# Improving interim decisions for single-arm trials by adjusting for baseline covariates and short-term endpoints

*Eline Anslot*

*Joint work with Kelly Van Lancker*

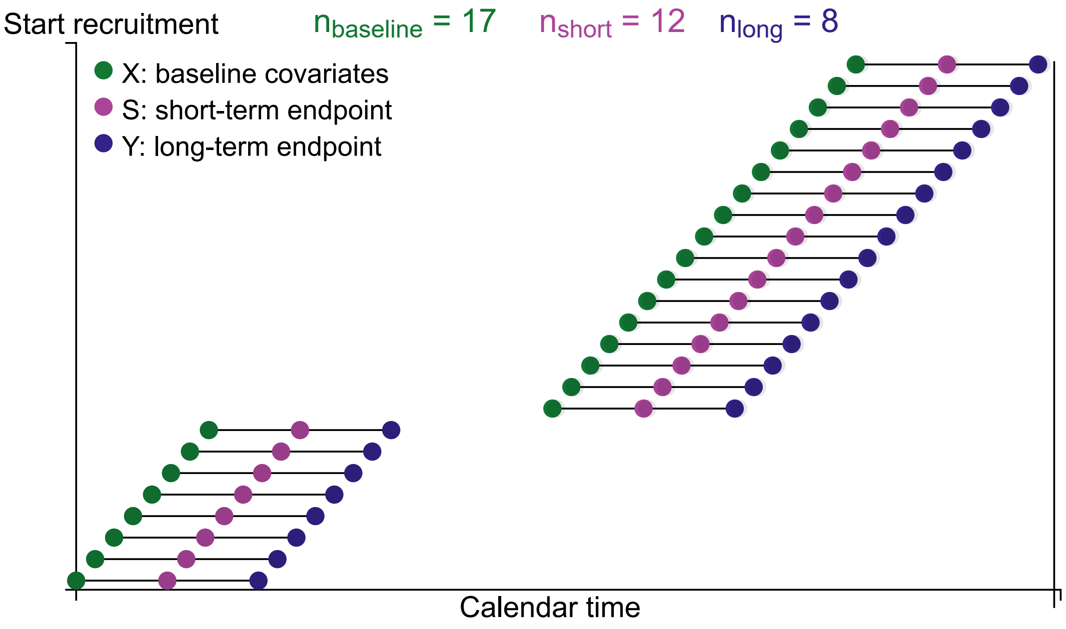
GHENT UNIVERSITY

# Table of content



GHENT UNIVERSITY

# Single-arm trials with multi-stage designs




Single-arm trials often used when:

- Few participants available
- Ethical reasons
- Effect under SoC well known, e.g., in oncology

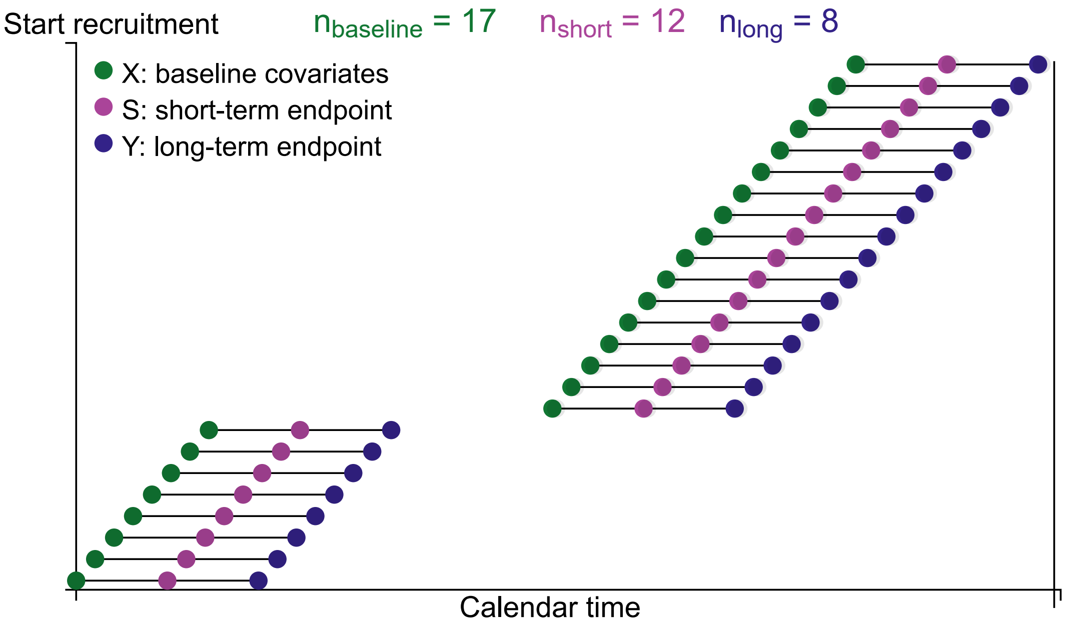
Multi-stage designs allow stopping an ongoing trial for futility and/or efficacy.

Commonly used designs:

- Group sequential designs
- Simon's two-stage designs

Continued/Paused

# Interim analysis of two-stage designs




Continued/Paused

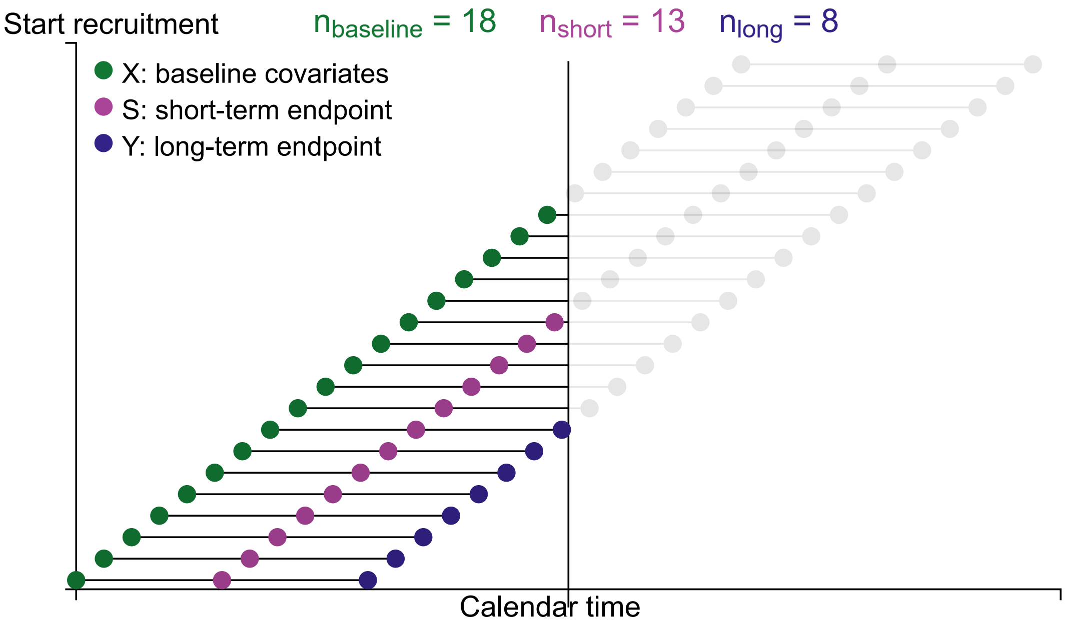
Interim analysis: based on the long-term endpoint ⇒ Unadjusted analysis

- **Simon's two-stage design:**

- **Group sequential designs:**

with compared to cut-off to stop a trial for futility or efficacy based on e.g., Pocock (1977), O'Brien and Fleming (1979) or Lan and DeMets (1983) or error spending functions

# Can we use more information?

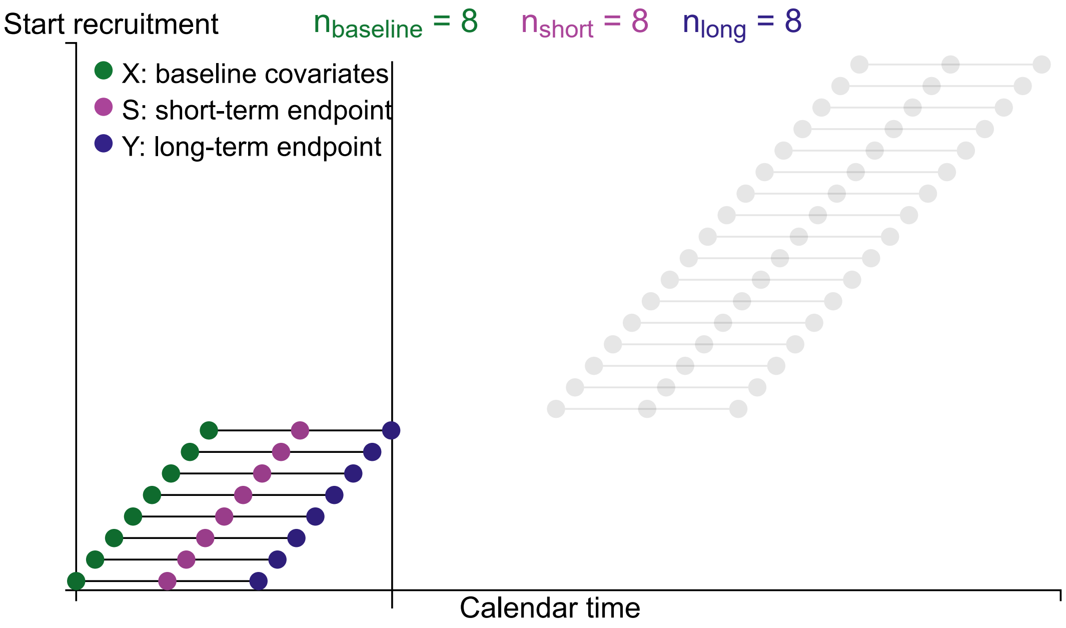



Unadjusted analysis

Reset

Adjusted analysis

# Improve decision by short-term endpoint



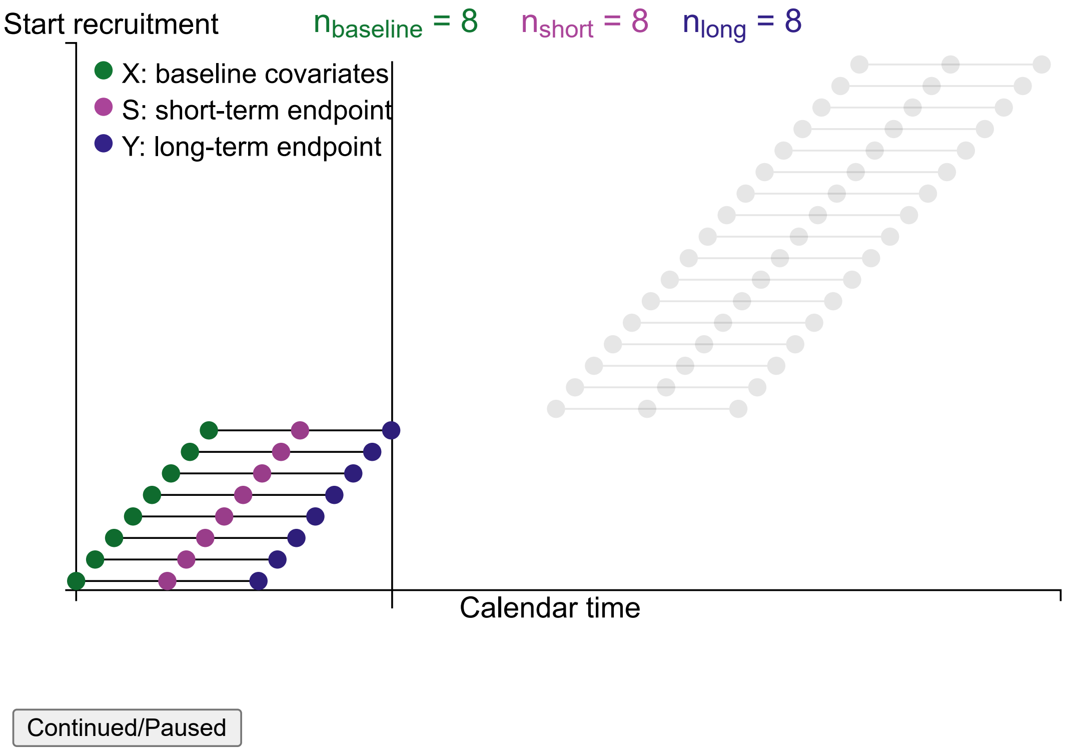

Unadjusted analysis

Kunz et al. (2017)

Zocholl et al. (2023)

Reset

# More precise interim estimator



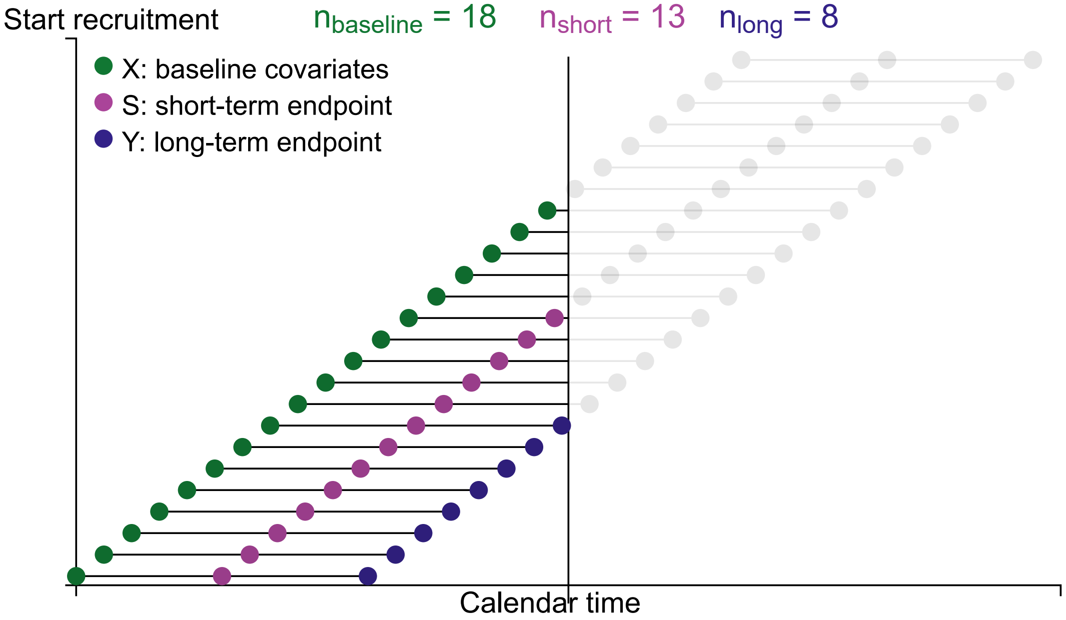

- **Recruitment with a pause:** Possible as in Kunz et al. (2017) and Zocholl et al. (2023)
- **Continuous recruitment**: ⇒ Focus of the talk

# Table of content



GHENT UNIVERSITY

# Proposed method

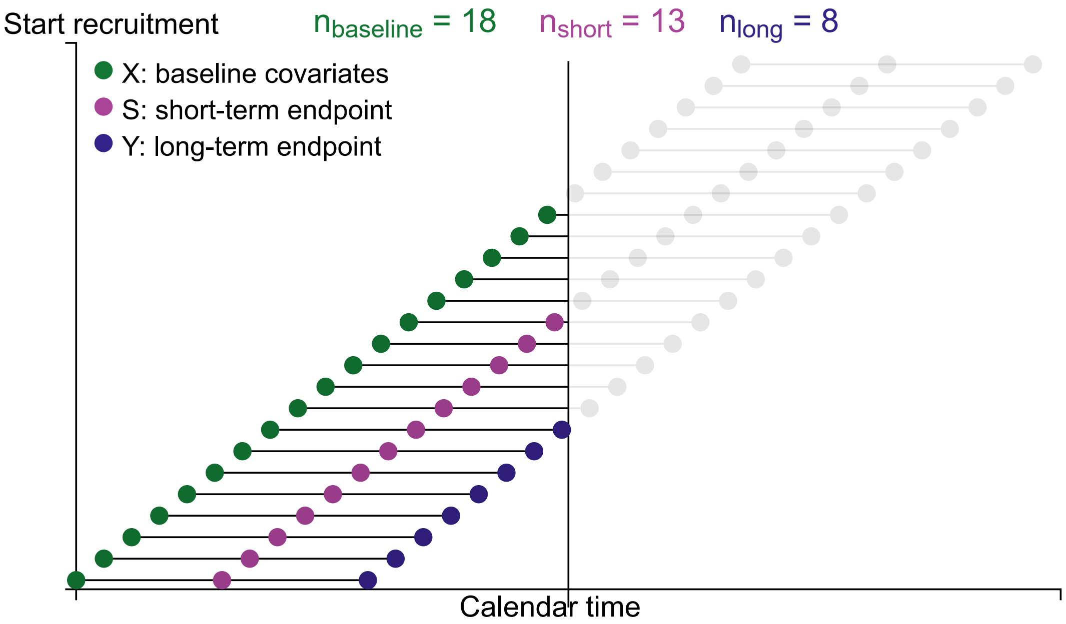



**Step 1: Model fitting in cohort 1:**

h(.): canonical link function

**Step 2: Predicting in cohort 1 and 2:**

# Proposed method




**Step 3: Model fitting in cohort 1 and 2:**

**Step 4: Predicting in cohort 1, 2 and 3:**

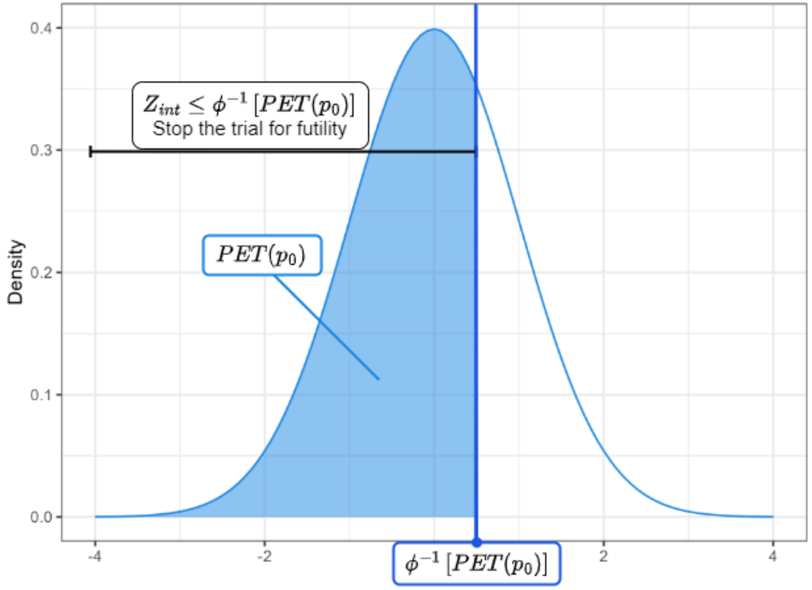
**Step 5: Averaging**

# Proposed method - Decision at interim

**Decision at interim:**

**In Simon's Two-Stage:**

with




**In Group Sequential Design:**

compared to cut-off to stop a trial for futility or efficacy based on e.g.,

Pocock (1977), O'Brien and Fleming (1979) or

Lan and DeMets (1983) or error spending functions

# Proposed method

**Decision at interim:**

- Adjusting for multiple short-term endpoints and baseline covariates
- Asymptotically unbiased even with misspecified models
    - Under random recruitment
- Asymptotically efficient when models are correct

# Table of content



GHENT UNIVERSITY

# Simulation settings

,

Design: two different optimal two-stage designs to generate e.g., and

| Design | $p_0$ | $p_A$ | $n_{final}$ | $n_{long}$ | $r_1$ | $PET(p_0)$ |
|---|---|---|---|---|---|---|
| 1 | .25 | .35 | **149** | **56** | 15 | **.6853** |
| 2 | .25 | .30 | **522** | **223** | 57 | **.6112** |

**Setting 1:**

| Design | Adjustment | | $n_{interim}$ | | Proportion $n_{short}$ | Degree of Predictivity |
|---|---|---|---|---|---|---|
| | Baseline covariate(s) | Short-term endpoint | $n_{long}$ | $n_{short}$ | | |
| 1 | / | 1 | 56 | 15 | 0.20 | Low to High |
| | | | | 25 | 0.30 | |
| | | | | 58 | 0.50 | |
| | | | | 86 | 0.60 | |
| 2 | / | 1 | 223 | 58 | 0.20 | Low to High |
| | | | | 99 | 0.30 | |
| | | | | 228 | 0.50 | |
| | | | | 299 | 0.60 | |

# Simulation settings

,

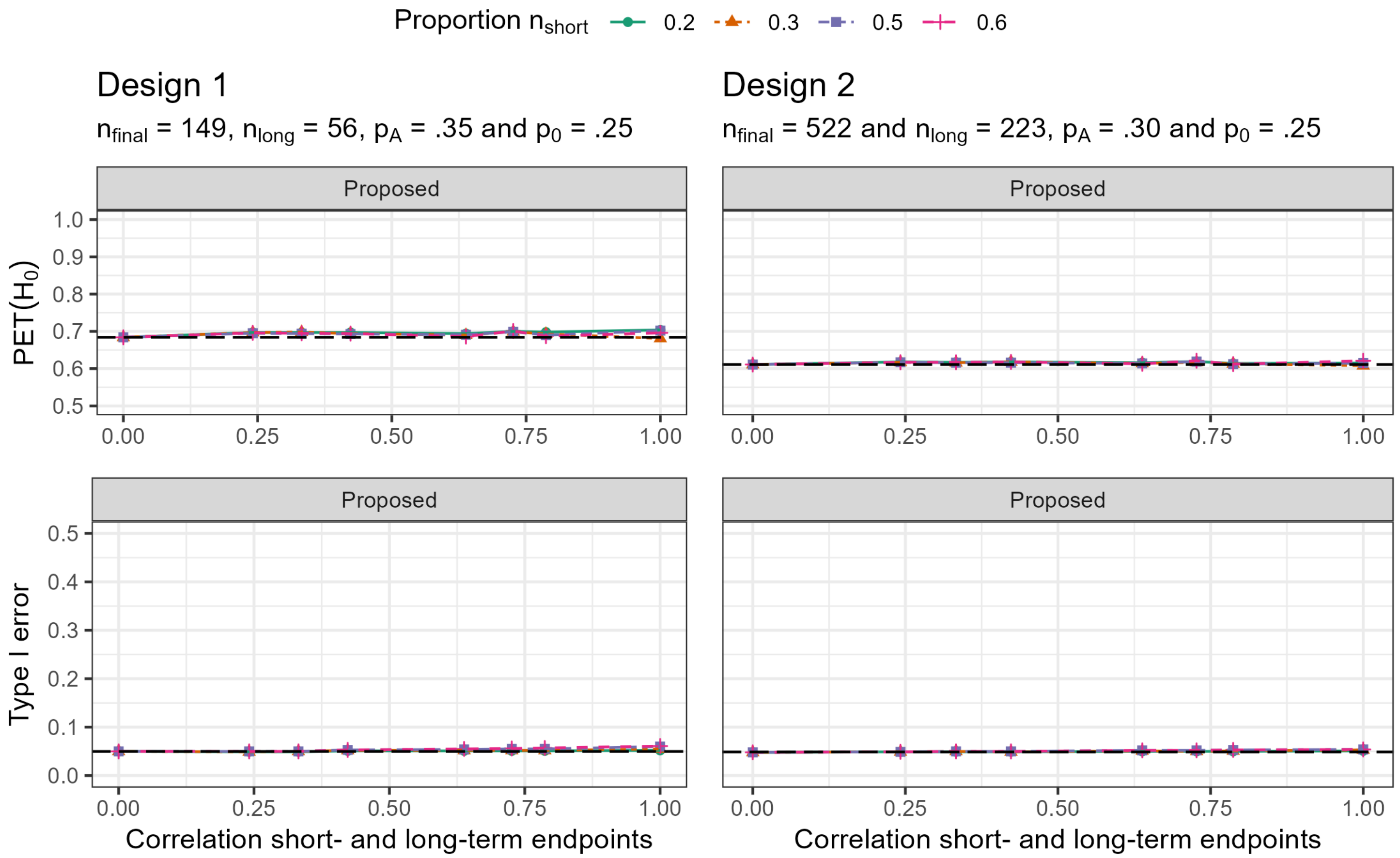
Design: two different optimal two-stage designs to generate e.g., and

| Design | $p_0$ | $p_A$ | $n_{final}$ | $n_{long}$ | $r_1$ | $PET(p_0)$ |
|---|---|---|---|---|---|---|
| 1 | .25 | .35 | **149** | **56** | 15 | **.6853** |

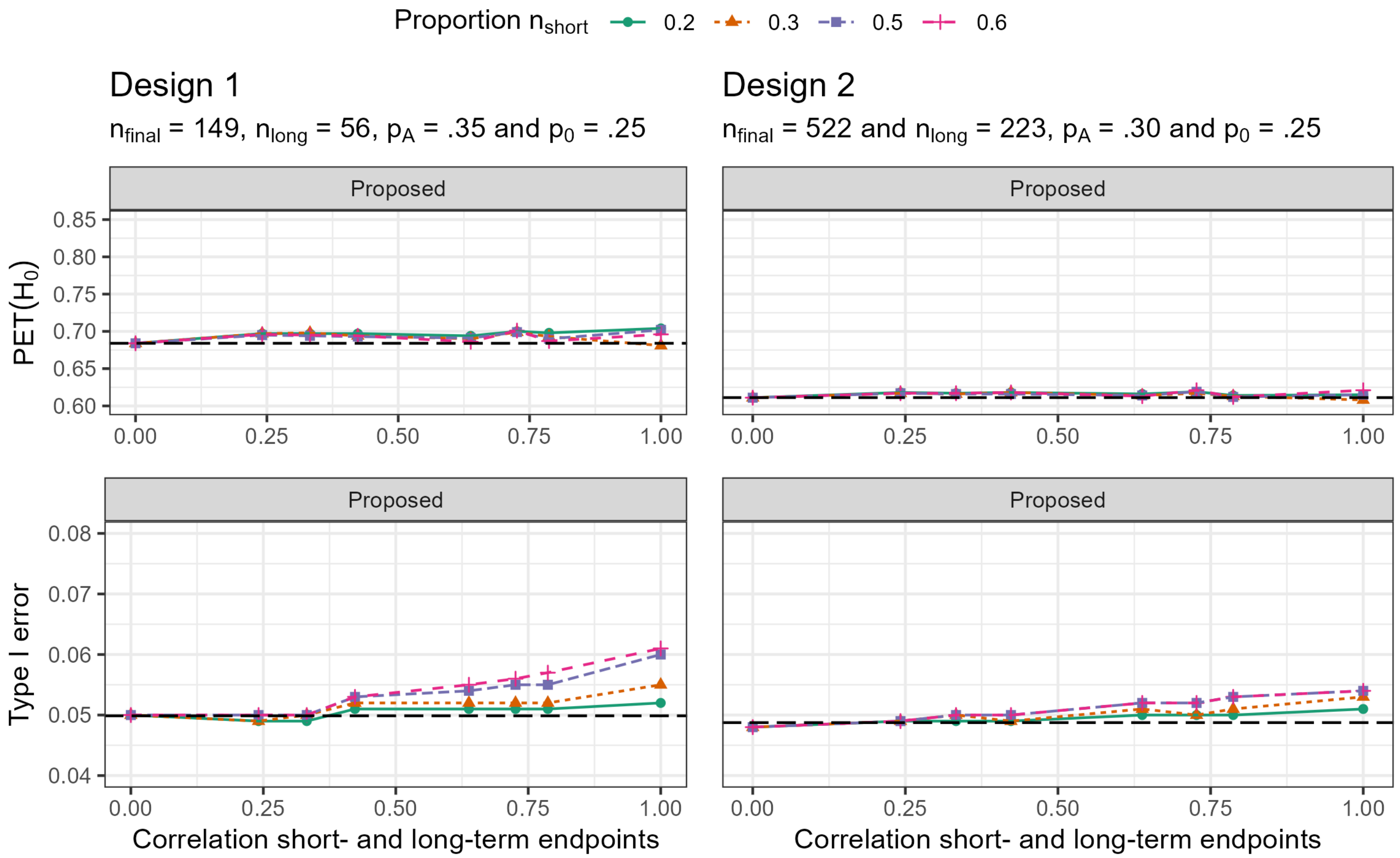
**Setting 2:**

| Design | Adjustment | | $n_{interim}$ | | | Degree of Predictivity | Models | |
|---|---|---|---|---|---|---|---|---|
| | Baseline covariate(s) | Short-term endpoint | $n_{cohort1}$ | $n_{cohort2}$ | $n_{cohort3}$ | | Correct | Misspecified |
| 1 | 3 | 1 | 56 | 29 | 29 | Low, Moderate, High | ✓ | ✓ |

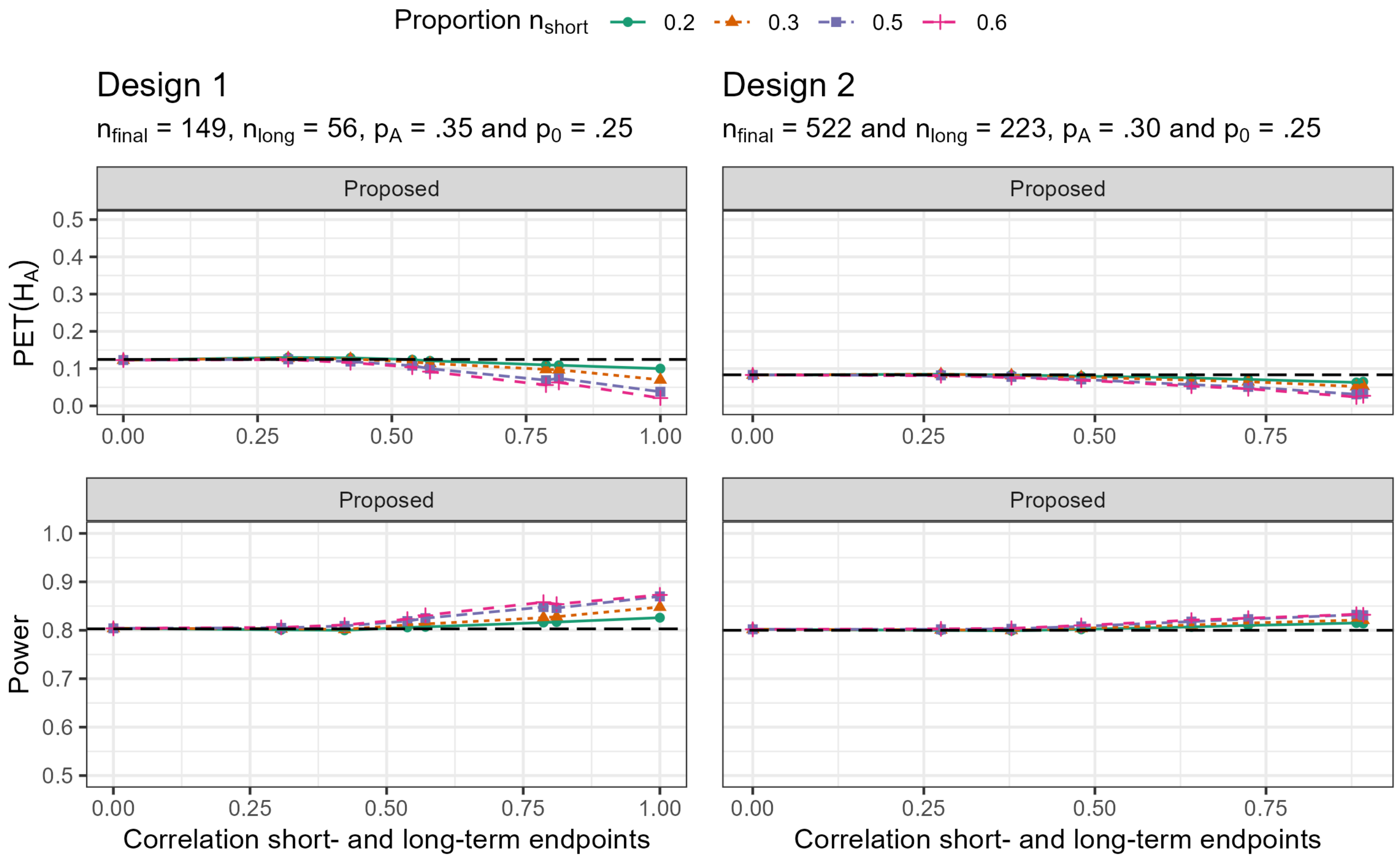
# Setting 1 - Under the null hypothesis



# Setting 1 - Under the null hypothesis



# Setting 1 - Under the alternative hypothesis



# Setting 2 - Model mispecification

| **Models** | **Degree of predictivity** | **power** |
|---|---|---|
| Correct | Not predictive | 79.7% |
| | Moderately predictive | 82.4% |
| | Highly predictive | 84.6% |
| Main | Not predictive | 79.6% |
| | Moderately predictive | 82.4% |
| | Highly predictive | 84.7% |
| $\lvert X_1 \rvert$ | Not predictive | 79.9% |
| | Moderately predictive | 80.2% |
| | Highly predictive | 80.8% |

# Table of content



GHENT UNIVERSITY

# Discussion

**Additional gain of the proposed method, depends on:**

- Proportion of additional participants in the pipeline
  - But ideally not everybody should be recruited at interim
- Predictivity of baseline covariates and short-term endpoint
- Model misspecification
  - Extension: data-adaptive methods to help build the models (see e.g., Van Lancker et al., 2024)

**Calculate sample size as if no power gain occured**

## Thank you for your attention

## Questions?

GHENT UNIVERSITY