolution of the patient	X	X				
Evaluation of adverse effects	X	X	X	X	X	X
SF36	X			X		X

[1] in case the patients have been treated with vancomycin with a 3 to 5 day regimen for the recurrence to be treated in the study and/or antibiotic for other medical indication, it is necessary that they undergo a 24h washout period.
[2]A first sample is collected (up to 72 hours before inclusion of the patient in the study) to confirm the diagnosis of CD and a second sample is collected once the patient has been included in the study and before receiving the medication to be stored for subsequent studies;[3] : A stool sample will be collected for subsequent studies.

2 Description of the statistical analysis.

2.1 Disposition and flow of patients.

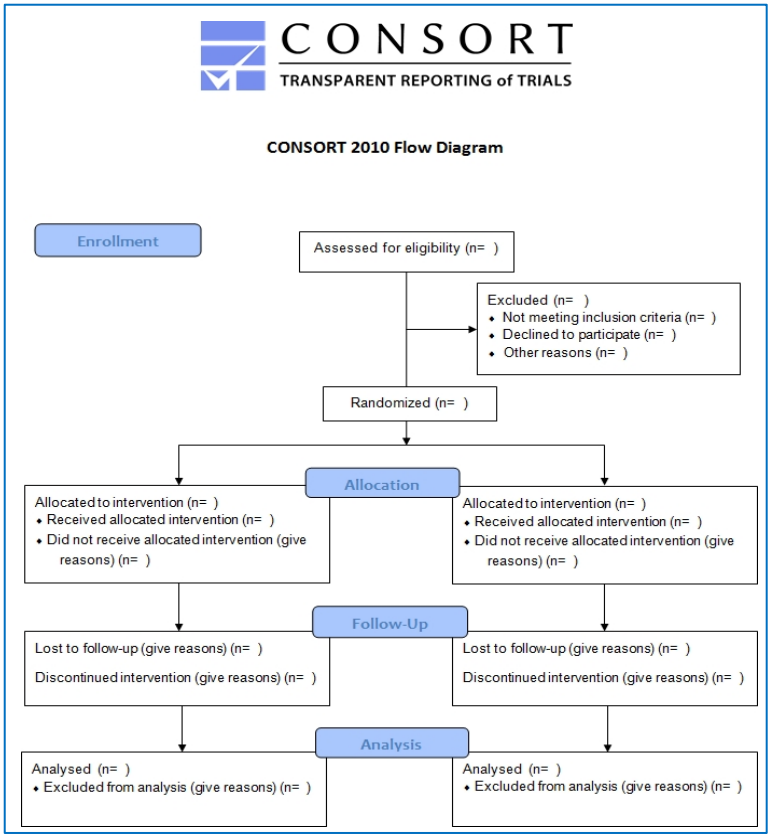
Regarding patient disposition, descriptive data on the following items will be included:

- Number of patients selected.
- Number of selection failures.
- Number and percentage of patients in each of the analysis populations.
- Number of patients excluded from each analysis population along with the reason for exclusion.
- Completion of the study:
    - Number and percentage of patients who completed treatment.
    - Number and percentage of patients who did not complete treatment.
    - Number of patients who completed the study.
    - Number of patients who did not complete the study.
    - Listing of all withdrawals along with the reason for termination.

A patient flow diagram following the structure of the CONSORT report including the following information will be made and presented in the statistical report:

- Selected patients.
- Selection errors. Reasons.
- Patients randomized to the different treatment groups.
- Patients withdrawn. Reasons.
- Patients analyzed.

An example of a patient flow format used by the CONSORT system is given below:



2.2 Statistical analysis according to protocol.

The protocol (pages 89-97) provides a general description of the planned analyses. In the present document, these analyses are detailed in their corresponding sections.

2.3 Descriptive statistics.

2.3.1 Baseline descriptive statistics.

The protocol (page 91) specifies: "*A descriptive analysis of all relevant variables collected at baseline will be performed, and this analysis will be divided by randomization groups.*

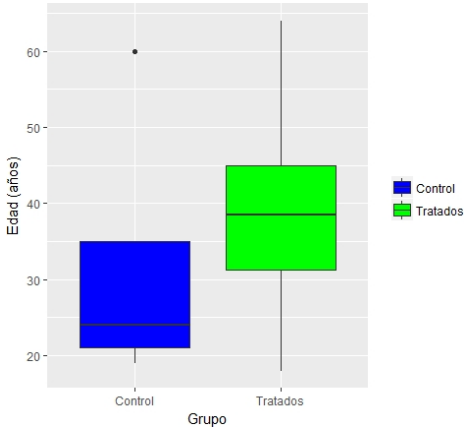
The treatment of both types of variables is illustrated as an example:

- **Categorical variables** (nominal, dichotomous or polytomous): Distribution of absolute and relative frequencies (transformed into percentages). Cumulative relative frequency distribution in the case of variables with ordered categories.

Frequency distribution of the sex variable (simulated data).					
Group	Man		Woman		p-value
	N	%	N	%	
Treaties	2	50.0	2	50.0	1
Controls	5	57.1	4	42.9	
Total	7	53.8	6	46.2	

- **Quantitative variables**: mean, standard deviation, minimum and maximum. If appropriate, other descriptive statistics such as skewness and kurtosis indices should be added, 95% confidence intervals should be added, and an analysis of the normality of the distribution should be included. In those variables that do not meet the assumption of normality, it will be convenient to use the median instead of the mean, and the interquartile range (or the AVERAGE) instead of the standard deviation. On occasions where the distribution of scores is peculiar, a histogram should be added for better understanding. When it is reasonable to compare groups of subjects (e.g., controls and treated), the graphical representation will be made by box-and-whisker plots. Where possible, a summary table in which each variable is on one line should be presented to facilitate tracking of that variable throughout the trial.

Age (in years) of participants (simulated data).								
Group	Media	Median	DT	MAD	As	K	Min.	Max.
Treaties	39.10	21.00	12.97	11.12	0.24	-0.78	18.00	64.00
Control	31.80	38.00	16.93	7.41	0.77	-1.35	19.00	60.00



Note. SD: Standard Deviation. MAD: *Mean Absolute Deviation*. As: Asymmetry. K: Kurtosis. Min:
Min. Max: Maximum.

2.3.2 Efficiency analysis.

The efficacy analysis involves the application of a descriptive analysis at the beginning, which will be complemented with the inferential analysis (if applicable). In addition, the possible *missing values* that exist in each variable being analyzed for the determination of efficacy will be accounted for.

For descriptive analysis, data will be presented as absolute frequencies and percentages in the case of qualitative variables. For quantitative variables, central tendency statistics, such as mean and median, and dispersion statistics, such as standard deviation and interquartile range, minimum and maximum values, will be used, and indicators of the shape of the distribution such as skewness and kurtosis indexes can be provided. In the case of ordinal variables, depending on the number of categories, one or another form of description will be used. In all cases, a column should be added to the data presentation table including the number of patients with available data. The analyses will be divided by randomization group.

2.3.3 Security analysis.

The protocol (pages 94-95) states that "*Safety variables are operationalized with the number of documented adverse events ranked by frequency, severity, seriousness and causality as assessed by subject interview and physical examination.*

For the analysis of the Adverse Events, a descriptive analysis (frequencies and percentages) of these events throughout the study will be performed, presenting a list of them, grouped according to severity and intensity. These events will be listed by patient and randomization group. Adverse reactions will be recorded using the Verbatim described in the eCRD.

Adverse Events and concomitant medication will be coded using a standardization system. The most current version of the MedDRA system will be used for Adverse Events and the ATC code will be used for concomitant medication.

The analysis of the physical examination variables will be done by frequencies and percentages and divided by randomization group.

*The **safety analysis** will be conducted on the **Safety Population** (SP).*"

Safety analyses will be carried out incorporating these lists:

- Adverse events (AEs).
- Serious adverse events (SAEs).
- Adverse Events of Special Interest (AESI).

- Events related to the investigational drug or comparator: Those whose causal relationship is "probable" or "related" will be considered.
- Events leading to discontinuation of study therapy.
- Events with grades of "serious", "life-threatening" or "death".

For all such events, this information will be provided:

- Number of events: this information will also be displayed by
  - SOC and PT codes.
  - Grade.
  - Progression.
  - Action taken.
  - Relationship with MBK-01 or with Fidaxomicin.
- Type of events: this information will be displayed by SOC and PT codes.
- Number of patients within each event category: this information will also be displayed by:
  - Organ classification system (OCS) and preferred term (PT).
  - Grade.
  - Progression.
  - Action taken.
  - Relationship with MBK-01 or with

Fidaxomicin. The variables related to patient exitus are:

  - Patients who died.
  - Cause of exitus.
  - Days since last dose of MBK-01 or Fidaxomicin. Variables

related to concomitant medication:

  - Number of concomitant medications.
  - Number of patients who required concomitant medication.
  - Number of concomitant medications per patient.
  - Number of concomitant medications related to Adverse Events for patients overall and per patient.

In addition, details on concomitant medications per patient will be displayed:

  - ATC code and name.
  - Dosage.
  - Route of administration.
  - Duration of treatment: difference in days between the end date and the start date.
  - Reason for treatment.
  - Relationship with adverse event.

2.4 Inferential analysis.

2.4.1 Efficiency analysis.

The protocol (page 92) specifies that the efficacy analysis will be conducted in the Intention-to-Treat Population.

Page 72 of the protocol describes that The probability of recurrence as a function of randomization group will be studied by binary logistic regression analysis. The odds ratio of recurrence will be reported, using the MBK-01 group as a reference. In addition, the confidence interval, Nagelkerke's R-squared and Yule's Q coefficient will be included, as well as the critical level. The reference group to be used will be MBK-01 (expecting that in Fidaxomicin the probability will be equal or higher compared to MBK-01).

The number of diarrhea episodes per visit and group will be analyzed, if assessable data are available, using analysis of variance mixed models or relevant robust/nonparametric models.

The good/bad evolution of patients will be analyzed using logistic regression procedures incorporating the same indicators as described for the analysis of the probability of recurrence.

Overall survival in weeks will be analyzed by Cox regression including the randomization group as an independent variable. Overall survival will be measured in weeks.

Due to the characteristics of the trial, it is possible to state the hypotheses as either non-inferiority or superiority. Given that the trial is a Phase III trial, both hypotheses can be tested in the same clinical trial as long as the α risk is controlled, or, in other words, to state that there are significant differences when in reality, at the population level, there are no such differences. It is important to remember that, in order to avoid false positives, the type I or α error must be distributed among the different hypotheses, which means that, in order to find statistically significant results, the critical level or p value must be lower (Ferreira-Gonzalez, 2014).

Although the overall significance level will be alpha=0.05, because more than one analysis of the data will be performed, it will be adjusted to 0.0294 to avoid inflation of the probability of type I error (Geller and Pocock, 1987). If multiple pairwise comparisons analyses are employed, the statistical report will describe the procedure used. Measures of the effect size of the results will be incorporated.

2.4.2 Security analysis.

The statistical significance between the randomization group and the intensity and treatment-relatedness of both Adverse Events and Serious Adverse Events will be studied, whenever feasible, using the χ^2 test for association between categorical variables. If possible, the *odds ratio* of occurrence of Serious Adverse Events will be calculated and the 95% confidence interval of this *odds ratio will* be provided. The Fidaxomicin group will be used as the reference group.

2.5 Statistical analysis steps.

The following tables describe the techniques to be applied according to the objectives and variables affected.

In the case of **qualitative variables**, a series of analyses based on the research objectives is established for the statistical analysis.

Objectives	Variables affected	Applied technique	Expected results
Determination of adverse events and serious adverse events.	▪ Grouping variables: - Randomization group. ▪ Variables to be analyzed: – Presence of AAG and AA. – Gravity. – Relationship to treatment.	Descriptive analysis: frequencies and percentages. Inferential analysis: Chi-square test. Odds ratio analysis.	The incidence will be minimal.
Determination of Adverse Events of Special Interest (AESI)	▪ Grouping variables: - Randomization group. ▪ Variables to be analyzed: – Presence of AESI. – Gravity. – Relationship to treatment.	Descriptive analysis: frequencies and percentages. Inferential analysis: Chi-square test. Odds ratio analysis.	The incidence will be minimal.
Determination of adverse events in relation to the ICD	▪ Grouping variables: - Randomization group. ▪ Variables to be analyzed: – Presence of adverse events related to ICD. – Gravity. – Relationship to treatment.	Descriptive analysis: frequencies and percentages. Inferential analysis: Chi-square test. Odds ratio analysis.	The incidence will be minimal.
Determination of mortality associated with CDI.	▪ Grouping variables: - Randomization group. ▪ Variables to be analyzed:	Descriptive analysis: frequencies and percentages. Inferential analysis: Chi-	Mortality will be minimal.

		square test.	

Objectives	Variables affected	Applied technique	Expected results
	- Determination of mortality associated with CDI.	Odds ratio analysis.	
ICU admissions	▪ Grouping variables: - Randomization group. ▪ Variables to be analyzed: - Determination of ICU admissions	Descriptive analysis: frequencies and percentages. Inferential analysis: Chi-square test. Odds ratio analysis.	Exploratory objective
Absence of CD diarrhea with no need for retreatment	▪ Grouping variables: – Randomization group. – Visit number. ▪ Variables to be analyzed: - Resolution of the diarrhea episode.	Descriptive analysis: frequencies and percentages. Inferential analysis: Chi-square test. Odds ratio analysis.	The incidence of diarrhea will be minimal.
Determination of good/bad patient outcome	▪ Grouping variables: – Randomization group. – Visit number. ▪ Variables to be analyzed: - Good/bad patient evolution	Descriptive analysis: frequencies and percentages. Inferential analysis: Chi-square test. Odds ratio analysis.	The evolution of the patients will be good.

For inferential analysis involving **quantitative variables**, a series of analyses are established based on the research objectives.

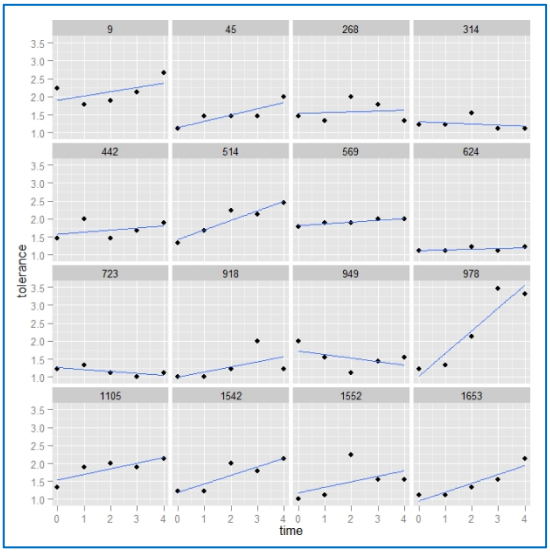
Objectives	Variables affected	Applied technique	Expected results
Duration of hospitalization	▪ Grouping variables: -Randomization group ▪ Variable under analysis: -Duration of hospitalization (time between the beginning of hospitalization and the date of discharge)	t-test for independent samples In case of non-compliance with assumptions: Mann-Whitney U test	The duration will be shorter for patients in the experimental group.
Duration of treatment	▪ Grouping variables: -Randomization group ▪ Variable under analysis: - Duration of treatment (time between the start of treatment and the end date of treatment)	t-test for independent samples In case of non-compliance with assumptions: Mann-Whitney U test	The duration will be shorter for patients in the experimental group.
Time to recurrence	▪ Grouping variables: -Randomization group. ▪ Variable under analysis: – Start date of treatment. – Recurrence. – Recurrence date. ▪ End date of study or date of patient withdrawal in case the patient has not had recurrence.	Survival analysis.	Exploratory objective

Overall survival (OS)	▪ Grouping variables: -Randomization group. ▪ Variable under analysis:	Survival analysis.	Exploratory objective

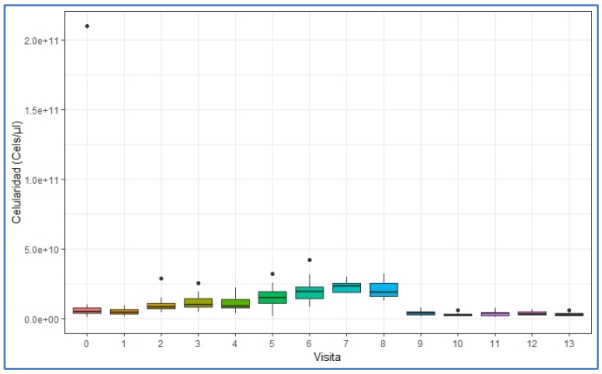
Objectives	Variables affected	Applied technique	Expected results
	– Start date of treatment. – Exitus. – Date of exitus. – End date of study or date of patient withdrawal in case the patient is not an exitus.		
Changes in quality of life variables	▪ Grouping variables: – Randomization group – Visit number ▪ Variable under analysis: ▪ SF-36 score and its dimensions	Mixed ANOVA. In case of non-compliance with assumptions, Brunner and Langer's test	Exploratory objective
Number of adverse events	▪ Grouping variables: -Randomization group ▪ Variable under analysis: -Number of adverse events	t-test for independent samples In case of non-compliance with assumptions: Mann-Whitney U test	The number of adverse events will be lower in the experimental group.
Number of serious adverse events	▪ Grouping variables: -Randomization group ▪ Variable under analysis: -Number of serious adverse events	t-test for independent samples In case of non-compliance with assumptions: U-test of Mann-Whitney	The number of serious adverse events will be lower in the experimental group.

Where possible, illustrative figures, such as those given below as examples, should be incorporated:

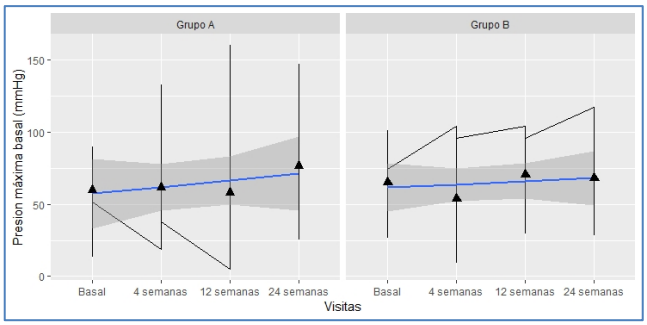
- Figure representing the evolution of each individual patient (simulated data) as a function of visits:



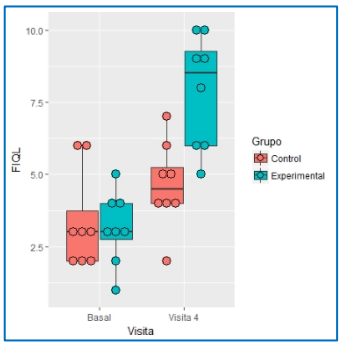
- Figure showing the overall results in terms of visits (simulated data):



- Figure showing the growth trend by group (simulated data):



- Figure showing the passage of time for each randomization group (simulated data):



2.6 Procedure for accounting for missing data, treatment of outliers and case imputation.

The number of missing data that cannot be retrieved or obtained (*missing values*) will be reported in the presentation of results. From the moment patients have completed their participation in the study, no additional data will be collected (except as required by the *Query plan* and for the resolution of possible new *queries arising from* the initial analysis of the database and study variables). The data collected will be used for efficacy and safety analysis.

Note: The protocol does not establish any procedure for case allocation.

The statistical report will incorporate information on the count of missing values for each variable studied and according to the study group. Optionally, the possible origin of the missing values will be studied to determine if they are completely random losses (MCAR), random losses (MAR) or non-random losses (MNAR).

In case of detecting the need to apply imputation methods in this study, both the method and its reference shall be specified and a study shall be made of the number of missing values and assess whether it is feasible to perform the imputation or not, and the decision taken shall be justified at all times. As far as possible, an indicator of the possible presence of bias in the imputation shall be provided by means of a sensitivity analysis. Whichever imputation method is selected, it shall be correctly identified and referenced.

On the other hand, statistical *outliers* will be detected from a graphical perspective, by means of box and whisker plots, and more complex mathematical procedures may be performed if necessary. The presence of these values may require the application of robust inferential and/or non-parametric tests that should be correctly specified and referenced in the final statistical report.

2.7 Procedures for multiplicity analysis.

The protocol does not specify a procedure for correcting the significance level associated with multiple comparison analyses. The more extrapolations that are made in simultaneous inferences, the more likely it is that erroneous inferences will occur. Several statistical techniques have been developed to prevent this from happening, allowing direct comparison of significance levels for single and multiple comparisons. These techniques generally require a tighter significance threshold for single comparisons to compensate for the number of inferences that are made. Methods that could be considered for multiple comparisons are, for example, Bonferroni correction or Holm's adjustment. For a detailed description of multiple comparison methods, see Chen, Feng, and Yi (2017).

Due to the characteristics of the trial, it is possible to state the hypotheses as either non-inferiority or superiority. Given that the trial is a Phase III trial, both hypotheses can be tested in the same clinical trial as long as the α risk is controlled, or, in other words, to state that there are significant differences when in reality, at the population level, there are no such differences. It is important to remember that, to avoid false positives, it is required to "distribute" the type I or α error between the different hypotheses, which means that, to find statistically significant results, the value of the critical level or p must be lower (Ferreira-Gonzalez, 2014).

In the case of using any procedure for multiple comparisons, it must be correctly specified and referenced in the final statistical report.

2.8 Intermediate statistical analysis.

Pages 77 and 78 of the protocol specify that two intermediate analyses are planned for this study.

1. The first interim analysis will be performed when approximately 50% of the patients have completed treatment.
2. The second interim analysis will be performed when approximately 80% of the patients have performed V4.

All interim analyses will incorporate a descriptive analysis of baseline patient characteristics, an efficacy analysis and a safety analysis. The analysis of baseline variables and the efficacy and safety analyses will be performed by comparing the randomization groups.

Both intermediate analyses will be done in the ITT population in the efficacy section and in the PS in the safety analysis section.

The results could lead to early discontinuation of the study if large differences between the two randomization groups are revealed after the study, or obvious futility is shown, so that there is no possibility of showing that continuing the study until 100% of enrolled patients are reached would result in a clinically meaningful effect. In addition to saving time and resources, this could reduce the exposure of study participants to a treatment that is inferior or shows no difference from the comparator.

If large and clear differences between the randomization groups or clear futility are observed, early discontinuation of the study should be considered. The study of differences between randomization groups should be accompanied by measures of effect size, confidence intervals and not only statistical significance.

Obviously, given the variety of statistical analyses to be carried out, the final decision to establish futility or obvious efficacy must be agreed upon after reviewing the results obtained from the intermediate analyses.

2.8.1 First interim analysis.

This analysis will include these variables:

- Efficiency analysis:
  - Recurrence probability.
  - Time to recurrence.

- Good/bad evolution of the patient.
- Analysis by subgroups.

- Security analysis:
  - Number of Adverse Events per randomization group.
  - Number of patients within each Adverse Event per randomization group.
  - Type of Adverse Events by randomization group.
  - Number of Adverse Events of Special Interest by randomization group.
  - Number of patients within each Adverse Event of Special Concern per randomization group.
  - Type of Adverse Events of Special Interest by randomization group.
  - Number of Serious Adverse Events by randomization group.
  - Type of Serious Adverse Events by randomization group.
  - Adverse events related to treatment.
  - Severity of Adverse Events.
  - Adverse Events in relation to ICD.
  - Mortality associated with CDI.
  - Admissions to the Intensive Care Unit (ICU).

This analysis will be versioned as version 1.0. If justified modifications are required, it will be versioned as 1.1, 1.2 and so on.

2.8.2 Second interim analysis.

This analysis will include these variables:

- Efficiency analysis:
  - Recurrence probability.
  - Time to recurrence.
  - Good/bad evolution of the patient.
  - Overall survival.
  - Analysis by subgroups.
- Security analysis:
  - Number of Adverse Events per randomization group.
  - Number of patients within each Adverse Event per randomization group.
  - Type of Adverse Events by randomization group.
  - Number of Adverse Events of Special Interest by randomization group.
  - Number of patients within each Adverse Event of Special Concern per randomization group.

- Type of Adverse Events of Special Interest by randomization group.
- Number of Serious Adverse Events by randomization group.
- Type of Serious Adverse Events by randomization group.
- Adverse events related to treatment.
- Severity of Adverse Events.
- Adverse Events in relation to ICD.
- Mortality associated with CDI.
- Admissions to the Intensive Care Unit (ICU).
- Analysis of concomitant medication:
  - Number of drugs per randomization group.
  - Number of medications per patient.
  - Type of drugs per randomization group and per patient.

- Analysis of quality of life: Analysis of the scores of the areas evaluated by the SF-36 questionnaire.

This analysis will be versioned as version 2.0. If justified modifications are required, it will be versioned as 2.1, 2.2 and so on.

2.9 Subgroup analysis.

The analysis of the efficacy variables will be studied in the following subgroups:

- Patients with primary episode of CDI.
- Patients with recurrent episodes of CDI.
- Patients not receiving prior vancomycin.

The analysis of the probability of recurrence and good/bad evolution of the patients will be performed by binary logistic regression analysis. The odds ratio of recurrence will be reported, using the MBK-01 group as reference. In addition, the confidence interval, Nagelkerke's R-squared and Yule's Q coefficient will be included, as well as the critical level.

Time to recurrence will be studied using Cox regression procedures.

2.10 Deviations from the SAP.

Any deviation from the original SAP must be duly described and justified in the statistical report.

3. Summary of statistical analysis.

The main objective of the study is to evaluate the efficacy of the investigational product (MBK-01) compared to the control (Fidaxomicin) at 8 weeks after initiation of treatment.

Secondary objectives involve assessing the safety of the investigational product and the quality of life of the patients participating in the study.

Summary of statistical analysis.					
Overall objectives	**Specific objectives**	**Independent variable**	**Operationalization and level of measurement of the independent variable in the database**	**Dependent variable**	**Operationalization and level of measurement of the dependent variable in the database**
Security	Determination of adverse events, their degree, and relationship with the product under investigation.	Randomization group	RANDGRP. Qualitative	AA.	*An active search for potential AESEV AAs will be conducted. AEREL
	Determination of serious adverse events, their severity, and relationship to the product in research.	Randomization group	RANDGRP. Qualitative	Presence of AAG.	AESER. AESEV AEREL
	Determination of adverse events in relation to the ICD.	Randomization group	RANDGRP. Qualitative	Presence of AA in relation to ICD.	*An active search will be made for possible AAs related to the ICD.
	Determination of adverse events of special interest, their degree, and relationship with the product in research.	Randomization group	RANDGRP. Qualitative	Presence of AESI.	*An active search for potential AESIs will be conducted.
	ICU admissions	Randomization group	RANDGRP. Qualitative	ICU admissions	*An active search will be made for those AEs/SAEs involving admission to the ICU.

Summary of statistical analysis.					
Overall objectives	**Specific objectives**	**Independent variable**	**Operationalization and level of measurement of the independent variable in the database**	**Dependent variable**	**Operationalization and level of measurement of the dependent variable in the database**
Efficiency	Resolution of the diarrhea episode	Group of randomization	RANDGRP. Qualitative	Resolution of the diarrhea episode	RESOL_FAYN
	Good/Bad evolution of patient	Randomization group	RANDGRP. Qualitative	Good/Bad evolution of patient	EVAL_RS
	Time to recurrence	Randomization group	RANDGRP. Qualitative	Presence of recurrence Date of recurrence recurrence	RECU_RSYN RECU_RSDAT
	Overall survival	Randomization group	RANDGRP. Qualitative	Exitus End date of study Exit date	*For calculating the time variable TRTDEC TRTDECDAT
	Quality of life variables	Randomization group	RANDGRP. Qualitative	Dimensions of quality of life	*For calculating each dimension based on the answers. to the variables

4. References.

Brunner E, Domhof S, Langer F (2002). Nonparametric Analysis of longitudinal Data in Factorial Experiments. John Wiley & Sons, New York.

Chen, S.-Y., Feng, Z., & Yi, X. (2017). A general introduction to adjustment for multiple comparisons. *Journal of Thoracic Disease, 9*(6), 1725-1729. doi:10.21037/jtd.2017.05.34.

Del Re, A. C. (2013). *compute.es: Compute Effect Sizes*. R package version 0.2-2. URL http://cran.r-project.org/web/packages/compute.es.

Grothendieck, G. (2017*). sqldf: Manipulate R Data Frames Using SQL*. R package version 0.4-11. https://CRAN.R-project.org/package=sqldf.

Kassambara, K. (2020). ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.4.0. https://CRAN.R-project.org/package=ggpubr.

Kassambara, A., Kosinski, M. & Biecek, P. (2021). *survminer: Drawing Survival Curves using 'ggplot2'*. R package version 0.4.9. URL https://CRAN.R-project.org/package=survminer.

Navarro, D. J. (2015) *Learning statistics with R: A tutorial for psychology students and other beginners*. (Version 0.5) University of Adelaide. Adelaide, Australia.

Noguchi, K., Gel Y. R., Brunner, E., and Konietschke, F. (2012). nparLD: An R Software Package for the Nonparametric Analysis of Longitudinal Data in Factorial Experiments. Journal of Statistical Software, 50(12), 1-23. URL http://www.jstatsoft.org/v50/i12/.

R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Therneau T (2021). *A Package for Survival Analysis in R*. R package version 3.2-13, URL https://CRAN.R-project.org/package=survival.

Terry M. Therneau, Patricia M. Grambsch (2000). *Modeling Survival Data: Extending the Cox Model*. Springer, New York. ISBN 0-387-98784-3.

Warnes, G. R., Bolker, B., Lumley, T., & Johnson, R. C. (2018). *gmodels: Various R Programming Tools for Model Fitting*. R package version 2.18.1. https://CRAN.R-project.org/package=gmodels.

Vilagut, G., M. Ferrer, L. Rajmil, P. Rebollo, G. Permanyer-Miralda, J. M. Quintana, R. Santed, J. M. Valderas, A. Ribera, A. Domingo-Salvany, and J. Alonso. "[the Spanish Version of the

Short Form 36 Health Survey: A Decade of Experience and New Developments]." Gac Sanit 19, no. 2 (Mar-Apr 2005): 135-50.

Vilagut, G., J. M. Valderas, M. Ferrer, O. Garin, E. Lopez-Garcia, and J. Alonso. "[Interpretation of Sf-36 and Sf-12 Questionnaires in Spain: Physical and Mental Components]." Med Clin (Barc) 130, no. 19 (May 24, 2008): 726-35.

Ware, J.E., K.K. Snow, M. Kosinski, and B. Gandek. "SF-36 Health Survey Manual & Interpretation Guide" MA: The Health Institute, New England Medical Center, 1993 (pp.6:1- 6:22).

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wickham; H., & Bryan, J. (2019). readxl: Read Excel Files. R package version 1.3.1. https://CRAN.R-project.org/package=readxl.