Clinical Trials

Adaptive Designs for Detecting Gene Signatures

Angeliki Skandali

September 13, 2019

Athens University of Economics and Business

Table of contents

Introduction to clinical trials

Introduction to clinical trials

- What is a clinical trial?

  An experiment testing the relationship between a medical intervention and health outcome on human volunteers.

Introduction to clinical trials

- What is a clinical trial?

  An experiment testing the relationship between a medical intervention and health outcome on human volunteers.

- Their contribution?

  They have contributed in a great extent to the evolution of medical research, through the examination of treatment efficacy.

Introduction to clinical trials

- What is a clinical trial?

  An experiment testing the relationship between a medical intervention and health outcome on human volunteers.

- Their contribution?

  They have contributed in a great extent to the evolution of medical research, through the examination of treatment efficacy.

- Since when?

  The first reported clinical trial was in 1747, when James Lind tried to cure twelve sailors from scurvy [Lind, 1980].

Introduction to clinical trials

- What is a clinical trial?

  An experiment testing the relationship between a medical intervention and health outcome on human volunteers.

- Their contribution?

  They have contributed in a great extent to the evolution of medical research, through the examination of treatment efficacy.

- Since when?

  The first reported clinical trial was in 1747, when James Lind tried to cure twelve sailors from scurvy [Lind, 1980].

- What about the types of clinical trials?

  Clinical trials can be treatment, prevention, diagnostic, screening or quality of life trials.

Phases of clinical trials

When clinical research is used to evaluate medications, devices or treatments, the experiment is conducted in phases:

Phase I

Researchers test an experimental treatment in a small group of people, so as to find the best dose of the drug and identify side effects.

Phase II

The experimental treatment is given to a larger group of people at a fixed dose in order to determine the efficacy and safety of the drug.

Phase III

The effectiveness of the new treatment is compared to to a standard one, also known as control arm.

Phase IV

Researchers track the treatment's safety in the general population, seeking more information about its benefits and optimal use.

Defining the research question

Before the beginning of a clinical trial, investigators have to define the **primary research question.**

Defining the research question

Before the beginning of a clinical trial, investigators have to define the **primary research question.**

How a research question is formulated?
It is formulated into a hypothesis stating what investigators expect to find. For instance: "Do subjects treated with agent A have better outcomes than those treated with agent B?"

Defining the research question

Before the beginning of a clinical trial, investigators have to define the **primary research question.**

How a research question is formulated?
It is formulated into a hypothesis stating what investigators expect to find. For instance: "Do subjects treated with agent A have better outcomes than those treated with agent B?"

What kind of hypothesis is used in cancer trials?
In cancer trials when new treatments are compared with existing ones, the hypothesis is defined by the difference that participants had for example in overall survival.

Defining the research question

Before the beginning of a clinical trial, investigators have to define the **primary research question.**

How a research question is formulated?
It is formulated into a hypothesis stating what investigators expect to find. For instance: "Do subjects treated with agent A have better outcomes than those treated with agent B?"

What kind of hypothesis is used in cancer trials?
In cancer trials when new treatments are compared with existing ones, the hypothesis is defined by the difference that participants had for example in overall survival.

Which design elements are associated with the research question?
The population, intervention, comparator, outcome, and time-frame are elements determined along with the research question. These elements are commonly referred to as the PICOT format [Giuffrida, 2016].

Description of the current study

Description of the problem

The problem

- In traditional randomized clinical trials, the way to examine the efficacy of a treatment is a hypothesis test over **all the population**.
- Due to the **heterogeneity** of most human cancers, only a **subset** of patients can get a positive impact from a given treatment.
- As a result, this situation often leads to the **withdrawal** of new treatments, without knowing whether there were truly beneficial or not.

Shift towards traditional designs

The **solution** to this problem is a **Biomarker Adaptive Design**.

What is a Biomarker Adaptive Design?

- Unlike traditional designs, a Biomarker Adaptive Design is a **2-stage design**.
- The design depends on prognostic covariates, like as a set of biomarkers or a gene-signature that contribute to the:
  1. Identification of a more suitable target subpopulation.
  2. The evaluation of the treatment efficacy in both the subpopulation and all population.

The goals of our study

Our goals

Our goal is to built a biomarker adaptive design that could:

1. Develop a **predictive gene signature**, capable of identifying a subpopulation, which benefits most from the new treatment.
2. Examine the treatment efficacy in this **subpopulation**, if the treatment effect is not identified in the overall population.

Material & Methods

Introduction to the design

- Our study will be a simulation over a phase III clinical trial with **survival outcome** that **randomly** assigns patients in two different arms (A: experimental & standard, B: standard treatment)
- It will use DNA microarrays expressions, like in the ASD design of Freidlin and Simon [Freidlin and Simon, 2005].
- The design consists of a **gene identification**, a **classifier development** and at last **performance assessment**.
- In order to validate the results of the study: If N patients are participating in the trial, N_1 patients will be used for the gene identification (i.e., learn set) and the rest $N_2 = N - N_1$ for the classifier development and performance assessment (i.e., confirm set).

Gene identification

- For a given patient $i \in 1, ..., N$ and for the $j \in 1, ..., L$ gene, let x_{ij} denote the expression level, and t the kind of treatment a patient is assigned:

$$t = \begin{cases} 1, & \text{new and standard treatment} \\ 0, & \text{standard treatment} \end{cases}.$$

- For every $j \in 1, ..., L$ and only for the patients included in the learn set, a Cox Model is fitted, including the **treatment** t and the **interaction** $x_{ij}t$.
- The rule that determines if the g gene will be included in the signature is whether the gene-by-treatment interaction term was **statistically significant**. The level of significance is set at 0.01.

Classifier development

At this stage, only the patients from the **confirm set** will be classified as **sensitive** or not. The classification is organized in several steps:

Classifier development

At this stage, only the patients from the **confirm set** will be classified as **sensitive** or not. The classification is organized in several steps:

Step 1: For every significant gene $j \in \mathcal{S}$, fit the same cox model and save the coefficients from the treatment effect (b_{1j}) and interaction term (b_{2j}).

Classifier development

At this stage, only the patients from the **confirm set** will be classified as **sensitive** or not. The classification is organized in several steps:

Step 1: For every significant gene $j \in \mathcal{S}$, fit the same cox model and save the coefficients from the treatment effect (b_{1j}) and interaction term (b_{2j}).

Step 2: For each patient in confirm set i and for every gene $j \in \mathcal{S}$ calculate the hazard ratio HR_{ij} using the b_{1j} and b_{2j} coefficients, with the formula

$$HR_{ij} = \exp^{b_{1j} + b_{2j} x_{ij}}$$

Classifier development

At this stage, only the patients from the **confirm set** will be classified as **sensitive** or not. The classification is organized in several steps:

Step 1: For every significant gene $j \in \mathcal{S}$, fit the same cox model and save the coefficients from the treatment effect (b_{1j}) and interaction term (b_{2j}).

Step 2: For each patient in confirm set i and for every gene $j \in \mathcal{S}$ calculate the hazard ratio HR_{ij} using the b_{1j} and b_{2j} coefficients, with the formula

$$HR_{ij} = \exp^{b_{1j} + b_{2j} x_{ij}}$$

Step 3: Classify the i-th patient as sensitive if following condition is satisfied

$$\sum_{j \in \mathcal{S}} \left(HR_{ij} \leq R \right) \geq G$$

Performance assessment

Which are the ways to examine the performance of the design?

First is the **accuracy** of the model to classify the patients properly and second the **power** to detect a meaningful difference.

The power of the design depends on **two comparisons** (i.e., hypothesis tests):

1. A hypothesis test for the overall comparison of the new treatment with the control.
2. A hypothesis test for the comparison of the new treatment versus control in the identified as sensitive subpopulation.

A p-value is calculated for **each test**, so the two p-values α_1 and α_2 will be adjusted in a way that the overall type I error is no more than 5%.

Data Generation

Gene expression simulation

- In the simulation from the L=10000 genes, K=10 were generated as **"predictive"** genes.
- Assume that N_s is the number of sensitive patients.
- The **gene expression** levels x_{ij} $i \in 1, \ldots, N$, $j \in 1, ..., L$ were generated as matrices using a multivariate normal distribution:
  1. One $N_s \times K$ matrix with values generated by **N(0,0.25)** for the **predictive genes** in **sensitive patients**
  2. One $(N - N_s) \times K$ matrix with values generated by **N(0,0.01)** for the **predictive genes** in **non-sensitive patients**
  3. One $N \times (L - K)$ matrix with values generated by **N(0,0.1)** for the **non-predictive genes** in all patients.

Gene expression simulation

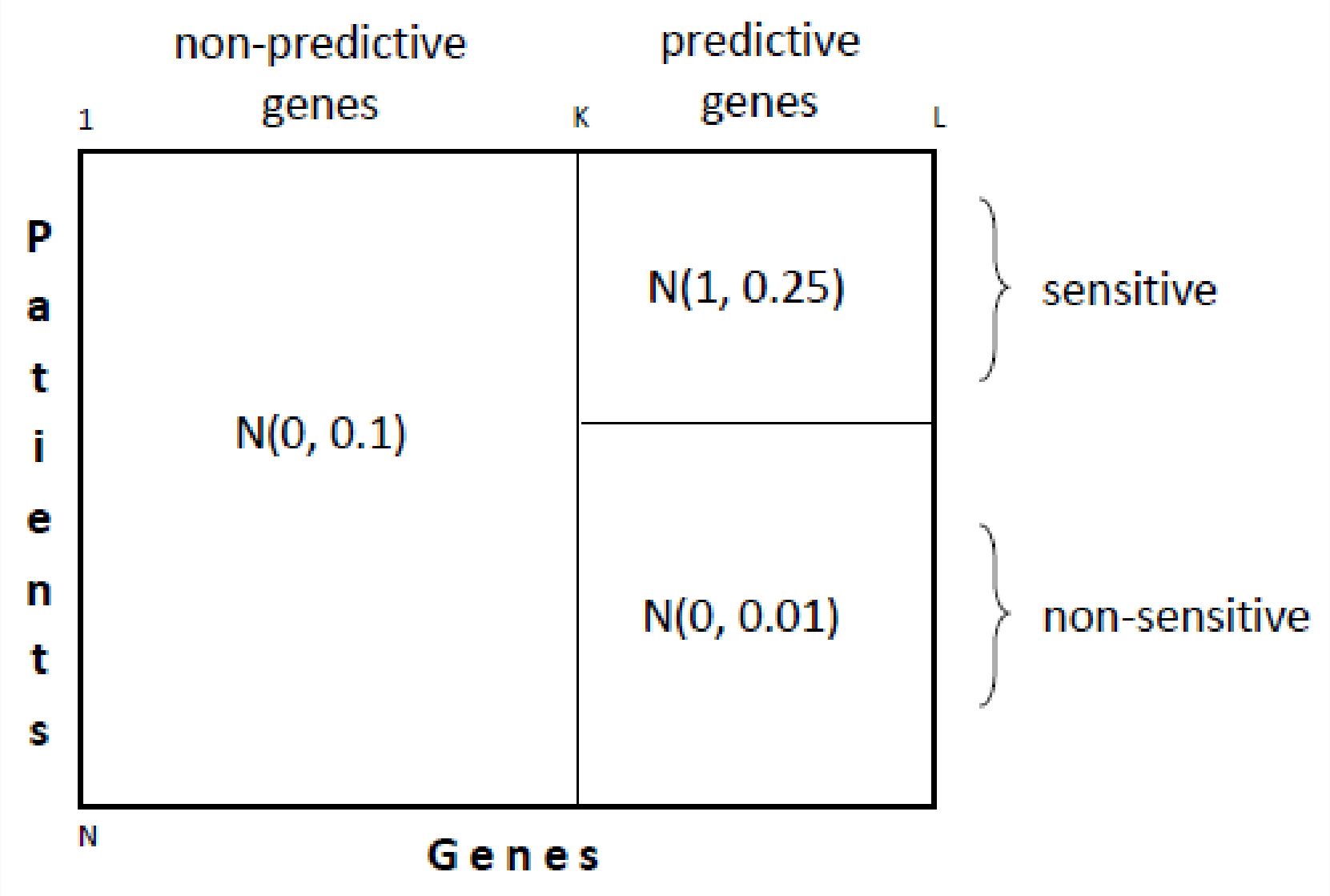


Figure 1: Generation of expression genes levels

Survival time for control group

We will simulate **survival time** with constant hazard in 3 segments and 2 breakpoints, with **piecewise exponential** distribution as a baseline **survival function**.

- For the control group:

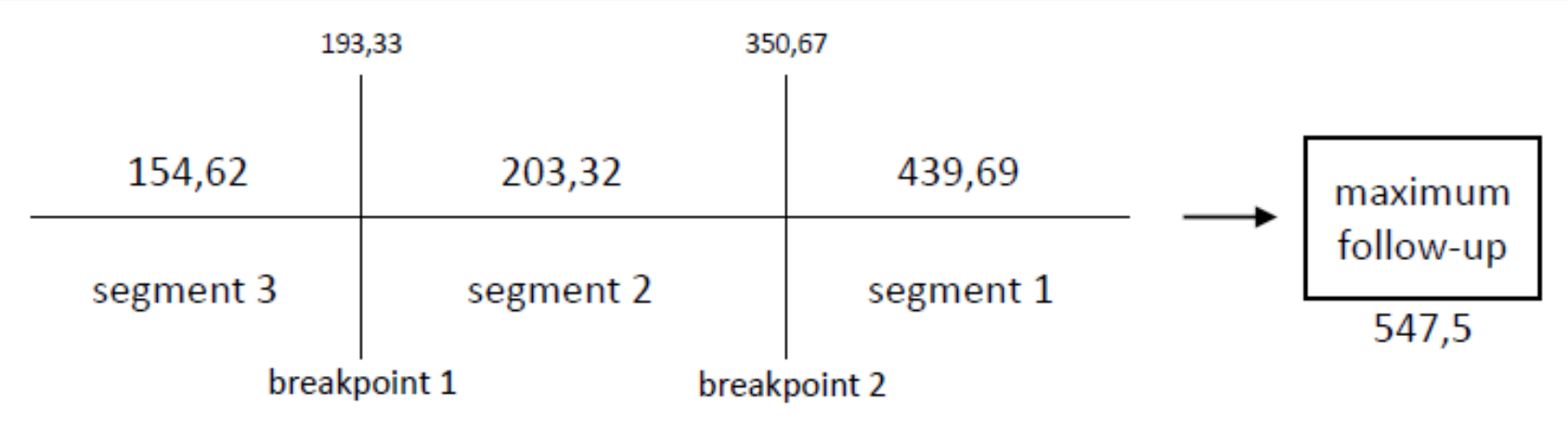


Figure 2: Median survival times in control group

Survival time for control group

From the median survival times, we can calculate the exponential rates in each segment:

$$\begin{bmatrix} m_1 = 439.64 \\ m_2 = 203.32 \\ m_3 = 154.62 \end{bmatrix} \Longrightarrow \begin{bmatrix} \rho_1 = \frac{ln(2)}{m_1} = 0.0015 \\ \rho_2 = \frac{ln(2)}{m_2} = 0.003 \\ \rho_3 = \frac{ln(2)}{m_3} = 0.004 \end{bmatrix}$$

and so for the **control group**, the rates are:

$$\rho(t) = \begin{cases} 0.004, & 0 < t \leq 193.33 \\ 0.003, & 193.33 < t \leq 350.67 \\ 0.0015, & t \geq 350.67 \end{cases}$$

Survival time for treatment group

- For the treatment group: Sensitive (M+) and non-sensitive (M-) patients have **different** survival times and this is determined by the HR over control group. We evaluate 2 scenarios:

Scenario	HR in M+	HR in M-
Strong	0.65	1.1
Moderate	0.65	1

Table 1: Hazard ratios in different predictive scenarios

Having the HR for M+ and M- group, it is possible to simulate from the piecewise exponential by **multiplying** the control rate with the HR for each group.

Censoring

1. The censoring rate is adjusted at 20% prior to maximum follow up, meaning that 20% of the patients left the study before experienced the event or for unknown reasons their behavior is not recorded.
2. If a patient is not predicted to experience the event after the maximum follow up, will be automatically censored and thus the percentage of censoring rises up to approximately 30%.

Results

The evaluated parameters

For both scenarios, results in terms of **statistical power** are evaluated by alternating the following parameters:

Parameters	Possible Values or Levels
Sample Size	1500 | 1800 | 2000 | 2300 | 2500
Learn/Confirm set allocation	40/60 | 50/50 | 60/40 | 70/30
Splits of α	0.04/0.01 | 0.03/0.02 | 0.025/0.025
R	0.7 | 0.75 | 0.8 | 0.85
G	2 | 3 | 5
Sensitive percentage	10 | 25 | 40

Table 2: The parameters evaluated in this study

Sample size evaluation

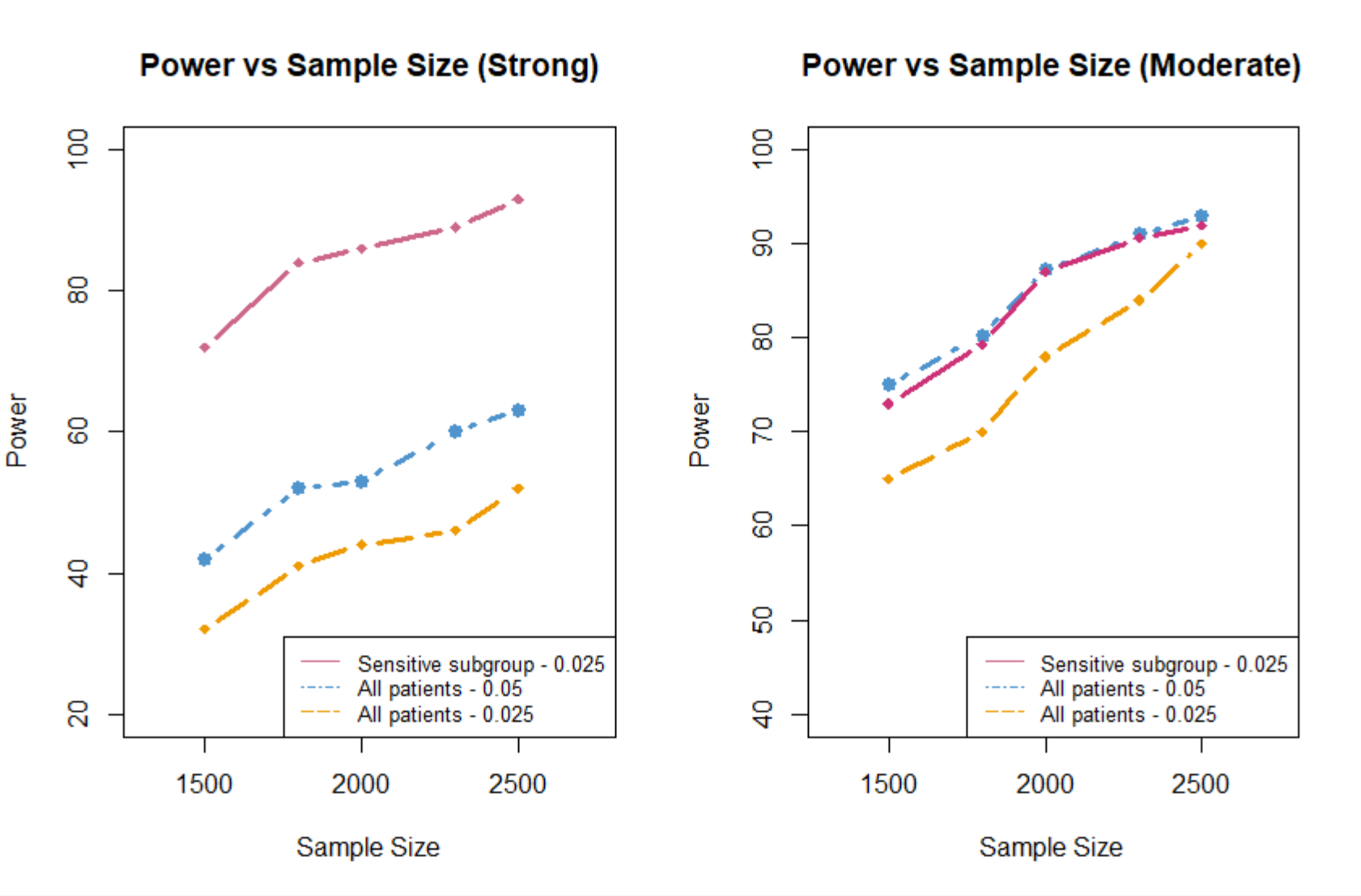


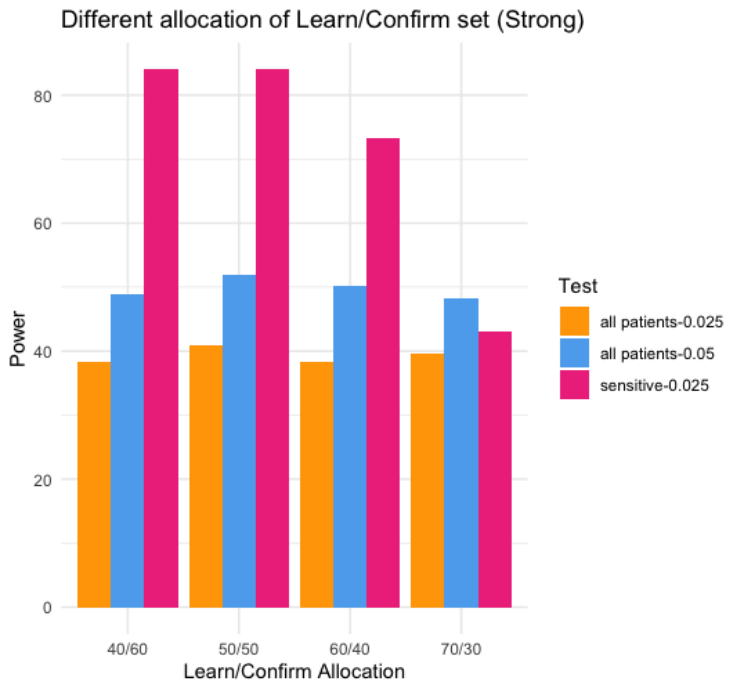
Figure 3: Power evaluation (%) for sample size alterations

Evaluation of R tuning parameter

	Scenario	Strong			Moderate		
	Patients	All	All	Sensitive	All	All	Sensitive
	P-value	0.05	0.025	0.025	0.05	0.025	0.025
R	0.7	44	34	71	78	70	71
	0.75	52	41	**84**	80	70	**79**
	0.8	48	36	57	82	72	59
	0.85	53	43	34	82	75	46

Table 3: Power evaluation (%) for different tuning parameter R

Evaluation of different learn/confirm allocations



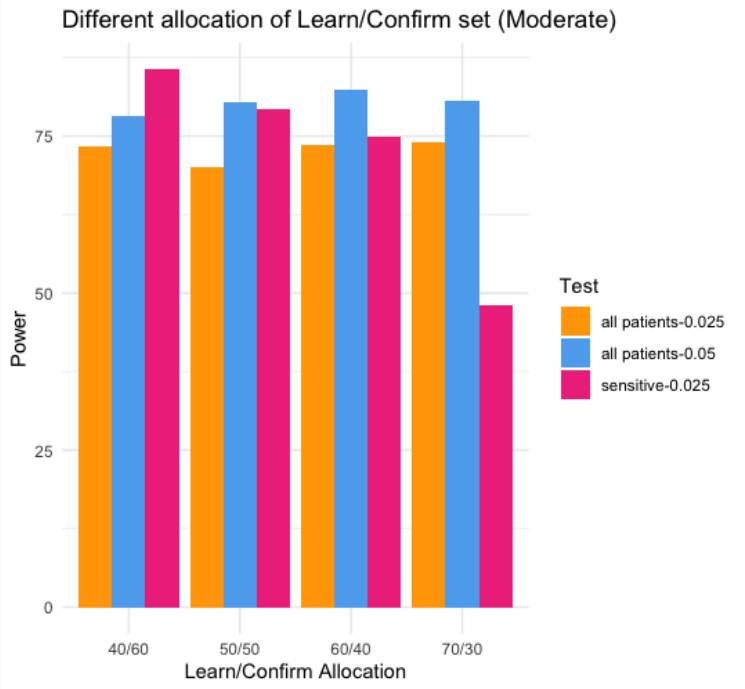


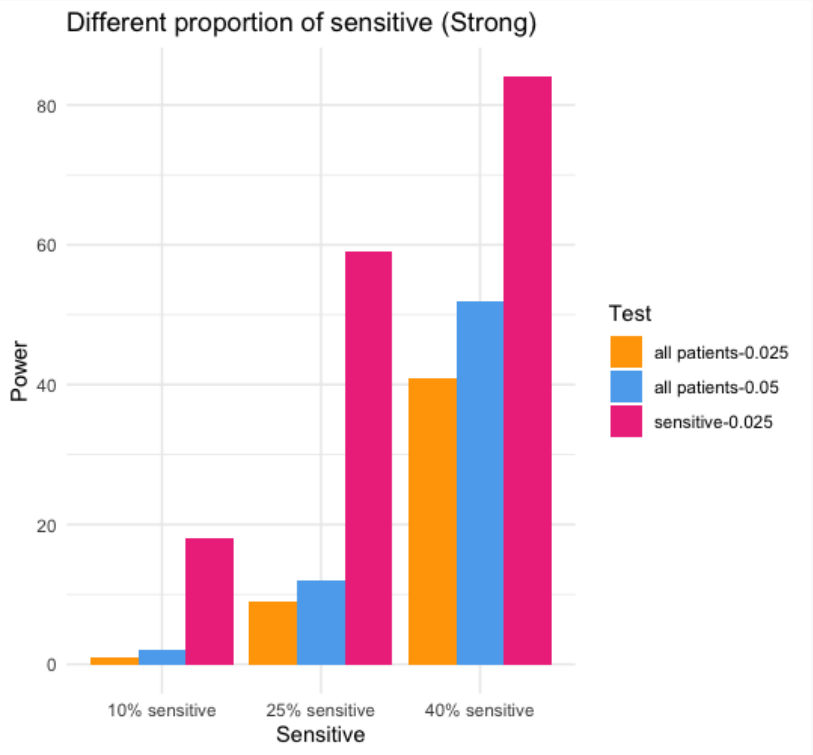
Figure 4: Power evaluation (%) for different learn/confirm allocation

Evaluation of different splits of α

	Scenario	Strong			Moderate		
	Patients	All	All	Sensitive	All	All	Sensitive
	P-value	0.05	α_1	α_2	0.05	α_1	α_2
$\alpha_1 \vert \alpha_2$	0.04|0.01	51	46	72	82	79	67
	0.03|0.02	46	38	79	85	80	**88**
	0.025|0.025	52	41	**84**	80	70	79

Table 4: Power evaluation (%) for different split of significance level

Evaluation of different percentage of sensitive patients



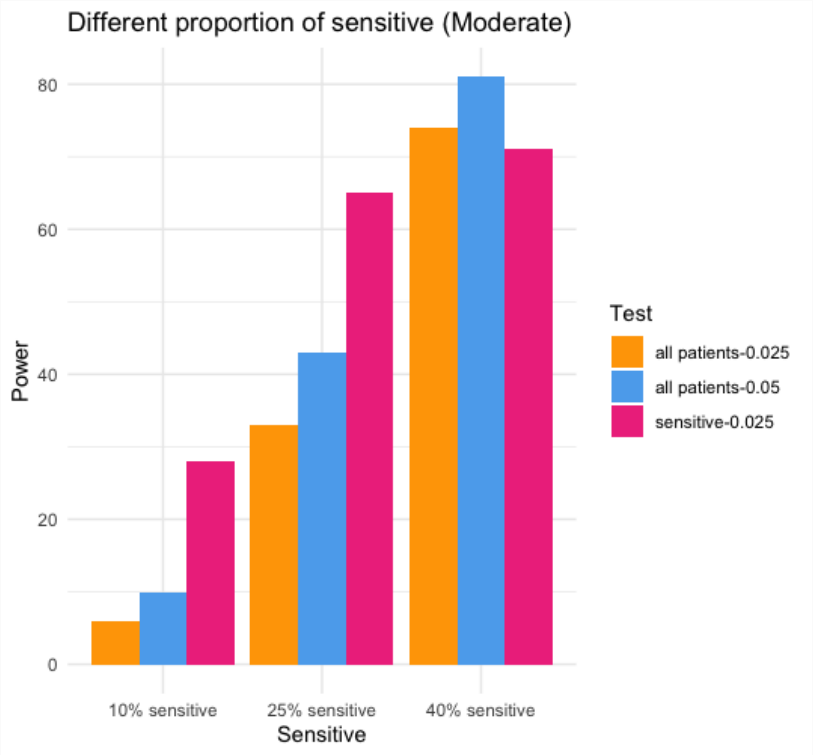


Figure 5: Power evaluation (%) for different percentage of sensitive patients

Evaluation of G tuning parameter

	Scenario	Strong			Moderate		
	Patients	All	All	Sensitive	All	All	Sensitive
	P-value	0.05	0.025	0.025	0.05	0.025	0.025
G	**2**	53	42	**86**	81	72	**84**
	3	52	41	84	80	70	79
	5	55	43	78	74	63	75

Table 5: Power evaluation (%) for different tuning parameter G

Discussion

Discussion

We implemented 2 scenarios and after examining several parameters their best performance for 1800 patients with R=0.75 and G=2 was:

Scenario	sensitive	α	learn/confirm	Power
Strong	40%	0.025/0.025	50/50	86% over 53%
Moderate	40%	0.03/0.02	40/60	92% over 84%

Table 6: Best performance for each scenario.

Further enhancement of the design:

1. Different combinations of L and K that could offer better results.
2. A more efficient way of internal and external validation than splitting the sample size in different proportions. That could be re-sampling techniques like cross-validation and bootstrap.

Thank you for your attention!

References


Freidlin, B. and Simon, R. (2005).
Adaptive signature design: an adaptive clinical trial design for generating and prospectively testing a gene expression signature for sensitive patients.
Clinical cancer research, 11(21):7872–7878.


Giuffrida, M. A. (2016).
Defining the primary research question in veterinary clinical studies.
Journal of the American Veterinary Medical Association, 249(5):547–551.


Lind, J. (1980).
A Treatise on the Scurvy.
Classics of Medicine Library.


Mi, G. (2017).
Enhancement of the adaptive signature design for learning and confirming in a single pivotal trial.