

---

# Sample size for the verification of possible Interleukin-6 circadian rhythm in Alpha-1 Antitrypsin deficient patients

DOCUMENT: SAP-2021-020-LH-v02

From: Felipe Figueiredo To: Lameese Alhaddah

2022-12-06

## TABLE OF CONTENTS

1	ABBREVIATIONS.....	2
2	CONTEXT.....	2
	2.1 Objectives.....	2
	2.2 Hypotheses.....	2
3	DATA.....	2
	3.1 Raw data.....	2
	3.2 Analytical dataset.....	3
4	STUDY PARAMETERS.....	3
	4.1 Study design.....	3
	4.2 Inclusion and exclusion criteria.....	3
	4.3 Exposures.....	3
	4.4 Outcomes.....	4
	4.5 Covariates.....	4
5	STATISTICAL METHODS.....	4
	5.1 Statistical analyses.....	4
	5.1.1 Descriptive analyses.....	4
	5.1.2 Inferential analyses.....	4
	5.1.3 Statistical modeling.....	4
	5.1.4 Missing data.....	4
	5.2 Significance and Confidence Intervals.....	5
	5.3 Study size and Power.....	5
	5.3.1 Large effect size.....	5
	5.3.2 Medium effect size.....	6
	5.3.3 Small effect size.....	6
	5.3.4 Effect size for a given sample size.....	7
	5.4 Statistical packages.....	7
6	OBSERVATIONS AND LIMITATIONS.....	8
7	REFERENCES.....	8
8	APPENDIX.....	8
	8.1 Availability.....	8

---

---

# Sample size for the verification of possible Interleukin-6 circadian rhythm in Alpha-1 Antitrypsin deficient patients

## Document version

Version	Alterations
01	Initial version
02	Added detectable effect for a given sample size

## 1 ABBREVIATIONS

- BMI: body mass index
- CI: confidence interval
- d: Cohen's D effect size
- IQR: interquartile range
- SD: standard deviation

## 2 CONTEXT

### 2.1 Objectives

Determine the sample size to detect the difference of the daily range of Interleukin-6 blood levels between Alpha-1 Antitrypsin deficient and healthy patients.

### 2.2 Hypotheses

The circadian rhythm in patients induces a detectable amplitude between peak and lower daily IL6 levels. The average IL6 daily amplitudes differ between Alpha-1 Antitrypsin deficient and healthy patients.

## 3 DATA

### 3.1 Raw data

A unique study ID will be created for each participant to allow tracking of consecutive IL6 blood level measurements. IL6 data of each participant will be collected at three different times per day. Each datum will have its respective date and time of specimen

Analytical Plan (SAP)

collection recorded. Participant demographic characteristics to be collected are the group to which they belong (AAT-deficient or healthy), the date of birth, sex and BMI at the date of inclusion.

After the cleaning process the raw data will include 13 variables. Table 1 shows the structure of the raw dataset.

**Table 1** Raw dataset structure.

id	group	sex	dob	bmi	date	il6_time1	il6_measure1	il6_time2	il6_measure2	il6_time3	il6_measure3	outcome
1												
2												
3												
...												
N												

### 3.2 Analytical dataset

Age at inclusion will be computed between the date of birth and the date of the first specimen collection.

For each patient the peak level and the lower level of IL6 measurements will be calculated. The peak level (respectively, lower level) of each patient's IL6 will be calculated as the average of all peak level (respectively, lower level) measurements. The amplitude of each patient's IL6 blood level will be calculated as the difference between peak level and lower level.

All variables in the analytical set will be labeled according to the raw data provided and values were labeled according to the data dictionary for the preparation of production-quality results tables and figures.

## 4 STUDY PARAMETERS

### 4.1 Study design

Cross-sectional.

### 4.2 Inclusion and exclusion criteria

N/A

### 4.3 Exposures

Patients with Antitrypsin deficiency will be compared to controls.

## 4.4 Outcomes

**Specification of outcome measures** (Zarin, 2011):

1. (Domain) Circadian rythm
2. (Specific measurement) Amplitude of IL6 blood level
3. (Specific metric) End value
4. (Method of aggregation) Average

### Primary outcome

The primary outcome for the detection of a circadian rhythm will be defined as the amplitude of the IL6 blood level which will be calculated as the difference between the peak level and the lower level of daily measurements.

The secondary outcome will be defined as the the difference between the average IL6 daily amplitude between healthy and AAT-deficient patients.

## 4.5 Covariates

No covariates will be considered for the sample size calculation. Age, sex and BMI will be used to adjust the estimates of the primary and secondary outcomes. Both raw and adjusted estimates will be computed.

# 5 STATISTICAL METHODS

## 5.1 Statistical analyses

### 5.1.1 Descriptive analyses

N/A

### 5.1.2 Inferential analyses

N/A

### 5.1.3 Statistical modeling

N/A

### 5.1.4 Missing data

No missing data imputation will be performed. All evaluations will be performed as complete case analyses.

## 5.2 Significance and Confidence Intervals

All analyses will be performed using the significance level of 5%. All significance hypothesis tests and confidence intervals computed will be two-tailed.

## 5.3 Study size and Power

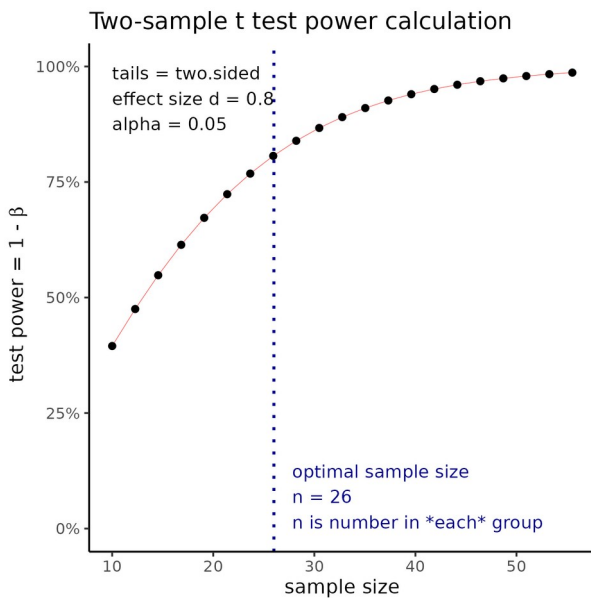
This sample size calculation assumes the exposed and control groups will be balanced in equal sizes.

We will consider three possible scenarios that the main researcher could choose from, all using Cohen's  $d$  relative effect size (Cohen, 1988). This is a relative metric, accounting for both the mean and SD of the observed clinical measurements, thus simplifying the interpretation of the clinical effect into a single figure. A large effect requires less samples to be detected than a small one, but is typically considered as rarer effect.

Here we consider the small effect size defined as Cohen's  $d = 0.2$ , the medium as  $d = 0.5$  and the large effect size as  $d = 0.8$ .

### 5.3.1 Large effect size

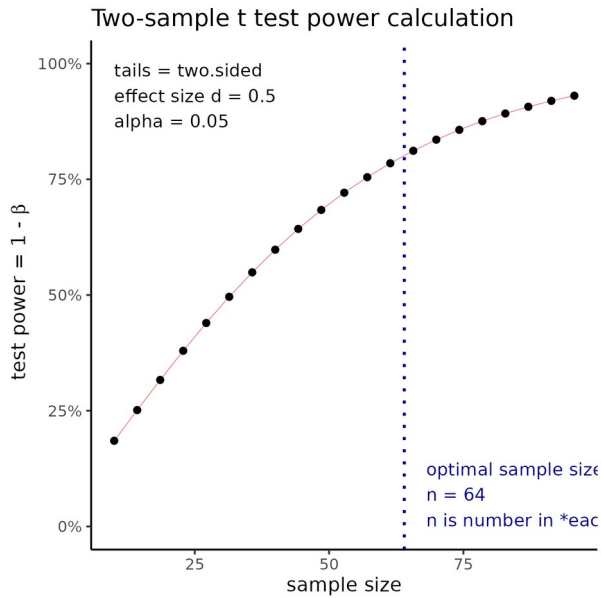
With sample size of 26 per group a t test would be able to detect an effect size of  $d = 0.8$  with 80% power at a 5% significance level ( $\alpha$ ). For this effect size the total study size would require the inclusion of 52 participants.



**Figure 1** Sample size required to detect a large effect size.

### 5.3.2 Medium effect size

With sample size of 64 per group a t test would be able to detect an effect size of  $d = 0.5$  with 80% power at a 5% significance level ( $\alpha$ ). For this effect size the total study size would require the inclusion of 128 participants.

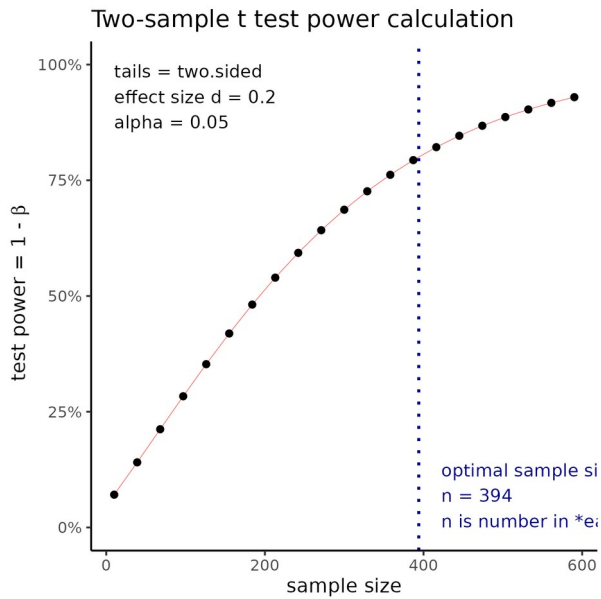


**Figure 2** Sample size for a medium effect size.

### 5.3.3 Small effect size

With sample size of 394 per group a t test would be able to detect an effect size of  $d = 0.2$  with 80% power at a 5% significance level ( $\alpha$ ). For this effect size the total study size would require the inclusion of 788 participants.

Analytical Plan (SAP)



**Figure 3** Sample size required to detect a small effect size.

### 5.3.4 Effect size for a given sample size

If the maximum sample size feasible is  $N = 10$  (divided into two groups of 5), then the minimum effect size that could be detected by this study would be  $d = 2.02$ .

This effect can be interpreted as the average amplitude (difference between peak value and low value) being twice as large as the SD for a t test to detect it.

## 5.4 Statistical packages

This analysis will be performed using statistical software R version 4.2.1.

## 6 OBSERVATIONS AND LIMITATIONS

### Direction of the difference

The AAT deficient patient IL6 blood levels may be hypothesized to be lower (higher) than the levels observed healthy patients. If this hypothesis holds, a left tailed (right tailed) test will have more statistical power to detect significant differences than a two-tailed test. If there is support in the literature for an educated guess on the direction of the difference this could improve the chance of having significant p-values in the final analysis, at the risk of not detecting a difference in the other direction.

For example, if the hypothesis is specified in the left direction (group of interest has lower levels of IL6 than the healthy group) and the actual observed levels indicate that the group of interest has higher levels of inflammatory biomarkers, then the actual difference will likely not be significant in the study. It is recommended that such decision be made with caution.

## 7 REFERENCES

- **SAR-2021-020-LH-v01** – Sample size for the verification of possible Interleukin-6 circadian rhythm in Alpha-1 Antitrypsin deficient patients
- Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd Ed.). New York: Routledge.
- Zarin DA, et al. The ClinicalTrials.gov results database – update and key issues. N Engl J Med 2011;364:852-60 (<https://doi.org/10.1056/NEJMsa1012065>).
- Gamble C, et al. Guidelines for the Content of Statistical Analysis Plans in Clinical Trials. JAMA. 2017;318(23):2337–2343 (<https://doi.org/10.1001/jama.2017.18556>).

## 8 APPENDIX

This document was elaborated following recommendations on the structure for Statistical Analysis Plans (Gamble, 2017) for better transparency and clarity.

### 8.1 Availability

All documents from this consultation were included in the consultant's Portfolio.

The portfolio is available at:

<https://philsf-biostat.github.io/SAR-2021-020-LH/>