V: 21/03/2023

# 3. Simulations and Results

we evaluate

In the following chapter, we assess the performance of the proposed methods in a simulation study. For each model, the type I error rate and statistical power ~~will be examined~~ under a wide range of scenarios. In Section 3.1, we describe the general design of the platform trials considered in the study and the chosen design parameters. In Section 3.2 we present the results from the study, discuss the ~~behaviour~~ properties of the examined methods and the influence of the design parameters.

influence with respect to what? or on what?

## 3.1. Considered Designs

We consider a platform trial with $K$ experimental treatment arms that enter the trial sequentially and a control group that is common to all treatment arms. The timing of adding of the treatment arms is given by $d = (d_1, \ldots, d_K)$, where $d_i$ indicates how many patients had already been enrolled to the trial by the time treatment $i$ entered the platform. $d_1$ is always set to 0 to ensure that the platform trial starts with at least one experimental treatment (for illustration see Fig. 2.2). We assume equal sample sizes in all treatment arms ($n_k = n = 250, \forall k = 1, \ldots, K$) and an allocation ratio of $1 : 1 : \ldots : 1$ in each period. Patients were assigned to arms following block randomization with block sizes of $2 \cdot (\#\text{active arms} + 1)$ in every period. The data was drawn from a normal distribution according to:

*

#

$$y_j \sim \mathcal{N}(\mu, \sigma^2)$$

with

$$\mu = \eta_0 + \sum_{k=1}^{K} \theta_k \cdot I(k_j = k) + f(j)$$

and

$$\sigma^2 = 1$$

where $\eta_0$ and $\theta_k$ are response in the control arm and the effect of treatment $k$. Moreover, time trends of various strengths and shapes may be present in the trial. The time trends are denoted by the function $f(j)$ and their magnitude is given by $\lambda$. Following time trend patterns are considered: ✓

(*)

We distinguish three settings in which we vary the design according to the objective

- Setting 1: Trial with $K = 10$ arms, where we evaluate arm $M=5$. In this setting we aim at evaluating the generalization of the model-based approach ...
- Setting 2: Trial with $K = 4$, where we ....
  In this setting, we aim at comparing the def. of time as time-int. with time periods...
- Setting 3: Trial with $K = 7$, where ...
  In this setting, ...

---

(#) To generate trial data, we ...

( then explain how to simulate data and the design assumptions )

- linear time trend: $f(j) = \lambda \cdot \frac{j-1}{N-1}$, where $N$ indicates the total sample size in the trial
- stepwise time trend: $f(j) = \lambda \cdot (w_j - 1)$, where $w_j$ indicates how many treatment arms have already entered the trial at the time patient $j$ was enrolled
- inverted-u trend: $f(j) = \lambda \cdot \frac{j-1}{N-1}(I(j \leq N_p) - I(j > N_p))$, where $N$ indicates the total sample size in the trial and $N_p$ is the point at which the trend turns from positive to negative in terms of the sample size
- seasonal trend: $f(j) = \lambda \cdot \sin\big(\psi \cdot 2\pi \cdot \frac{j-1}{N-1}\big)$, where $N$ indicates the total sample size in the trial and $\psi$ determines how many cycles the time trend should have

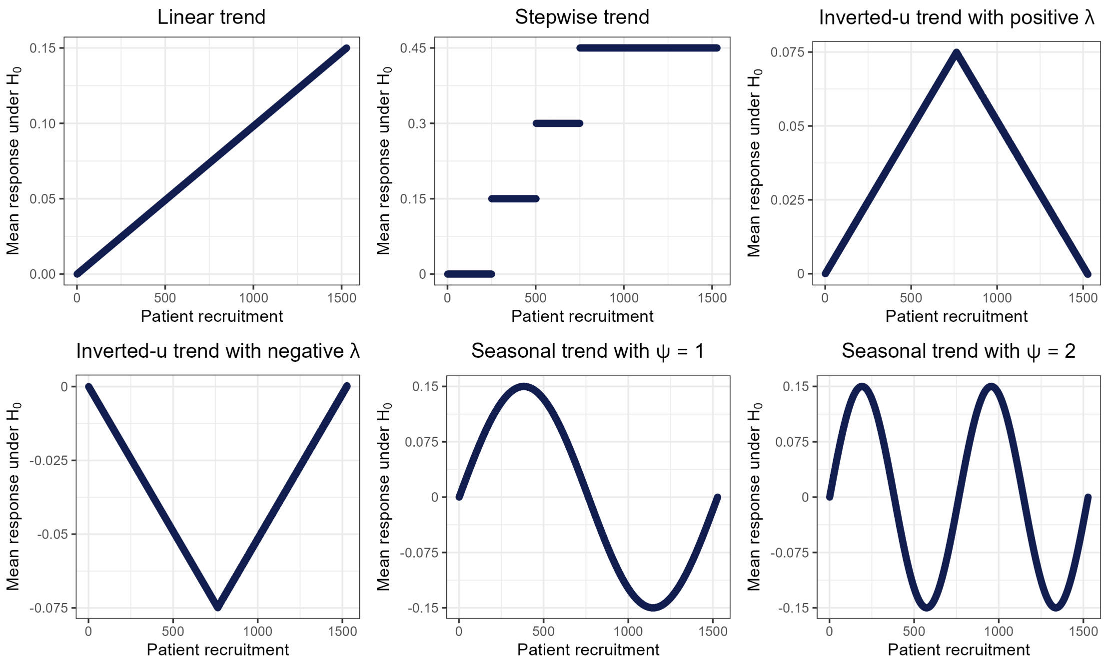


Figure 3.1.: Mean responses under the null hypothesis under time trends of different patterns and strength of $\lambda = 0.15$.

Under the linear time trend the mean response linearly increases with slope $\lambda$ over time ~~throughout the whole~~ trial, while under the stepwise time trend there is a jump in the mean response of size $\lambda$ every time a new arm is added to the trial. In the case of inverted-U time trend, the mean response linearly increases (with slope $\lambda$) until the sample size has reached $N_p$, and linearly decreases afterwards. Seasonal trend may consists of multiple cycles, where the response increases at first and decreases afterwards, while the respective peaks of the cycles correspond to $\lambda$ or $-\lambda$. Mean responses under the null hypothesis under the considered time trend patterns are illustrated in Fig. 3.1.

Maybe you can explain this further.

It is assumed that the equality assumption holds, hence that the time trends affect all arms in the same way.

In all cases we use underlying response of zero for the control arm ($\eta_0 = 0$) and a treatment effect of $\theta_k = 0.25$ for treatment arms under the alternative hypothesis. This effect size was chosen such that the separate approach achieves approximately 80% power with the given sample size at a one-sided significance level $\alpha = 0.025$.

The strength of the time trend $\lambda$, the timings of adding individual treatment arms $d$, as well as the length of the calendar time units were varied across the scenarios in order to investigate the impact of these parameters on the considered metrics. which values?

We simulated 100,000 replicates of each scenario to estimate the type I error rate and statistical power.

→ Add table to summarize the simulation designs and parameters considered

## 3.2. Results

We present the results in three sections, corresponding to three different aims of the simulation study. For each aim, we consider a different simulation scenario.

Firstly, in Section 3.2.1, we generalize the model-based approach proposed in [2] to trials with more than two experimental arms and examine in particular the impact of overlaps between arms on the operating characteristics of the extended linear model. ← here you are evaluating the extension was in CH2

Secondly, we evaluate the calendar time definition of the time covariate in Section 3.2.2. Here, we compare the performance of models with calendar time and period adjustment and discuss the optimal length of the calendar time intervals.

Lastly, we assess the performance of the newly proposed flexible modeling approaches for incorporating non-concurrent controls, i.e. linear mixed models and spline regression. There results are shown in Section 3.2.3. ✓

[illegible] (*)

SETTING 1:

### 3.2.1. Extension of regression model to trials with multiple arms

Consider a platform trial with 10 experimental treatment arms and a shared control arm, where the experimental arm $i$ enters the trial after $d_i = d \cdot (i-1)$ patients have been recruited to the platform, as illustrated in Fig. 3.2. Note that $d$ determines the amount of overlapping sample size between the treatment arms. If $d = 0$, all arms join and leave the trial simultaneously, resulting in a standard multi-arm trial and a total overlap between the arms. If $d = 2n$, a new treatment arm enters the trial once the previous one finishes, so that there is only one active experimental arm at a time, and hence no overlap between them.

We vary the $d$ from 0 to $2n$ ($d \in \{0, 100, 200, 300, 400, 500\}$) in order to evaluate the effect of the overlap on the type I error rate and statistical power. Moreover, we vary the strength of the time trend $\lambda$ ($\lambda \in [-0.5; 0.5]$) and its pattern. We compare the linear regression model with period adjustment to the pooled and separate analysis approaches.

Figure 3.3 shows the impact of the strength of the time trend on the operating characteristics when evaluating the 5th experimental arm. The regression model, as well as the separate analysis asymptotically control the type I error rate, regardless of the

→ you can explain again the design as in 3.2.2, but I would move the main explanation of the design to the prev. sect

Link better these ideas with CH2.

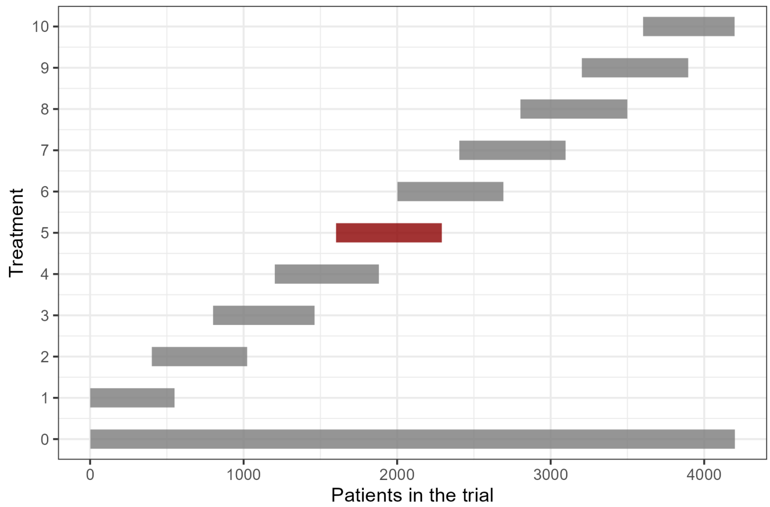


Figure 3.2.: Illustration of the scenario with 10 experimental arms entering sequentially with $d = 400$.

strength and pattern of the time trend. The pooled analysis leads to inflation of the type I error in the presence of positive time trends and deflation in case of negative time trends. Additionally, the regression model leads to gain in power as compared to the separate analysis.

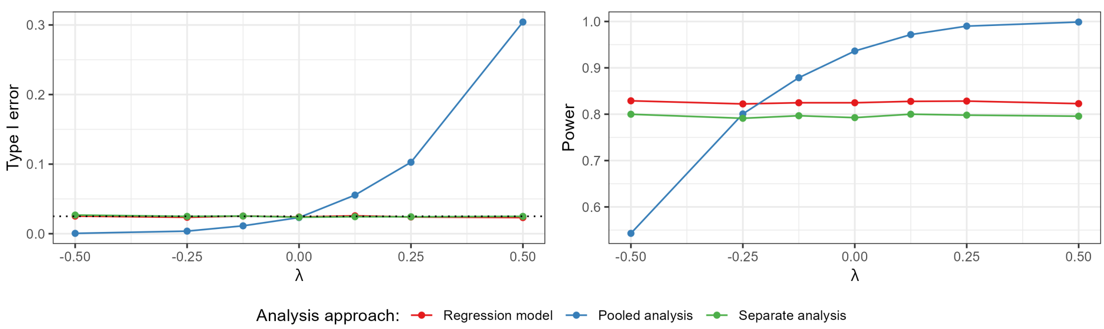


Figure 3.3.: Type I error rate and power of the regression model with period adjustment compared to the pooled and separate analyses with respect to the strength of the time trend $\lambda$. In this example, $d = 400$ and linear shape of the time trend are used and the 5th experimental arm is being evaluated.

The effect of the amount of overlapping sample size is shown in Figure 3.4. The overlaps have no effect on the type I error rate control by the regression model and separate approach, which is guaranteed (asymptotically) in all the cases. The inflation of the pooled analysis gets stronger with increasing $d$. This is because larger $d$'s result in longer platform trials and larger size of the NCC data. The power of the regression model, however, depends on the overlap between treatment arms. In the extreme cases with $d = 0$ and $d = 2n$, the regression model leads to identical power as the separate analysis. If $d = 0$, there is no NCC data, as all the arms join the trial at the beginning. Thus, the control group used for the treatment-control comparison is the same for both, the separate analysis and the regression model. In case of no overlap between the arms ($d = 2n$), there is insufficient amount of data to estimate the period effect. Hence, simultaneous presence of the experimental arms in the trial is crucial for power gains when using the regression model.

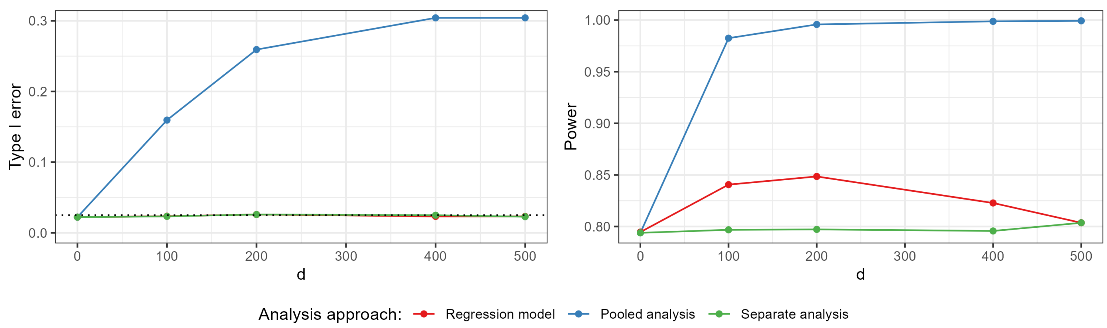


Figure 3.4.: Type I error rate and power of the regression model with period adjustment compared to the pooled and separate analyses with respect to the timing of adding the treatment arms. In this example, a linear time trend with strength $\lambda = 0.5$ is considered and the 5th experimental arm is being evaluated.

SETTING 2:

### 3.2.2. Alternative definition of the time covariate

OK!

To examine the regression model with calendar time adjustment, we consider a platform trial with 4 experimental treatment arms, where arm $i$ enters after $d_i = 250 \cdot (i - 1)$, leading to a total sample size of 1528 patients. The considered scenario is shown in Figure 3.5. The calendar time length varies from 15 to 800, while the range of considered strengths of the time trend is $[-0.5; 0.5]$.

The effect of increasing calendar time length on the type I error rate and power under linear time trend is illustrated in Figure 3.7. The type I error is maintained for moderately sized calendar time units, and slightly inflated for units larger than 600 patients. However, the type I error rate control for different time unit lengths is dependent on the time trend pattern - for instance, in case of the stepwise trend, the control is only given for very small units (<50 patients). Depending on the unit size, the model with calendar time adjustment can lead to power improvements as compared to the period adjustment. Nevertheless, as the type I error rate is not maintained in all cases, the choice of the interval length and the resulting trade-off between type I and type II errors needs to be carefully assessed.

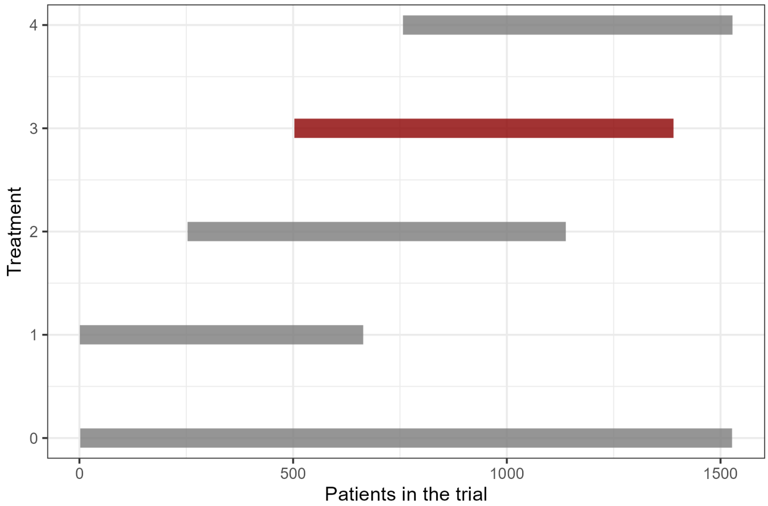


Figure 3.5.: Illustration of the scenario with 4 experimental arms entering sequentially with $d = 250$.

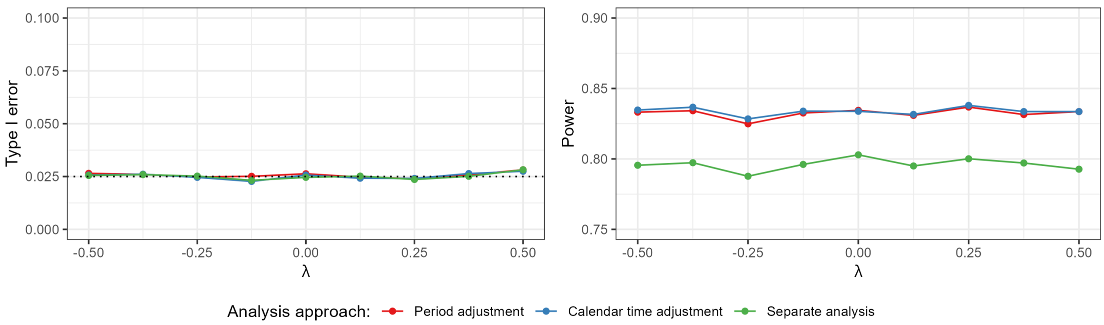


Figure 3.6.: Type I error rate and power of the regression model with calendar time adjustment compared to the regression model with period adjustment and separate analysis with respect to the strength of the time trend $\lambda$. In this example, calendar time unit length of 25 and linear shape of the time trend are used and the 3rd experimental arm is being evaluated.

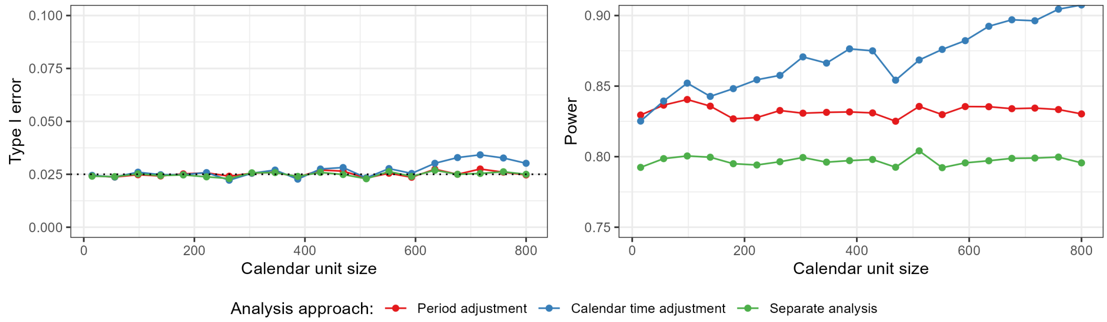


Figure 3.7.: Type I error rate and power of the regression model with calendar time adjustment compared to the regression model with period adjustment and separate analysis with respect to the size of the calendar time unit. In this example, a linear time trend with strength $\lambda = 0.15$ is considered and the 3rd experimental arm is being evaluated.

### 3.2.3. More flexible modelling approaches

The performance of the spline regression models is assessed in a scenario with 7 experimental treatment arms, where some of them enter and leave the trial simultaneously (see Figure 3.8 for illustration). In particular, entry times $d = (0, 250, 250, 500, 500, 750, 750)$ are used. The models are examined under different time trend patterns with varying strength ($\lambda \in [-0.5; 0.5]$).

The type I error rate for the cubic spline regressions under the considered scenarios is presented and compared to the fixed regression model with period adjustment in Figure 3.9. If the time trend pattern is given by a continuous function, the spline regression maintains the type I error rate. However, in case of the stepwise time trend, we observe an inflation in the type I error rate. This is because the spline regression estimates the time effect by a smooth function, which is not an optimal approach if there are sudden jumps in the time trend.

**PENDING:**

- Discuss mixed models
- Discuss degree of polynomial splines
- Discuss power of spline regression
- TBD

Good first version!

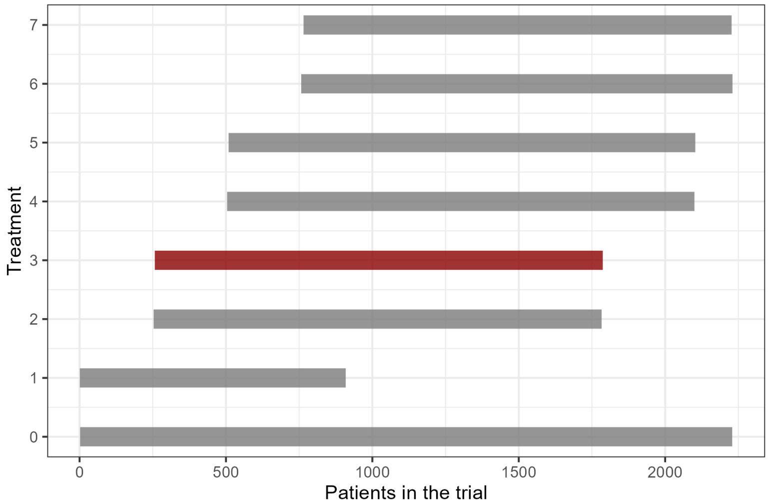


Figure 3.8.: Illustration of the scenario with 7 experimental arms entering sequentially with $d = (0, 250, 250, 500, 500, 750, 750)$.

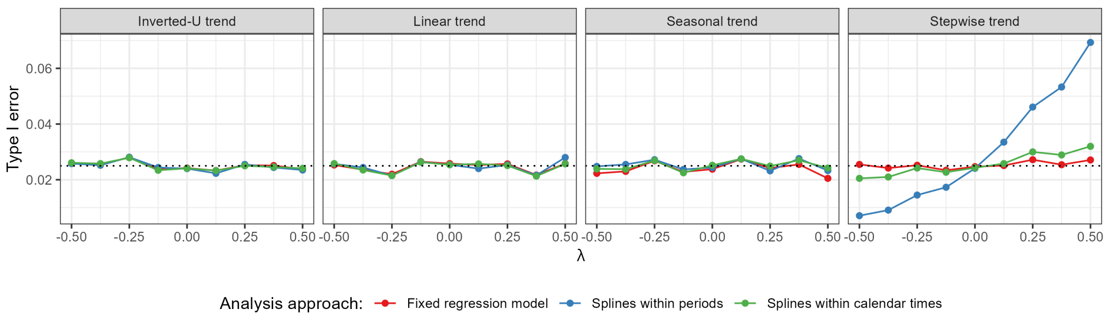


Figure 3.9.: Type I error rate for the cubic spline regression model with knots according to periods or calendar time units compared to the regression model with period adjustment with respect to the strength of the time trend $\lambda$ using different time trend patterns. In this example, the 3rd experimental arm is being evaluated.