# New inference methods for adaptive Phase II designs with a binary endpoint

**Arsénio Nhacolo** & **Werner Brannath**
Competence Centre for Clinical Trials
University of Bremen

**Contact Information:**
KKSB
Linzer Str. 4, 28359 Bremen
Germany

Email:
anhacolo@uni-bremen.de

**Abstract**

We propose point and interval estimation for adaptive designs. We consider the recently proposed oncology Phase II two-stage single-arm adaptive designs with binary endpoint, in which the second stage sample size is a pre-defined function of the first stage's number of responses. Our approach is based on sample space orderings, from which we derive p-values, and point and interval estimates. Simulation studies showed that our proposed methods perform better, in terms of bias and root mean square error, than the naïve (fixed-sample) maximum likelihood estimator.

## Introduction

Adaptive group-sequential designs offer the possibility of making mid-course data-driven changes without jeopardizing the integrity of the clinical trials. Due to this flexibility and gains in sample size as compared to the fixed-sample designs, adaptive designs are becoming popular. However, new proposals of these designs are mainly concerned with hypothesis testing and often come without the respective methods for the efficacy parameter estimation. We propose estimation methods for a class of adaptive single-arm group-sequential designs with binary endpoint, in which the sample size of the second stage is a pre-defined function of the number of responses in the first stage[2, 5]. These designs are intended to Phase II oncology trials.

## Main Objectives

1. Propose alternative sample space orderings;
2. Derive over-all p-value;
3. Propose interval and point estimators.

## Materials and Methods

### Design

We consider adaptive single-arm two-stage design with a binary endpoint [2]:

- Represented by $\{n_1, l_1, u_1, n_2(x_1), D(x_1), l(x_1)\}$, where $x_1$, $n_1$, $l_1$ and $u_1$ are stage 1 number of responses, sample size, and futility and efficacy boundaries, and $n_2(x_1)$, $D(x_1)$ and $l(x_1)$ are stage 2 sample size, conditional error and decision boundary.
- Hypotheses: $H_0 : \pi \leq \pi_0$ vs $H_1 : \pi \geq \pi_1$
- Trial stops at stage 1 with failure to reject $H_0$ if $x_1 \leq l_1$ or with rejection of $H_0$ if $x_1 \geq u_1$. Otherwise it proceeds to stage 2, at which $H_0$ is rejected if $p_2 \leq D(x_1)$ or, equivalently, $x > l(x_1)$, where $p_2$ is the second stage $p$-value and $x$ is the total number of responses.

### Estimation methods

Our proposed estimators [4] for the design above are as follows. Denote the trial outcome by $(m, x_1, x)$, where $m$ is the stopping stage. Based on stage-wise ordering, we defined a sample space ordering that take into account the design's adaptation rule, as follows. A trial outcome $(m', x_1', x')$ is at least as extreme (against $H_0$) as the observed trial outcome $(m, x_1, x)$ if one of the following conditions is met:

(A1) $m' = m = 1$ and $x' \geq x$
(A2) $m' = m = 2$ and $\delta(x_1', x_2') \geq \delta(x_1, x_2)$
(B) $m' = 1, m = 2$ and $x' \geq u_1$
(C) $m' = 2, m = 1$ and $x \leq l_1$

with $\delta$ defined as

- Method 1: $\delta(x_1, x_2) = x - l(x_1)$
- Method 2: $\delta(x_1, x_2) = D(x_1) - p_2(x_2)$
- Method 3: $\delta(x_1, x_2) = 1 - C(p_{1b}, p_2)$

where $C$ is the weighted inverse normal combination function represented as $C(p_1, p_2) = 1 - \Phi\left[w_1\Phi^{-1}(1-p_1) + w_2\Phi^{-1}(1-p_2)\right]$, with

$$w_1 = \sqrt{\frac{n_1}{n_1 + n_2(x_1)}}, w_2 = \sqrt{\frac{n_2(x_1)}{n_1 + n_2(x_1)}}$$

and

$$p_{1b}(x_1) = 1 - \Phi\left\{\frac{\Phi^{-1}(1-c) - w_2\Phi^{-1}[1 - D(x_1)]}{w_1}\right\}.$$

Based on these sample space ordering, we derived the overall $p$-value $Q$:

$$Q = \begin{cases} 1 - B(x_1 - 1, n_1, \tilde{\pi}_0) & \text{if } m = 1 \\ 1 - B(u_1 - 1, n_1, \tilde{\pi}_0) + & \\ \sum_{X_1 = l_1+1}^{u_1 - 1} b(X_1, n_1, \tilde{\pi}_0) \Pr_{\tilde{\pi}_0}[\delta(X_1, X_2) \geq \delta(x_1, x_2)] & \text{if } m = 2 \end{cases}$$

where $B(x, n, \pi)$ and $b(x, n, \pi)$ are the binomial cumulative distribution function and probability mass function.

Exploiting the duality between confidence intervals (CI) and hypothesis tests, we construct the CI by considering all the null hypotheses

$$H_0^{\tilde{\pi}_0} : \pi \leq \tilde{\pi}_0, \text{ with } 0 \leq \tilde{\pi}_0 \leq 1$$

The one-sided $(1-\alpha)100\%$ CI for $\pi$ is $\left[\pi_L^\alpha; 1\right]$ where

$$\pi_L^\alpha = \{\tilde{\pi}_0 : \Pr_{\tilde{\pi}_0}((M, X_1, X) \succeq (m_o, x_{1o}, x_o)) = Q(\tilde{\pi}_0) = \alpha\}$$

The point estimate (median) is

$$\hat{\pi} = \pi_L^{0.5}$$

Constraint: $Q$ must be increasing in $\tilde{\pi}_0$

### Simulation study

- Objective: Compare performance of the proposed estimates with the naïve maximum likelihood estimate (MLE), in terms of bias and root mean square error (RMSE).
- Design 1: $(\pi_0, \pi_1, \alpha, \beta, n_1) = (0.2, 0.4, 0.05, 0.1, 20)$.
- Design 2 : $(\pi_0, \pi_1, \alpha, \beta, n_1) = (0.4, 0.6, 0.05, 0.1, 22)$.
- True $\pi$: from 0 to 1 by increments of 0.01.
- Simulation runs: 50 000.

## Results

The results for bias is shown in Figures 1 and 2, and for RMSE in Figures 3 and 4. The vertical line represents $\pi = \pi_1$. We used two versions of the naïve MLE, one that uses all the trial data, $\hat{\pi}_p = [x_1 + x_2]/[n_1 + n_2(x_1)]$, and the other that uses the first stage data only, $\hat{\pi}_{p1} = x_1/n_1$. The reason for including $\hat{\pi}_{p1}$ is that since it is unbiased, it will serve as benchmark for comparison with respect to RMSE, i.e., a new estimator would not be desirable if it would be outperformed by $\hat{\pi}_{p1}$ in terms of RMSE. We denote the estimated response probability by $\hat{\pi}_{m1}$ for Method 1, $\hat{\pi}_{m2}$ for Method 2, $\hat{\pi}_{m2v2}$ for Method 2v2, and $\hat{\pi}_{m3}$ for Method 3. Method 2v2 is the same as Method 2 but with $p$-value calculated using approximations.

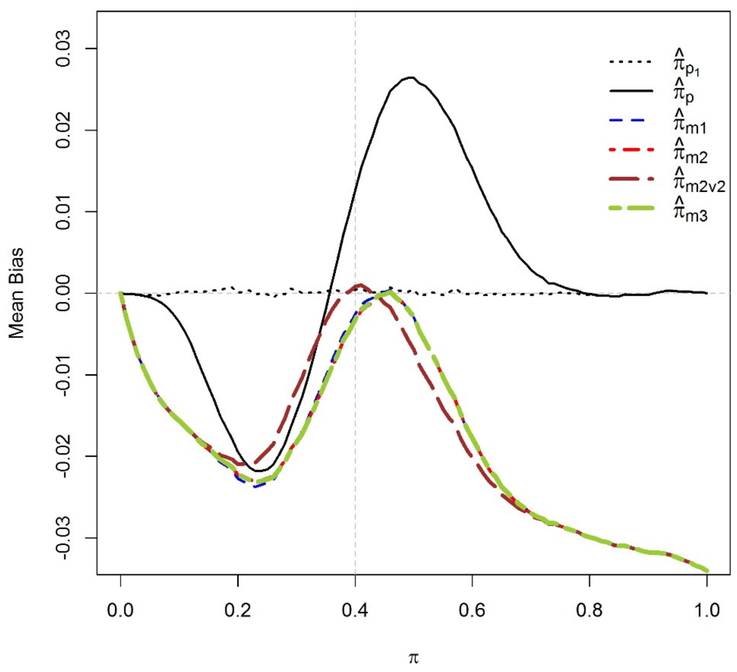


**Figure 1:** Mean bias of the estimators for the Design 1

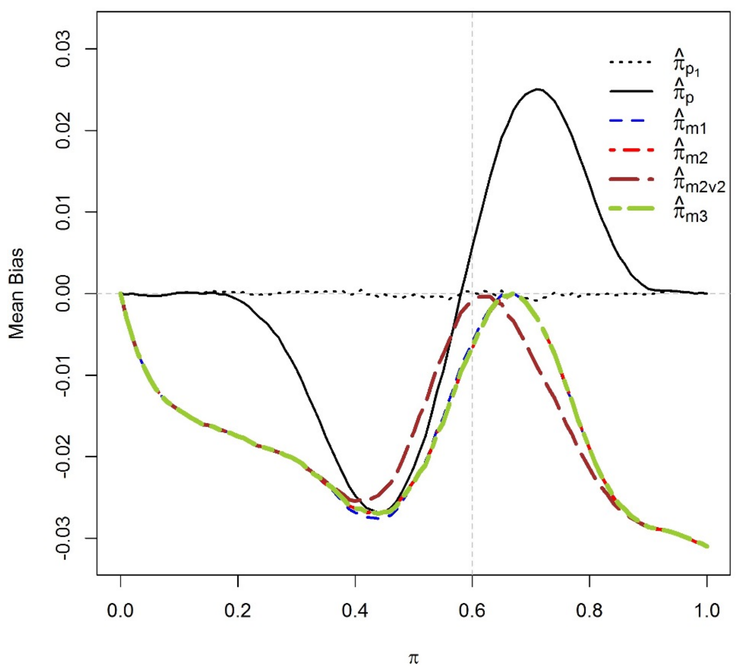


**Figure 2:** Mean bias of the estimators for the Design 2.

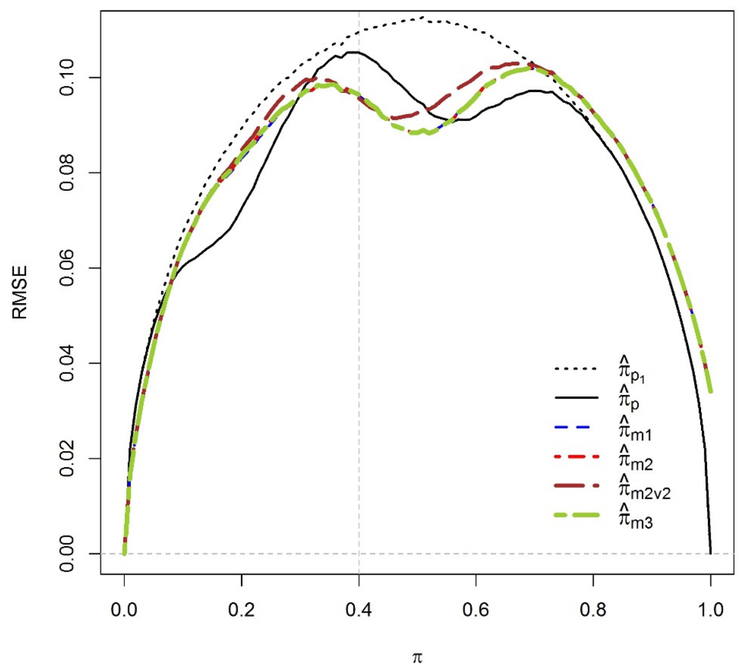


**Figure 3:** RMSE of the estimators for the Design 1.

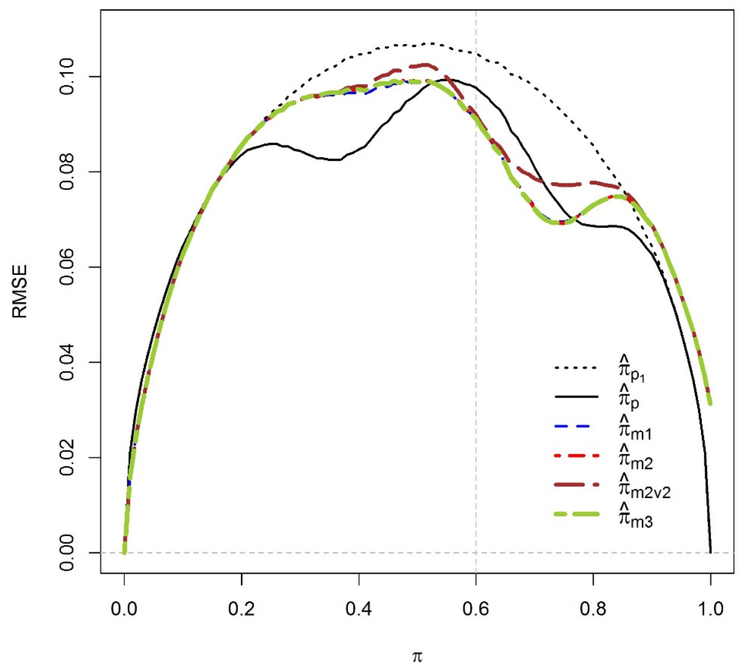


**Figure 4:** RMSE of the estimators for the Design 2.

## Conclusions

- The proposed estimators outperform the naïve MLE when the true response rate is in the vicinity of the response rate under $H_1$
- The proposed methods, unlike he naïve MLE, don't overestimate the true response rate

## References

[1] Chul Ahn, Moonseoung Heo, and Song Zhang. *Sample Size Calculations for Clustered and Longitudinal Outcomes in Clinical Research*. CRC Press, 2014.

[2] Stefan Englert and Meinhard Kieser. Optimal adaptive two-stage designs for phase II cancer clinical trials. *Biometrical Journal*, 55(6):955–968, 2013.

[3] Susan Halabi. Statistical considerations for the design and analysis of phase III clinical trials in prostate cancer. *Urologic Oncology: Seminars and Original Investigations*, 26(3):300 – 307, 2008. A Clinician's Guide to Statistical Methods in Urologic Oncology.

[4] Arsénio Nhacolo and Werner Brannath. Interval and point estimation in adaptive Phase II trials with binary endpoint. *Statistical Methods in Medical Research*, page 096228021878141, jun 2018.

[5] Guogen Shan, Gregory E. Wilding, Alan D. Hutson, and Shawn Gerstenberger. Optimal adaptive two-stage designs for early phase II clinical trials. *Statistics in Medicine*, 35(8):1257–1266, 2016. sim.6794.

## Acknowledgements

This project has received funding from the European Unions Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 633567.