Wemedoo
Clinical Information Specialists

AORA Health

Ezymax Forte

Study Design and Sample Size Estimation

Table of Contents

1. Introduction and General Considerations

Pancreatic Exocrine Insufficiency (PERT) can be caused by numerous factors all which are treated with Pancreatic Enzyme Replacement Therapy (PERT).

A literature review was performed with the goal of proposing potential study designs.

The following indications, with PEI as an etiology, can be the focus of this clinical trial:

- **Functional dyspepsia** – frequent condition, various symptoms, usually patients with good general health, no risky patients, younger participants, dominantly women
- **Irritable bowel syndrome diarrhea (IBS-D)** – frequent, usually middle aged or younger participants
- **Chronic pancreatitis** – condition after chronic pancreatitis may lead to exocrine insufficiency and cause dyspepsia
- **Cystic fibrosis** – NOT CONSIDERED for this study

The most feasible studies to evaluate Enzymax would be functional dyspepsia or PEI after chronic pancreatitis.

2. Functional Dyspepsia

Studies can be designed to:

1. Estimate some characteristics in one group of participants with a specific condition (for example, estimate the quality of life of patients with exocrine pancreatic insufficiency);
2. Examine the superiority of one product to another product (for example, prove that the examined product is better than placebo);
3. Examine the non-inferiority, i.e. show that the test product is “not much worse” than reference product (for example, Enzymax has very similar effect to Creon);
4. Establish the equivalence, i.e. prove that the test product is very similar to the reference product, not worse, not better.

In most cases, we examine either superiority (test vs. placebo) or noninferiority (test vs. reference product).

2.1. Superiority Approach

Functional dyspepsia is the most frequent condition, and a clinical study is easy to conduct. The reason is that in this case we compare effect of drug vs. placebo regarding the symptoms in specified period, and no specific laboratory analyses are necessary.

Most of the studies regarding functional dyspepsia are double blind, randomized, parallel design with Patient Reported Outcome (PRO), contrary to chronic pancreatitis studies with stool tests as the primary outcome (Coefficient of Fat Absorption, CFA).

Two studies obtained by Ana are Ullah et al, and Majeed et al.

Ullah et al. – placebo controlled, parallel design

Using two-sided alpha=0.05, mean change of NDI SF1 5 and -5 (common SD=19), 58 participants per group is sufficient to achieve 80% study power to detect significant difference between study groups. If we take into consideration 20% dropouts, then 146 participants (73 per arm) are sufficient to preserve study power in case of maximal allowable dropout.

	Placebo group ($n = 60$)			Treatment group ($n = 60$)		
	t0	t1	t2	t0	t1	t2
NDI-SF 1	42.1 ± 21.1 11 – 80	45.0 ± 21.4 13 – 81	47.9 ± 22.4 13 – 85	46.2 ± 20.4 13 – 80	43.6 ± 18.7 13 – 77	41.3 ± 17.1 13 – 73
NDI-SF 2–5	220.5 ± 28.9 164.6 – 281.9	207.7 ± 27.7 149.9 – 271.6	200.7 ± 32.5 143.9 – 292.6	223.1 ± 25.6 166.5 – 279.8	240.9 ± 30 160.9 – 301.1	274.6 ± 37.8 142.7 – 322.6
NDI-SF 6	28.3 ± 10 11 – 52	30.5 ± 11.9 12 – 80	31.4 ± 11.1 12 – 53	25.9 ± 8.2 13 – 40	25.8 ± 8 12 – 41	25.5 ± 8.2 12 – 44
VAS	5.7 ± 2.3 1 – 10	5.8 ± 2.3 1 – 9	6.2 ± 2.2 2 – 10	6.0 ± 2.6 1 – 9	5.0 ± 2.4 1 – 9	4.3 ± 2.4 1 – 9
PSQI	14.2 ± 4.4 7 – 21	14.3 ± 4.2 7 – 21	14.8 ± 4.4 6 – 21	14.4 ± 4.1 7 – 21	12.1 ± 4.6 2 – 20	10.7 ± 4.4 3 – 22

The table below presents preliminary sample size (SS) calculations with Test NDI SF1 score change of -5 and -4 (reduction of 5 and 4 units), and placebo NDI SF1 score change from -1 to 5 units.

	Change in NDI SF1 score		
	Test	Placebo	SS per arm
Common SD=19	-5	-1	355
	-5	0	228
	-5	1	159
	-5	2	117
	-5	3	90
	-5	4	71
	-5	5	58
	-4	-1	630
	-4	0	355
	-4	1	228
	-4	2	159
	-4	3	117
	-4	4	90

Majeed et al. – placebo controlled parallel design

Majeed et al also tested functional dyspepsia in a parallel, randomized, placebo-controlled trial (table below).

	MEC		*Placebo*	
Parameters	*Baseline*	*Final visit*	*Baseline*	*Final visit*
SF-LDQ	24.5±3.15	17.9±4.87	24.9±3.92	21.5±5.95
NDI-SF	26.7±5.73	15.7±6.79	28.5±4.73	22.7±8.28
CGI-S	17.8±1.12	8.3±5.73	16.3±4.06	13.8±5.91
VAS	7.3±1.22	2.6±2.48	7.6±1.10	5.6±3.12
GDSS	6.8±1.28	3.8±1.91	7.3±1.74	5.8±2.78

In the table below, we present average changes in scores from baseline with common SD, as well as the sample size per arm:

	Change in score from baseline			
	MEC	PLAC	Common SD	SS per Arm
SF-LDQ	6	3	5	45
NDI-SF	11	6	7	32
CFI-S	9	3	4	9
VAS	5	2	2.5	12
GDSS	3	2	2	64

The most adequate measurement for the indication would be NDI, but this calculation of adequate certainty, as it may have to small variability and with a larger effect size than was observed in other studies. As such, the data from that reason Ullah et al, study is more appropriate to use for the sample size calculation.

2.2. Non-Inferiority Studies

Non-inferiority studies must show that the score of symptoms or QoL are not significantly lower than the reference product. The not significantly lower threshold is called non-inferiority margin (NIM).

The problem is how to define the NIM since it is necessary to properly explain the NIM and to precisely define it. Defining the NIM is not simple requires detailed explanation with supporting literature and subject matter expert (SEM) opinions.

Using this approach the reduction of score in Enzymax (test product) would be compared to the reduction of score with standard of care (reference product). The NIM (as limit that suggests the lowest acceptable effect of test product compared to reference) is defined as -10 units. It must be emphasized that the NIM is for presentation as any literature review or expert consultations have not been performed.

If we consider the effect of the reference product and the potential NIM, using 80% power and alpha=0.05, the following sample sizes can be considered:

	Reduction of NDI SF score				NIM	SS per arm
NDI SF 1	Reference	Test	Test-Reference	SD		
Ullah et al	5±18	5±18	0	18	-10	52
	5±18	4±18	-1	18	-10	64
	5±18	3±18	-2	18	-10	81
	5±18	2±18	-3	18	-10	105
Majeed et al	10±7	10±7	0	7	-10	9
	10±7	9±7	-1	7	-10	11
	10±7	8±7	-2	7	-10	14
	10±7	7±7	-3	7	-10	17

The Majeed et al, study might be too optimistic, therefore Ullah et al, study would be a more realistic scenario with 64 participants or 81 per arm. If we add 20% dropouts, that would be 80 or 100 participants per arm (160 or 200 total participants).

3. Chronic Pancreatitis

In most cases, studies with chronic pancreatitis as the primary condition evaluate symptoms and stool as the primary outcome.

We performed a quick draft evaluation of registered studies on clinicaltrials.gov and the following primary endpoints are the most frequent:

- **Single arm**
  - **EPI symptoms using questionnaire** - NCT04949828
- **Two arm parallel**
  - **Fecal fat percentage** - NCT01430234, NCT00400842, NCT00630279, NCT00705978, NCT00788593, and NCT04315311
  - **Pain** - NCT01159119

The provided studies (from Ana) have complex methodologies with screening, run-in period, double blind phase and open label extension. In the double-blind phase, participants will be allocated to drug or placebo, and after 7 days, the stool examination would be performed. The primary outcome Coefficient of Fat Absorption (CFA) is measured on day 7 and the safety will be evaluated through 52 weeks.

3.1. Superiority Approach

If we want to prove that the test product is superior to placebo (Enzymax vs. Placebo) this approach would be used. It would not be appropriate to assess the noninferiority of any product to placebo.

The Throat et al, study was used to assess the possible sample size for our study and to include possible variations.

Screenshots presents CFA which is the primary endpoint in most studies evaluating stool as the primary endpoint.

Table 3 | Unadjusted mean ± standard deviation values at baseline and end of double-blind period for CFA, CNA, and stool characteristics (full analysis sample)*

	Pancreatin (n = 34)*		Placebo (n = 27)*	
	n		n	
CFA (%)				
Baseline	34	66.5 ± 14.1	27	67.0 ± 14.0
End of double-blind phase	32	86.1 ± 7.5	24	72.9 ± 11.5
CNA (%)				
Baseline	32	78.8 ± 10.0	23	79.7 ± 7.2
End of double-blind phase	32	83.8 ± 6.9	24	81.7 ± 7.3

As shown, the CFA increases significantly in the Pancreatin arm and shows a small increase in the Placebo arm. The increase is statistically significantly higher in the Pancreatin arm.

Table 2 | Least squares mean and 95% CIs for change from baseline to end of double-blind phase (ANCOVA with baseline as covariate and treatment as a factor; full analysis sample)

	Pancreatin (n = 34)*	Placebo (n = 27)*	Treatment difference	P-value
CFA (%)	18.5 (15.8, 21.2)	4.1 (1.0, 7.2)	14.4 (10.3, 18.5)	0.001
CNA (%)	4.7 (3.0, 6.5)	0.8 (−1.3, 2.9)	4.0 (1.2, 6.7)†	0.005
Stool fat (g/day)	−19.8 (−23.0, −16.6)	−4.3 (−8.0, −0.6)	−15.5 (−20.4, −10.6)	0.001

ANCOVA, analysis of covariance; CFA, coefficient of fat absorption; CNA, coefficient of nitrogen absorption.

* Owing to missing stool samples, two subjects in the pancreatin group and three subjects in the placebo group were excluded from the CFA and stool fat analysis; four subjects in the pancreatin group and seven subjects in the placebo group were excluded from the CNA analysis.

† Slight difference in value compared with pancreatin–placebo due to rounding.

Using two-sided 0.05 alpha and 80% power, the following sample sizes per arm are:

	CFA Change from baseline		
	Test	Placebo	SS per arm
Common SD=20	18	3	29
	18	4	34
	18	5	39
	17	3	34
	17	4	39
	17	5	45
	16	3	39
	16	4	45
	16	5	53

As can be seen in the table above, using a conservative approach with high SD and several possible differences to the placebo, the sample size ranges from 29 to 53 per arm. The most realistic scenario is 39 or 45 per arm. The dropout rate in this study is expected to be smaller due to short time periods for the primary endpoint (7 days) so only a few more patients more would be sufficient to preserve 80% power.

3.2. Non-inferiority Approach

If we want to compare the test drug with the reference (for example Enzymax vs Creon), then the non-inferiority approach would be another option. In this case, we need to determine the NIM (presumably reference being 10% worse compared to the test, which means that the lower bound of 95% CI for the difference between test and reference product will not exceed -10%). The reference product (gold standard) is assumed to have effect provided in tables above (blue tables presented as figures on previous pages).

Using one sided 0.025 alpha, and 80% power the following sample sizes per arm are:

NIM=-10	Test-Ref	SS per arm
Common SD=14	0	32
	-1	39
	-2	50
	-3	64
Common SD=16	0	42
	-1	51
	-2	64
	-3	83
Common SD=18	0	52
	-1	64
	-2	81
	-3	105

Using SD=16, and effect of test one unit lower compared to reference, the sample size of 51 per arm is sufficient to prove non-inferiority of the test product vs. standard of care using -10% margin.

4. Conclusion

In conclusion, the two most feasible options for Enzymax Forte clinical trial design are summarized below.

4.1. Functional Dyspepsia

The use of Enzymax Forte in treatment of functional dyspepsia clinical trial is probably less complex, because the primary endpoint is symptom reduction and quality of life. Both are measured with questionnaires and the follow up is mostly based on patient reported outcome, especially some laboratory analyses included at the beginning and the end of trial. The duration should be at least several months, and the number of subjects depends on methodological approach, superiority trial or noninferiority trial. In superiority trial we need to prove that Enzymax is better than placebo and in noninferiority we must prove that Enzymax is similar or not worse than gold standard (for example, Creon). The estimated sample size based on the collected information is approximately 70 to 80 per group, or 140 to 160 participants.

4.2. Exocrine Pancreatic Insufficiency due to Chronic Pancreatitis

In chronic pancreatitis patients, clinical trial is probably more complex because we have stool examination as primary endpoint (the studies mostly measure Coefficient of Fat Absorption), but the primary endpoint is in short period (7 days in provided studies). Contrary to previous model (functional dyspepsia), laboratory analysis is the primary evidence of efficacy and at least two laboratory examinations must be performed, preferably even more. On the other hand, all studies in patients with chronic pancreatitis patients have long follow up (51 weeks in provided studies) to assess safety of the product. Patients who are on placebo should be transferred to test product due to ethical reasons and the symptoms and signs should be followed up only on test group, not placebo, till the end of trial. The regulatory question is can we use efficacy data from 7 days efficacy epoch of the trial, or do we need to finish the trial and then to claim efficacy and safety. The first option is very fast, but the second is almost one year, according to the studies provided. The estimated sample size based on the collected information is approximately 45-50 per group, or 90 to 100 participants.

4.3. Final Considerations and Limitations

All the calculations and assumptions are just for informative purposes. When we define the methodology of the study and collect more information from the clinical trials and publications and perform additional research, we can calculate sample size more precisely.

Additionally, the suggestions above consider a conservative approach with a larger standard deviation that in the studies analyzed, resulting in a larger sample size, and decreased risk of the trial not meeting the primary endpoint (trial failure).