AN INTRODUCTION TO ADAPTIVE DESIGNS

Group sequential designs and early stopping in clinical trials

Julien SAUSER, Service prévention et contrôle de l'infection, HUG

Colloque du CRC – 27 mai 2019

Outline

- Definitions
- Operational considerations
- Group sequential designs
- Efficacy stopping boundaries
- Futility stopping boundaries
- Alpha-spending approach

Definition(s)

European Medicines Agency

London, 18 October 2007
Doc. Ref. CHMP/EWP/2459/02

COMMITTEE FOR MEDICINAL PRODUCTS FOR HUMAN USE (CHMP)

REFLECTION PAPER ON METHODOLOGICAL ISSUES IN CONFIRMATORY CLINICAL TRIALS PLANNED WITH AN ADAPTIVE DESIGN

DRAFT AGREED BY THE EFFICACY WORKING PARTY	11 January 2006
ADOPTION BY CHMP FOR RELEASE FOR CONSULTATION	23 March 2006
END OF CONSULTATION (DEADLINE FOR COMMENTS)	30 September 2006
AGREED BY THE EFFICACY WORKING PARTY	September 2007
ADOPTION BY CHMP	18 October 2007

KEYWORDS	Adaptive Design, Interim Analyses; Design Modifications; Randomised Clinical Trials; Confirmatory Clinical Trials; Biostatistics

Adaptive Designs for Clinical Trials of Drugs and Biologics

Guidance for Industry

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

Comments and suggestions regarding this draft document should be submitted within 60 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit electronic comments to https://www.regulations.gov. Submit written comments to the Dockets Management Staff (HFA-305), Food and Drug Administration, 5630 Fishers Lane, Rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions regarding this draft document, contact (CDER) Scott N. Goldie at 301-796-2055, or (CBER) Office of Communication, Outreach and Development, 800-835-4709 or 240-402-8010.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

September 2018
Clinical/Medical

Definition(s)

- A study design is called "*adaptive*" if statistical methodology allows the modification of a design element (e.g. sample-size, randomisation ratio, number of treatment arms) at an **interim analysis** with full **control of the type I error**. (EMA, 2007)

- An *adaptive design* is defined as a clinical trial design that allows for **prospectively planned** modifications to one or more aspects of the design based on **accumulating data** from subjects in the trial. (FDA, 2018)

Why adaptive designs?

- **Ethical** : Ensure subjects are not exposed to unsafe, inferior or innefective treatments

- **Administrative** : Ensure the trial is conducted as planned (correct population, eligibility criteria)

- **Economical** : Exploit a new treatment earlier if a positive result is observed. Save resources if negative result is observed.

Interim Analysis

- Any analysis intended to compare treatment arms with respect to efficacy or safety at any time prior to the formal completion of a trial (ICH E9, 1998)

- Any examination of data obtained from subjects in a trial while that trial is ongoing, and is not restricted to cases in which there are formal between-group comparisons (FDA, 2018)

Unblinding of interim results

- Usually requires **unblinding** of the data

→ Knowledge of interim results (based on comparisons of treatment groups) can change the behaviour of the investigators and subjects

- Trial integrity must be maintained
- Sponsors, investigators, subjects, people involved in the trial conduct must remain unaware of the product codes
- Confidentiality of the interim results must be ensured

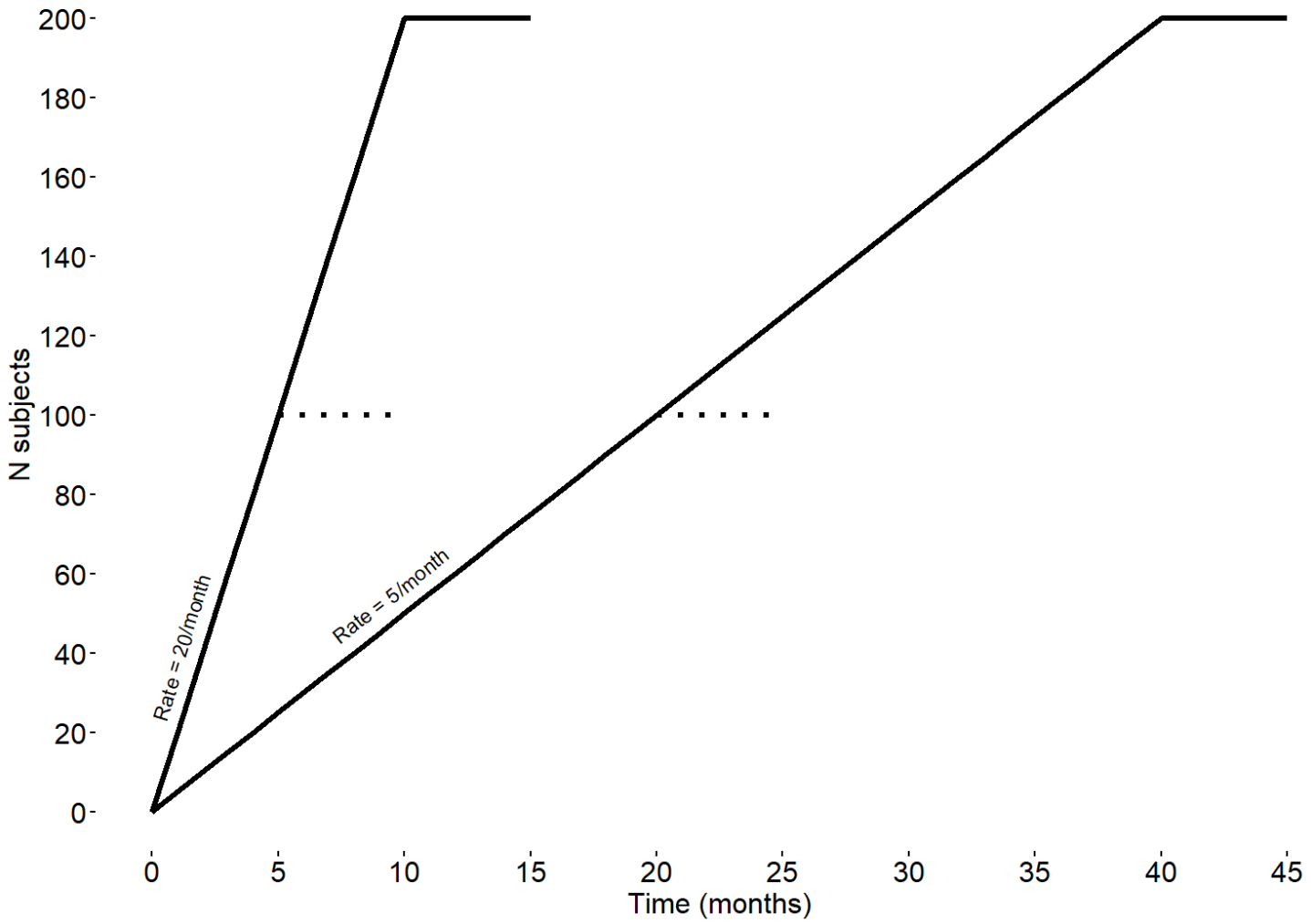
Data monitoring committee

- An independent Data Monitoring Committee (DMC) should be constitued.
  - Indepdendent group of experts mandated to review the interim results
  - At least two (independent) clinical experts and one (independent) statistician are members
  - DMC charter and interim SAP must be written
  - Provide only recommendations based on pre-planned decision rules
  - At completion of the trial only, interim results can be disclosed.
  - Can request additional analyses

Recruitment

- Sample size = 200
- Follow-up period = 5m
- Recruit. rate = 5 and 20 / m
- Interim analysis at N = 100

→ **Overrunning**



Examples of adaptive designs

a) adaptive randomization design,
b) group sequential design,
c) sample size re-estimation design,
d) drop-the-loser design,
e) adaptive dose finding (e.g., dose escalation) design,
f) biomarker-adaptive design,
g) adaptive treatment-switching design,
h) hypothesis-adaptive design,
i) adaptive seamless phase II/III trial design, and
j) multiple adaptive design.

Group sequential designs

- So-called “group sequential” designs have been developed that **avoid inflating the pre-specified type I error** associated with the repeated testing of the treatment effect based on accumulating data.(EMEA, 2007)

- Group sequential trials allow for one or more **prospectively planned interim analyses** of comparative data with **prespecified criteria for stopping the trial** (FDA, 2018)

Group sequential designs

- **What**: Look at accumulating data during the course of trial
- **Why**: Take decisions to continue or stop the trial
- **How**: K-1 interim analyses planned and one final analysis.

Maximum sample size

With K analyses planned (K-1 interim and a final analysis), and n_i observations per stage, the maximum sample size N is

$$N_{max} = n_1 + n_2 + \cdots + n_K$$

With equally sized stages,

$$N_{max} = Kn$$

Fixed design studies

- In the fixed design case the sample size is

$$N_{fix} = N_{fix}(\alpha, \beta, \Delta, \sigma)$$

- When the study is completed, compute the test statistic T and reject H_0, if (two-sided case)

$$|T| \geq z_{1-\alpha/2}$$

Inflation factor

- Monitoring data during intetrim analyses comes at cost
- The **maximum sample size gets inflated**
- It can be shown that

$$N_{max} = N_{fix} \times IF(K, \alpha, \beta)$$

- N_{fix} = sample size in fixed case
- IF = Inflation factor

Average sample size

- The maximum sample size is inflated, but the average sample size ASN is reduced

$$ASN_{H_1} = n_1 + n_2 P_{H_1}(|Z_1| < c_1) + n_3 P_{H_1}(|Z_1| < c_1, |Z_2| < c_2) + \ \ldots$$

- Group sequential designs allow to stop the trial early. → The average stopping time is decreased

Efficacy stopping boundaries

- $c_1, c_2, \ldots, c_K$ are a set of critical values corresponding to stopping boundaries.
- The trial is stopped at first interim i were $|Z_i| \geq c_i$
- To preserve the familywise error rate (FWER), i.e the probability to reject at least one true null hypothesis, c_i's must satisfy:

$$P_{H_0}(|Z_1| \geq c_1 \; or \; |Z_2| \geq c_2 \; or \; \ldots or \; |Z_K| \geq c_K) = \alpha$$

- There are different solutions.

Efficacy stopping boundaries

- Pocock (1977) proposed constant boundaries $c_i = c_P(K, \alpha)$

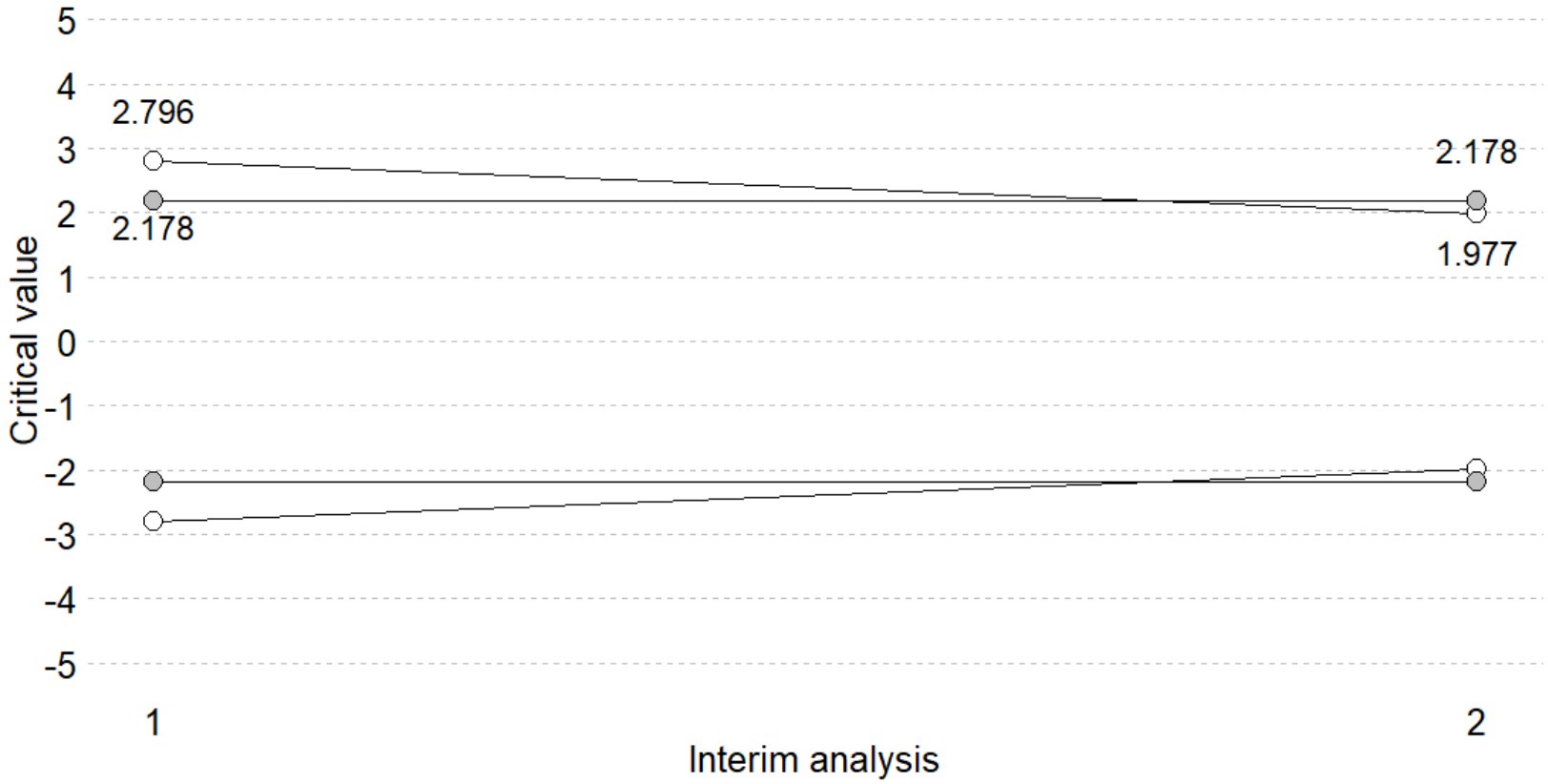
$$P_{H_0}(|Z_1| \geq c_P \; or \; |Z_2| \geq c_P \; or \; \ldots or \; |Z_K| \geq c_P) = \alpha$$

- O'Brien and Fleming (1979) proposed decreasing boundaries $c_i = \frac{c_{OBF}(K,\alpha)}{\sqrt{i}}$

$$P_{H_0}\left(|Z_1| \geq c_{OBF} \; or \; |Z_2| \geq \frac{c_{OBF}}{\sqrt{2}} \; or \; \ldots or \; |Z_K| \geq \frac{c_{OBF}}{\sqrt{K}}\right) = \alpha$$

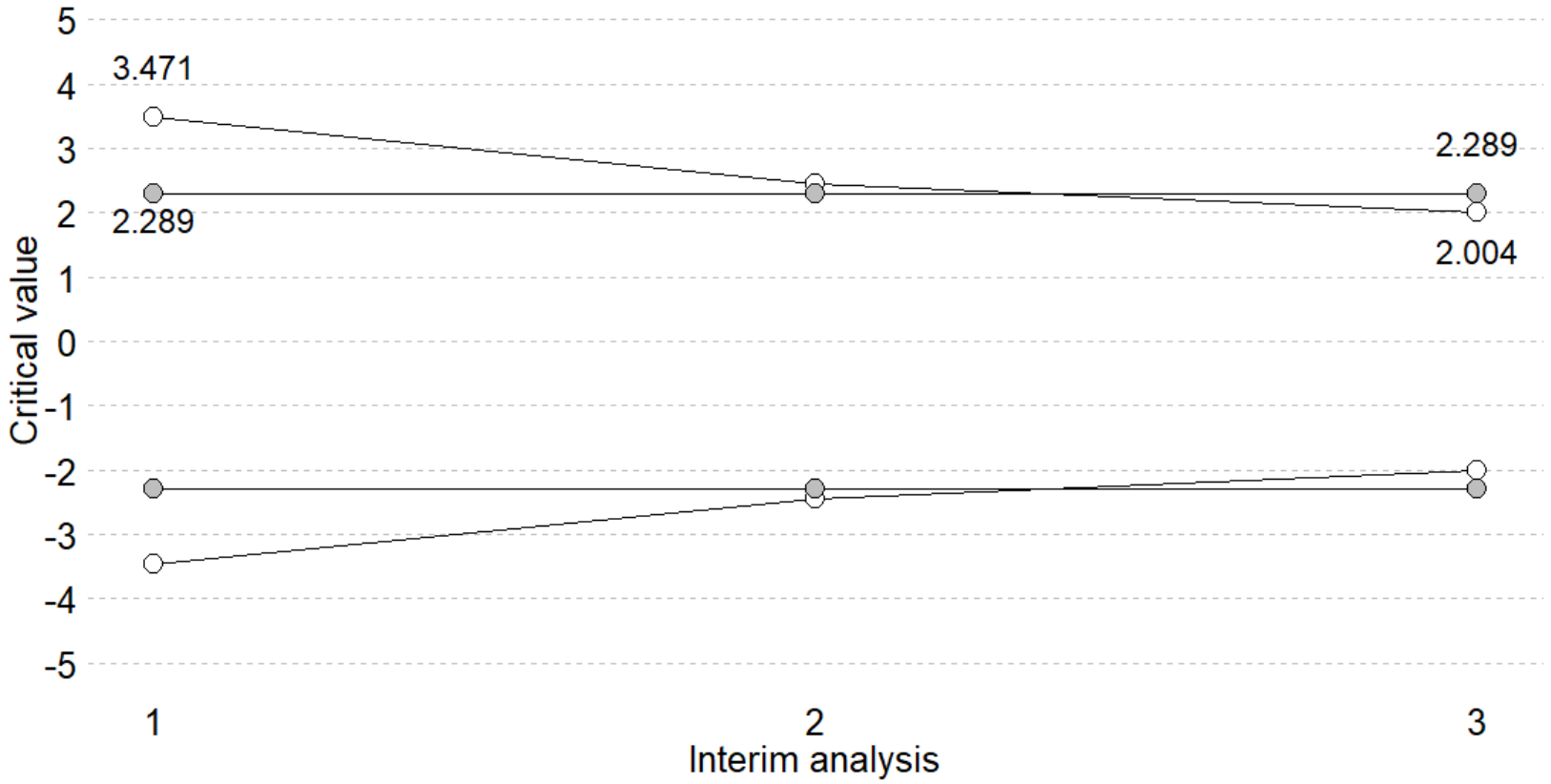
- Both methods require equally sized stages

Stopping boundaries



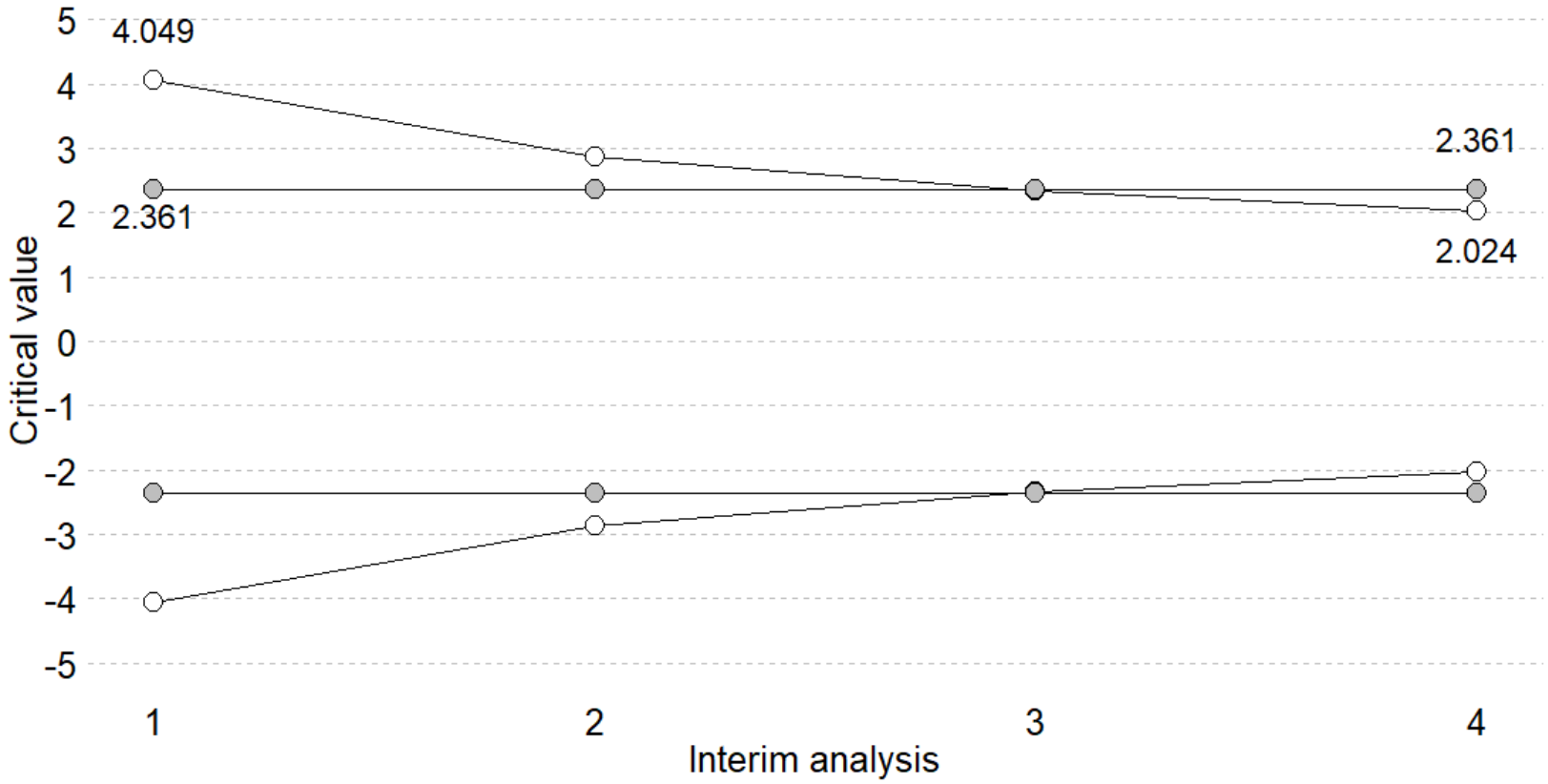
Z-value	P-value
1.977	0.0480
2.178	0.0294
2.796	0.0052

Stopping boundaries



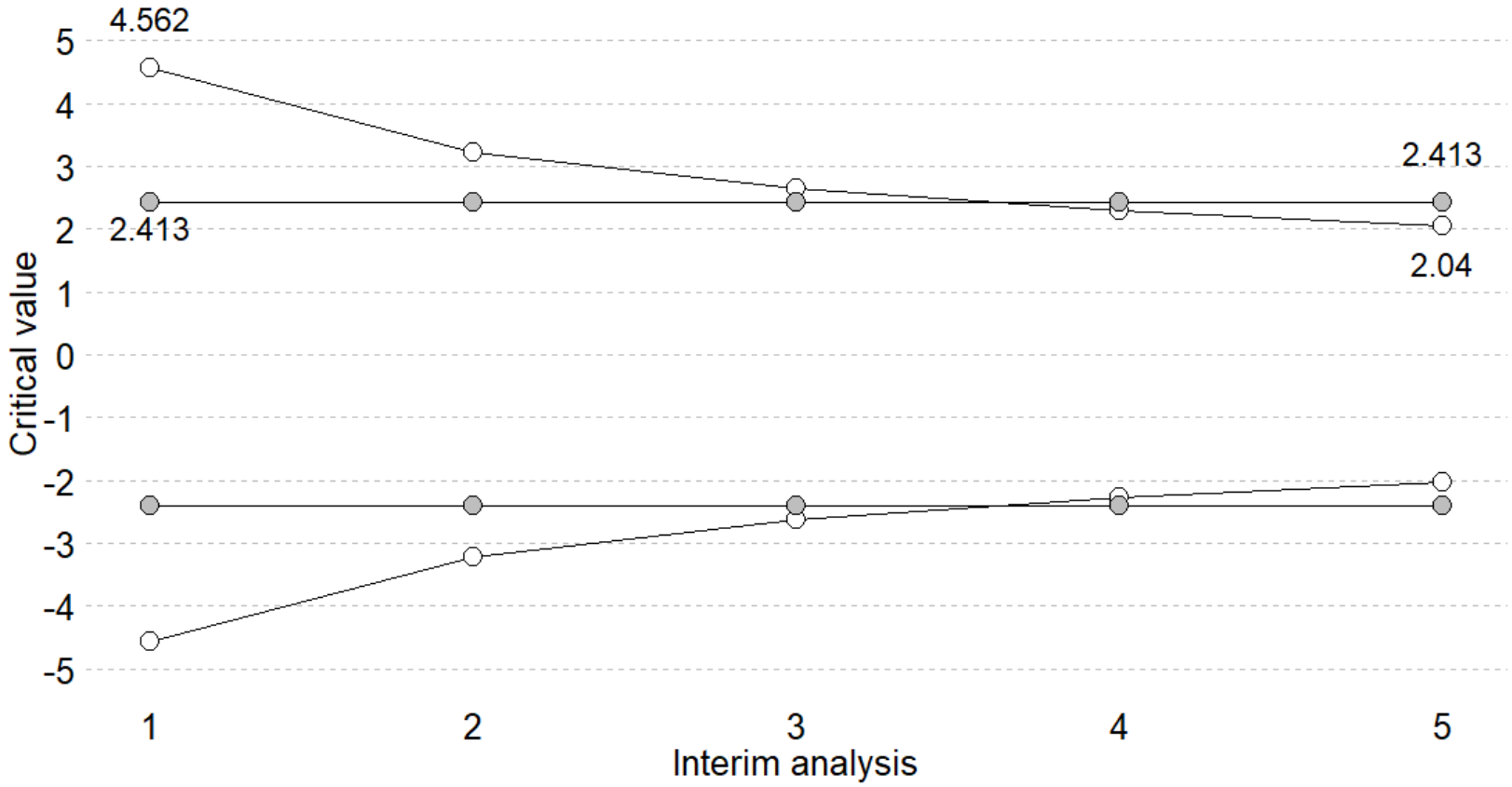
Z-value	P-value
2.004	0.0451
2.289	0.0221
3.471	0.0005

Stopping boundaries



Z-value	P-value
2.024	0.0429
2.361	0.0182
4.049	0.00005

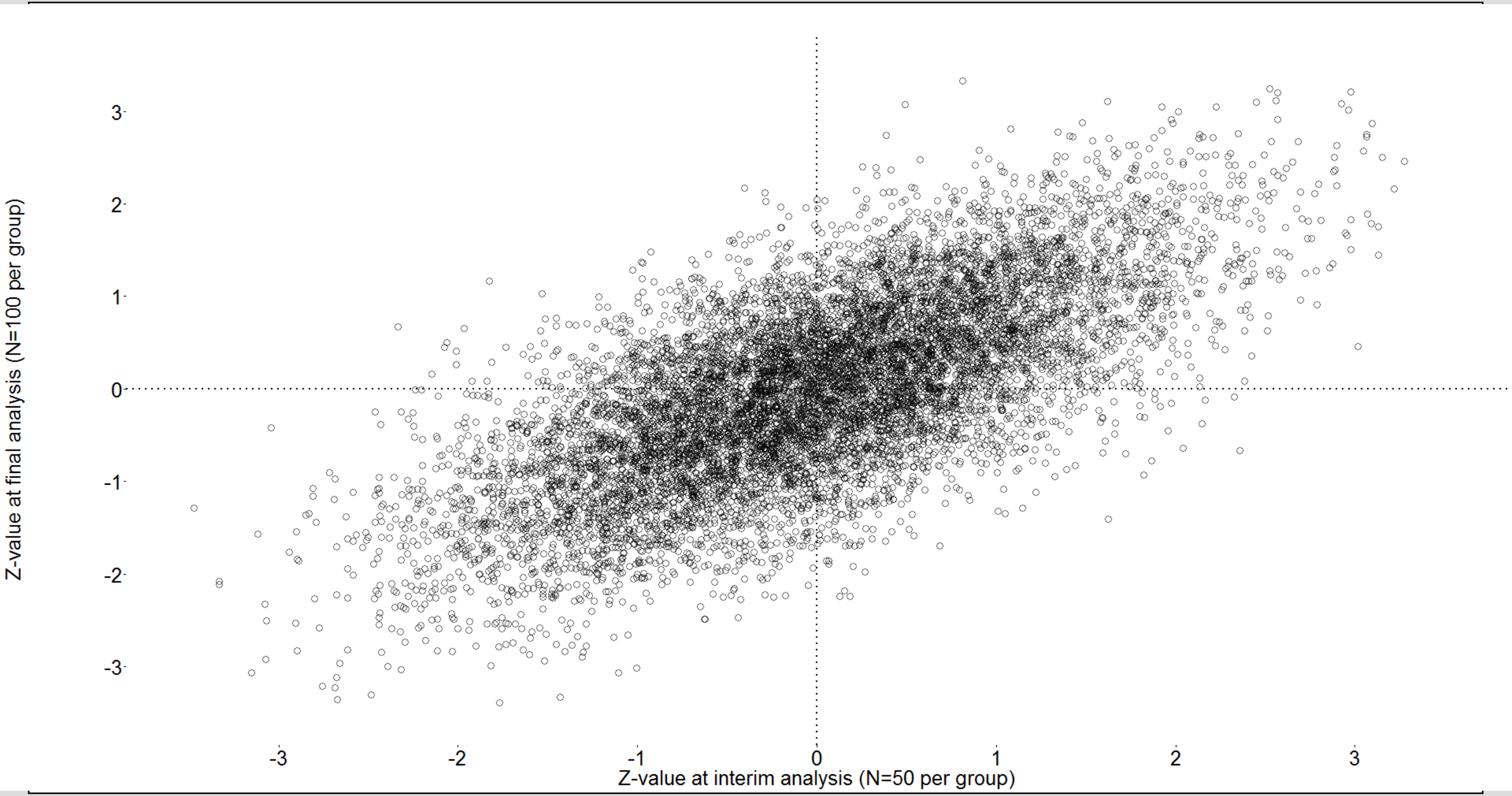
Stopping boundaries

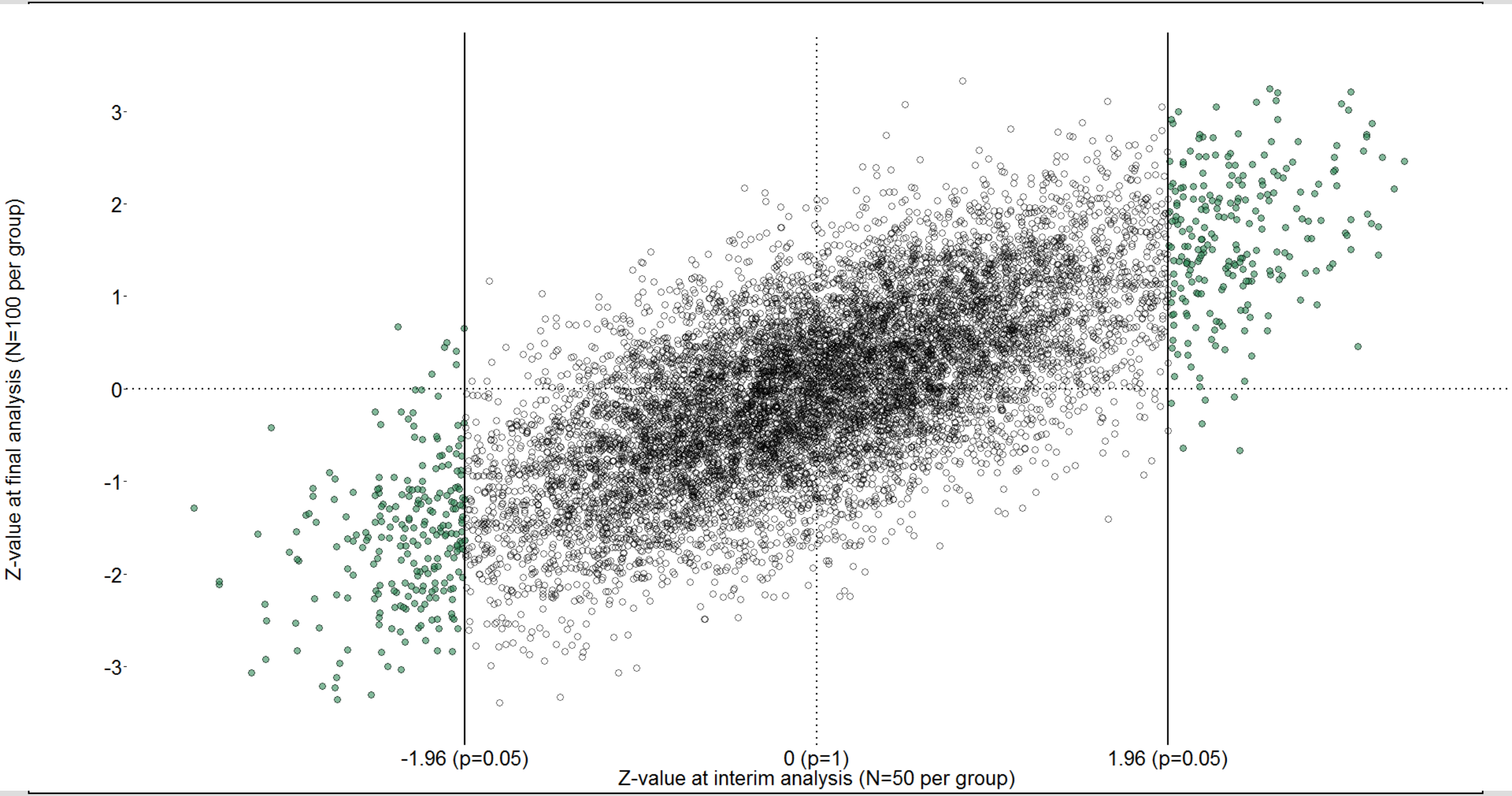


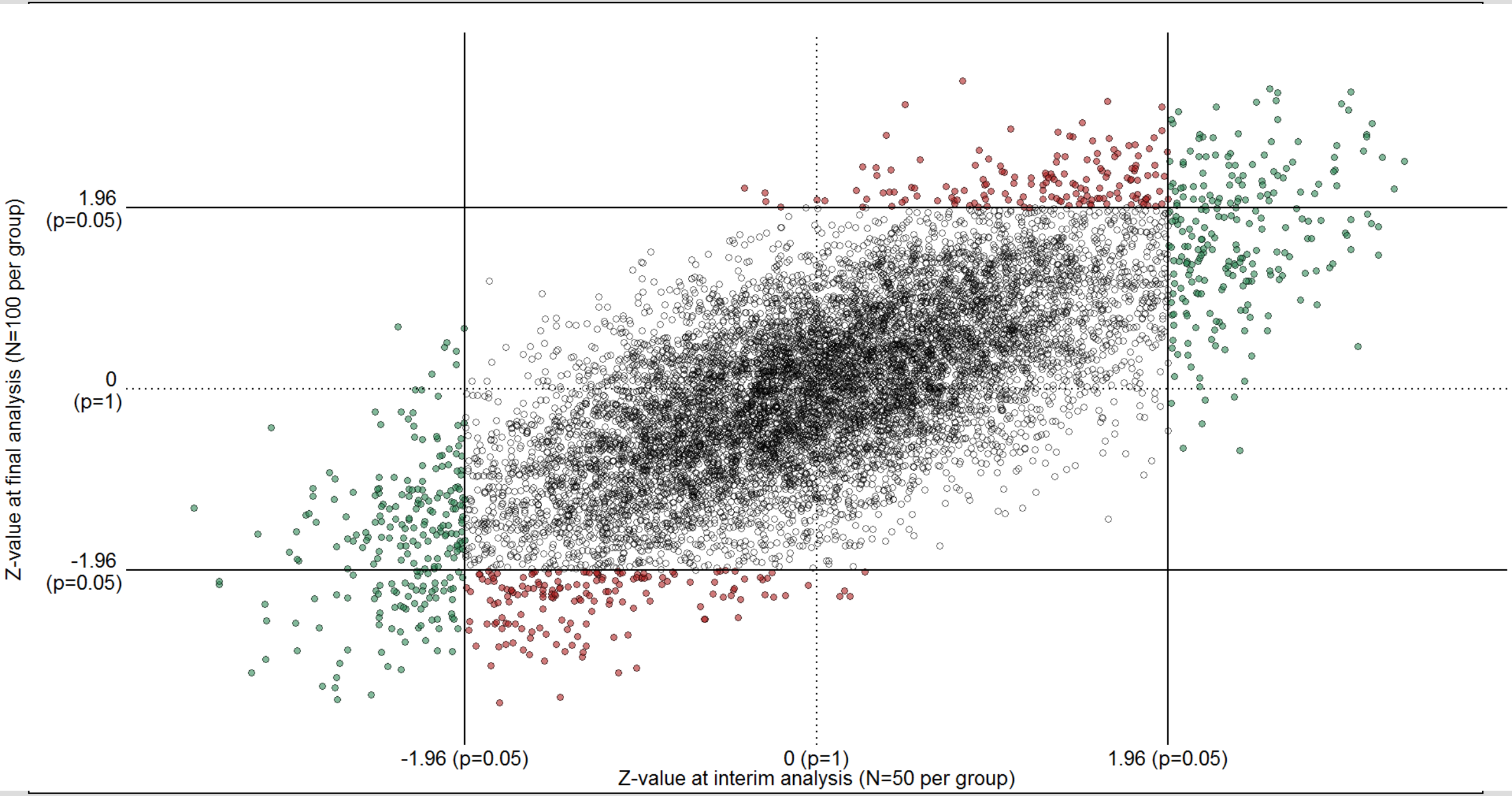
Z-value	P-value
2.04	0.04135
2.413	0.0158
4.562	0.000005

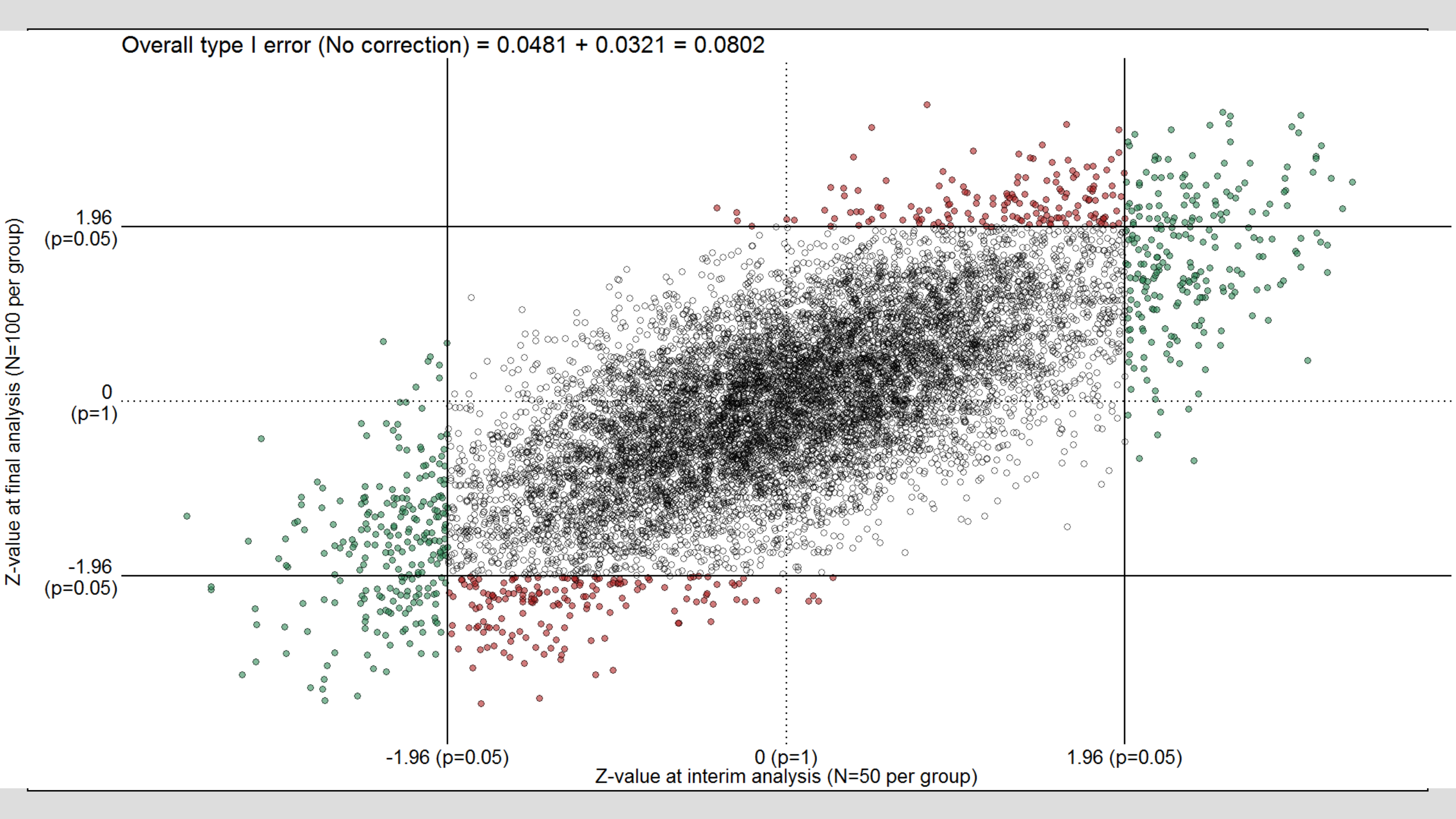
GSD – simulations

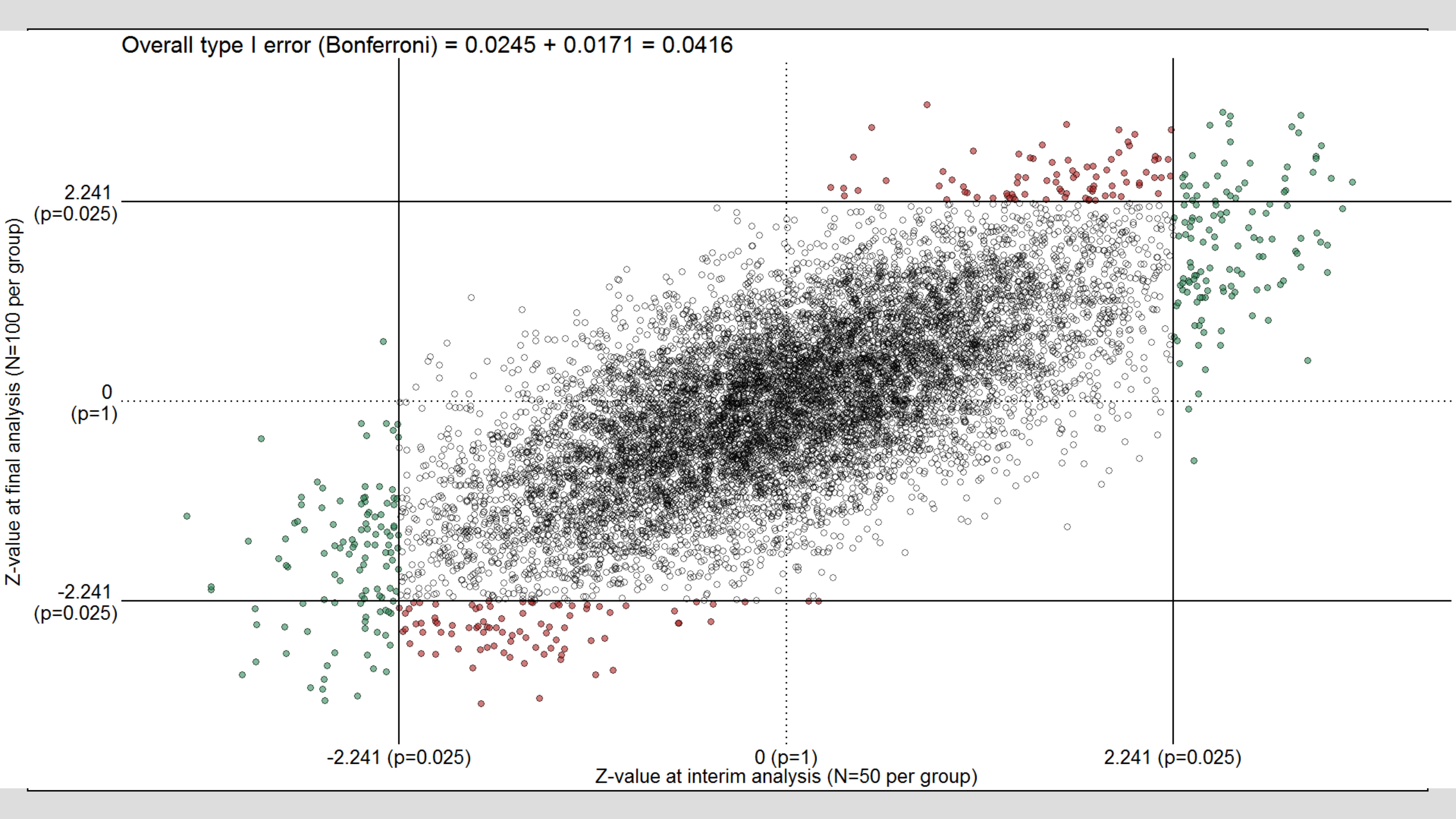
- How type I error is inflated with repeated significance testing
- Example
  - mean of treatment A : $\mu_A = 0$
  - mean of treatment B : $\mu_B = 0$
- 100 observations (subjects) in each group distributed as
  - $X_{Ai} \sim N(0,1)$
  - $X_{Bi} \sim N(0,1)$
- Hypothesis (two-sided): $H_0 : \mu_A = \mu_B$ and $H_A : \mu_A \neq \mu_B$
- Interim analysis when 50 subjects are recruited in each group
- 10’000 replications

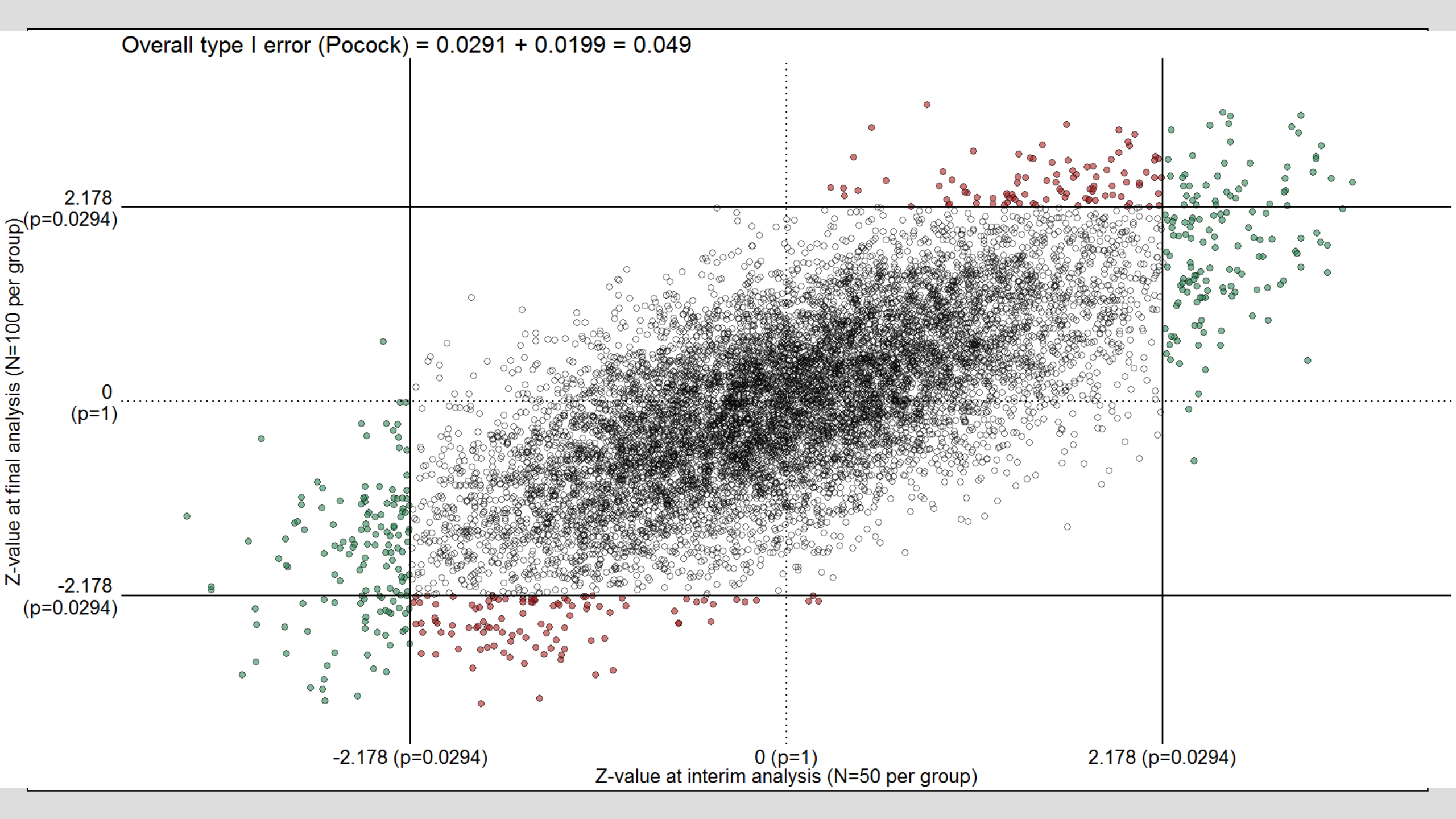

Z-value at final analysis (N=100 per group)
3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3
Z-value at interim analysis (N=50 per group)

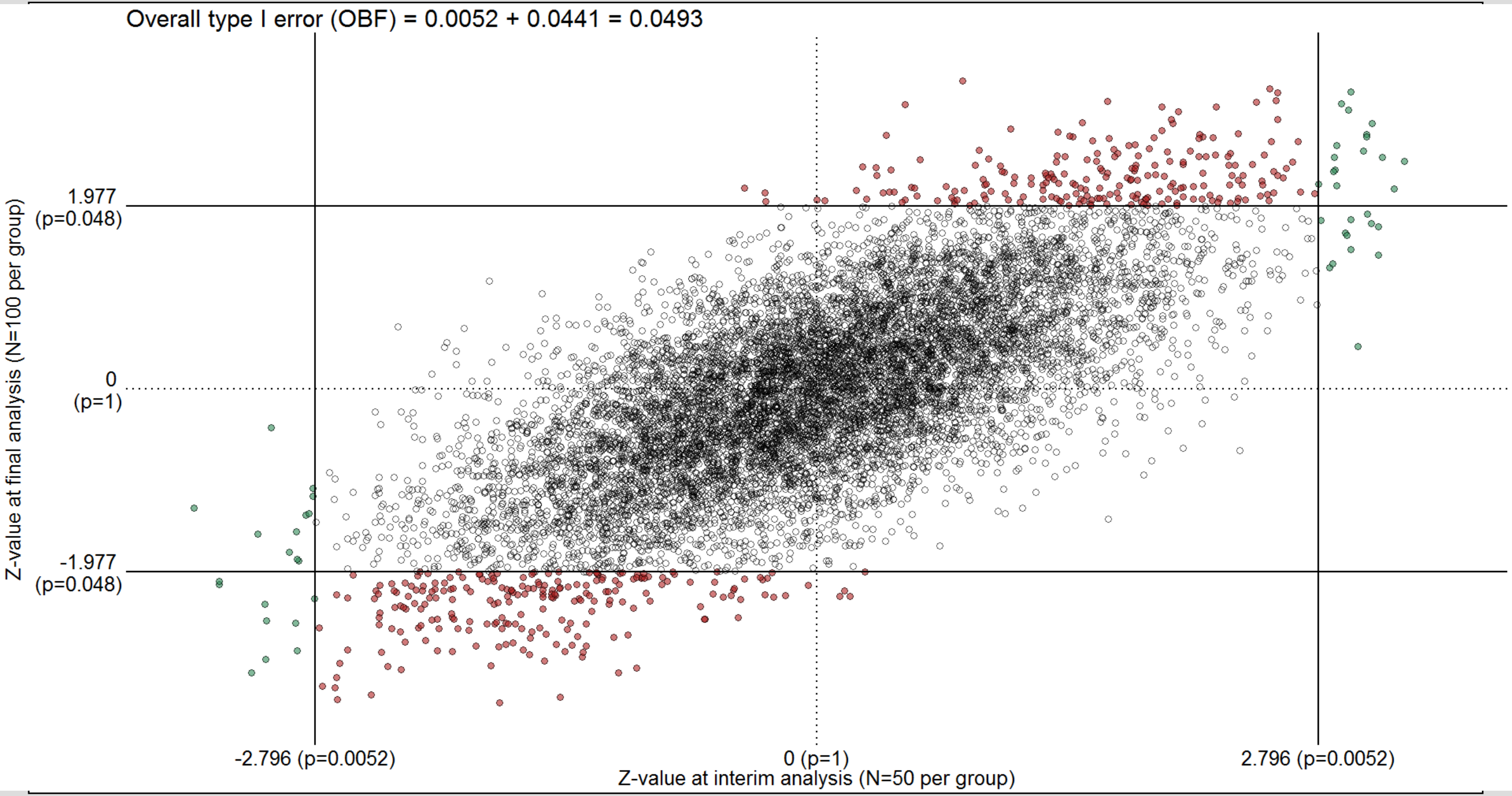

3
2
1
0
-1
-2
-3
Z-value at final analysis (N=100 per group)
-1.96 (p=0.05)
0 (p=1)
1.96 (p=0.05)
Z-value at interim analysis (N=50 per group)


1.96
(p=0.05)
0
(p=1)
-1.96
(p=0.05)
Z-value at final analysis (N=100 per group)
-1.96 (p=0.05)
0 (p=1)
1.96 (p=0.05)
Z-value at interim analysis (N=50 per group)


Overall type I error (No correction) = 0.0481 + 0.0321 = 0.0802
1.96
(p=0.05)
0
(p=1)
-1.96
(p=0.05)
Z-value at final analysis (N=100 per group)
-1.96 (p=0.05)
0 (p=1)
1.96 (p=0.05)
Z-value at interim analysis (N=50 per group)


Overall type I error (Bonferroni) = 0.0245 + 0.0171 = 0.0416
2.241
(p=0.025)
0
(p=1)
-2.241
(p=0.025)
Z-value at final analysis (N=100 per group)
-2.241 (p=0.025)
0 (p=1)
2.241 (p=0.025)
Z-value at interim analysis (N=50 per group)


Overall type I error (Pocock) = 0.0291 + 0.0199 = 0.049
2.178
(p=0.0294)
0
(p=1)
-2.178
(p=0.0294)
Z-value at final analysis (N=100 per group)
-2.178 (p=0.0294)
0 (p=1)
2.178 (p=0.0294)
Z-value at interim analysis (N=50 per group)


Overall type I error (OBF) = 0.0052 + 0.0441 = 0.0493
1.977
(p=0.048)
0
(p=1)
-1.977
(p=0.048)
Z-value at final analysis (N=100 per group)
-2.796 (p=0.0052)
0 (p=1)
2.796 (p=0.0052)
Z-value at interim analysis (N=50 per group)

Wang and Tsiatis power family (1987)

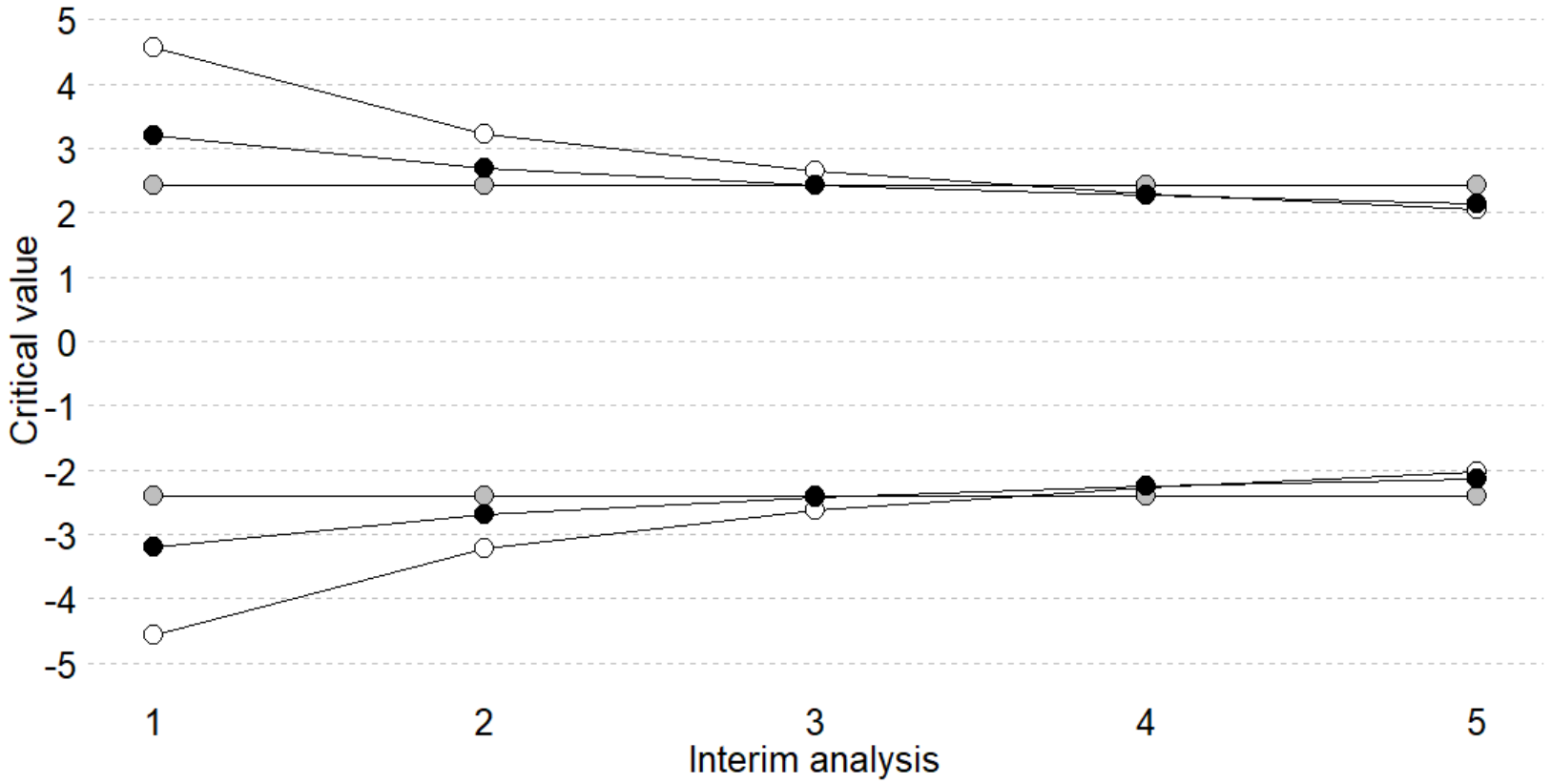
- Class of boundaries indexed by a power parameter Δ

$$c_i = c_{WT}(K, \alpha, \Delta) i^{\Delta - 0.5}$$

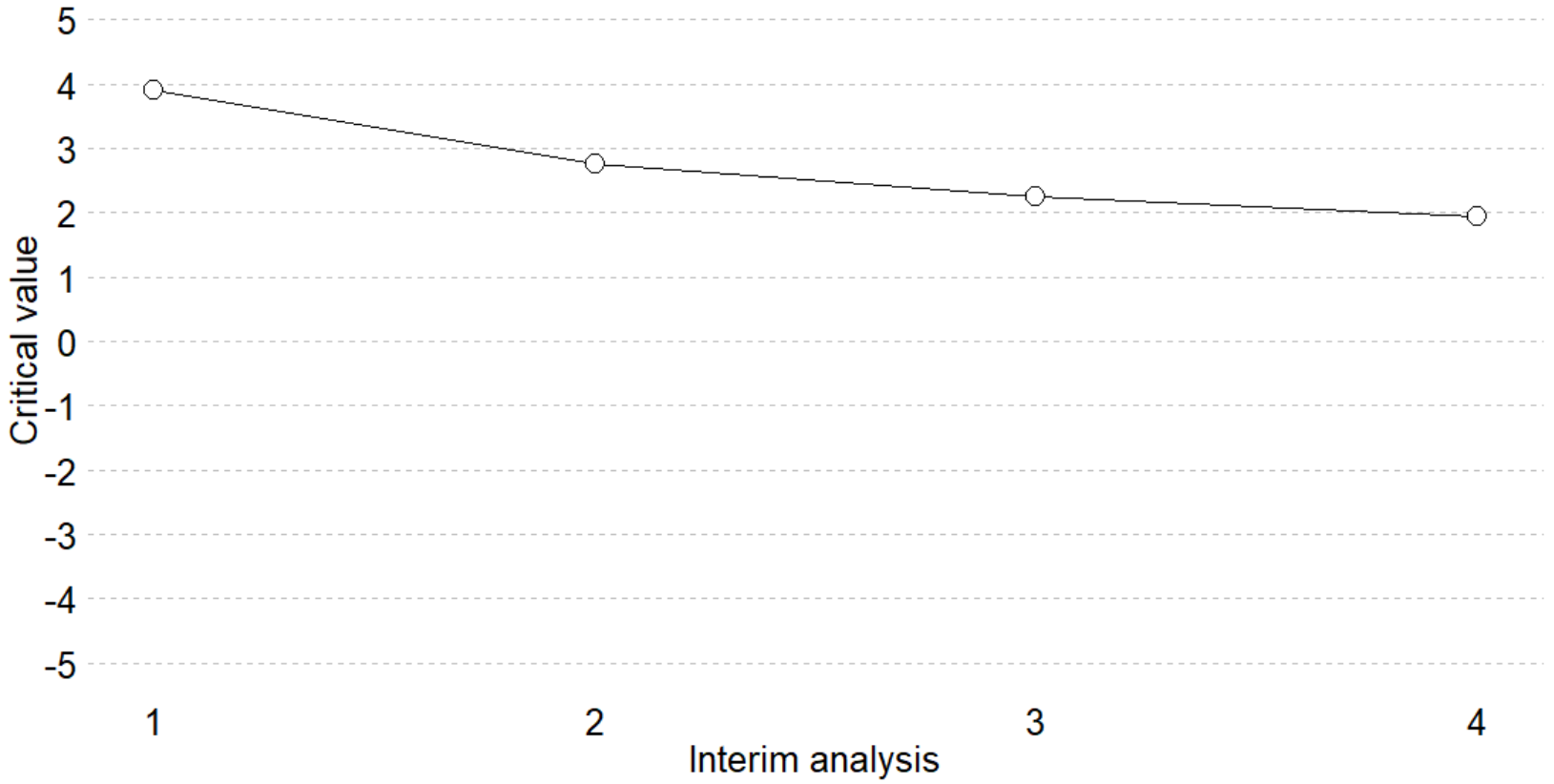
- $\Delta = 0$ → O'brien and Fleming. $\Delta = 0.5$ → Pocock

- Can use Δ to minimize the ASN for determining maximum sample size

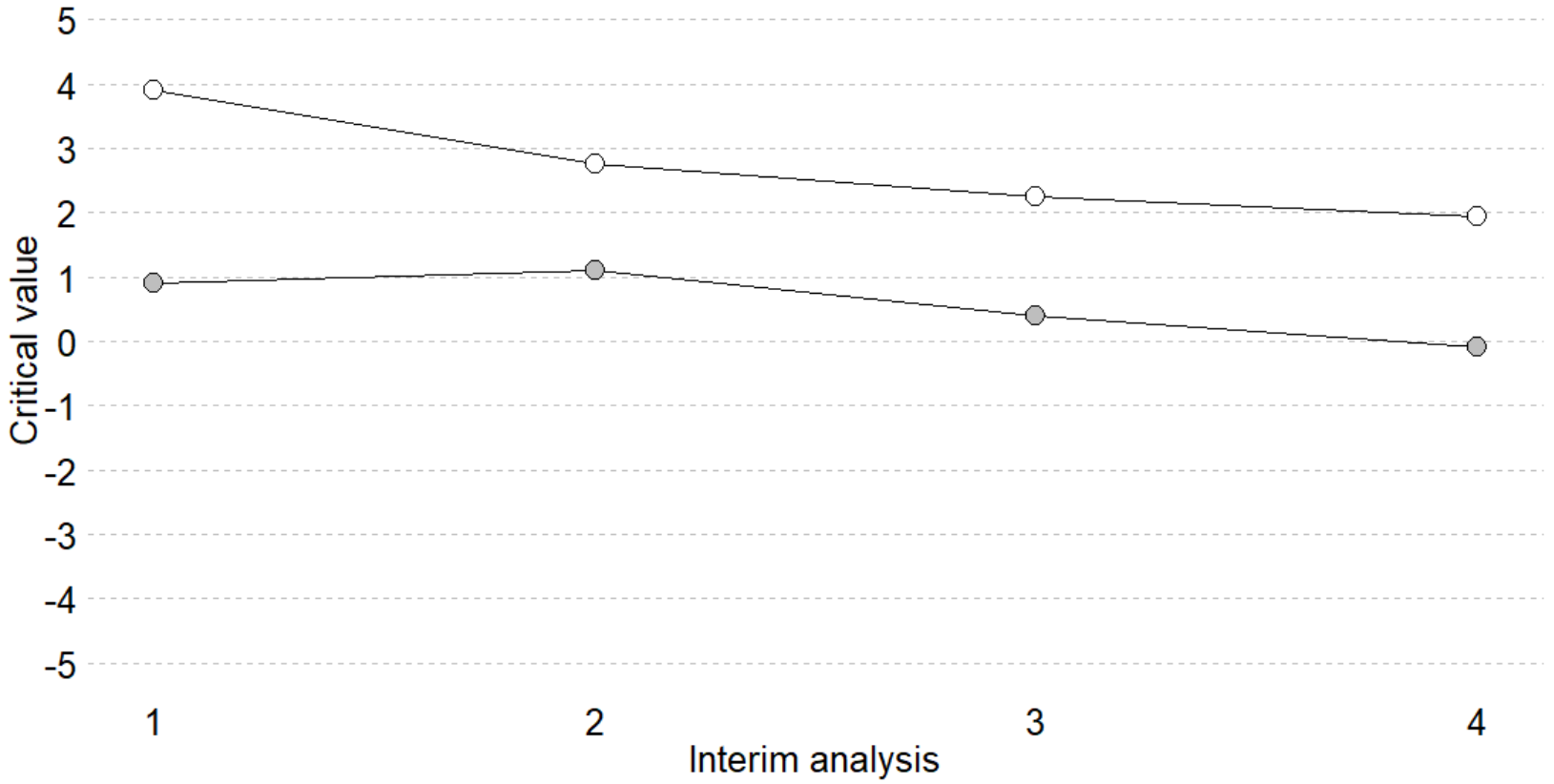
Stopping boundaries (WT : Δ=0.25)



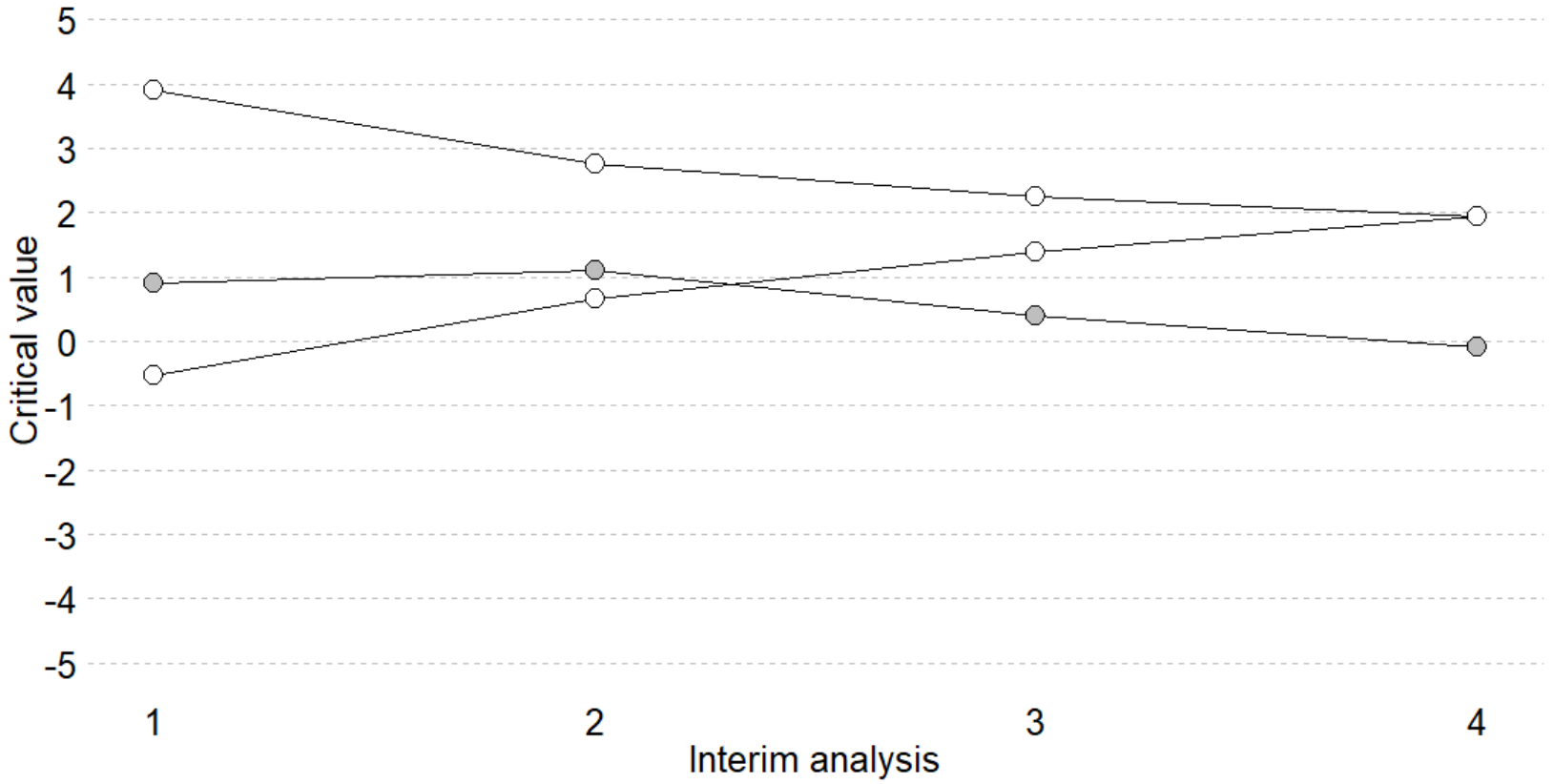
One-sided test and futility stopping



One-sided test and futility stopping



One-sided test and futility stopping



Binding and non-binding futility boundaries

- Analoglously to efficacy boundaries, futility boundaries attempt to control the type II error
- Type II error is under H_a, which assume a treatment effect Δ
- When adding futility boundaries, the critical efficacy boundaries can be slightly relaxed
- In this case, the futility boundaries are *binding* → stopping is mandatory when the boundary is crossed to preserve type I error
- Recommend to use *non-binding* boundaries → Futility boundaries serve as guidelines

Conditional power

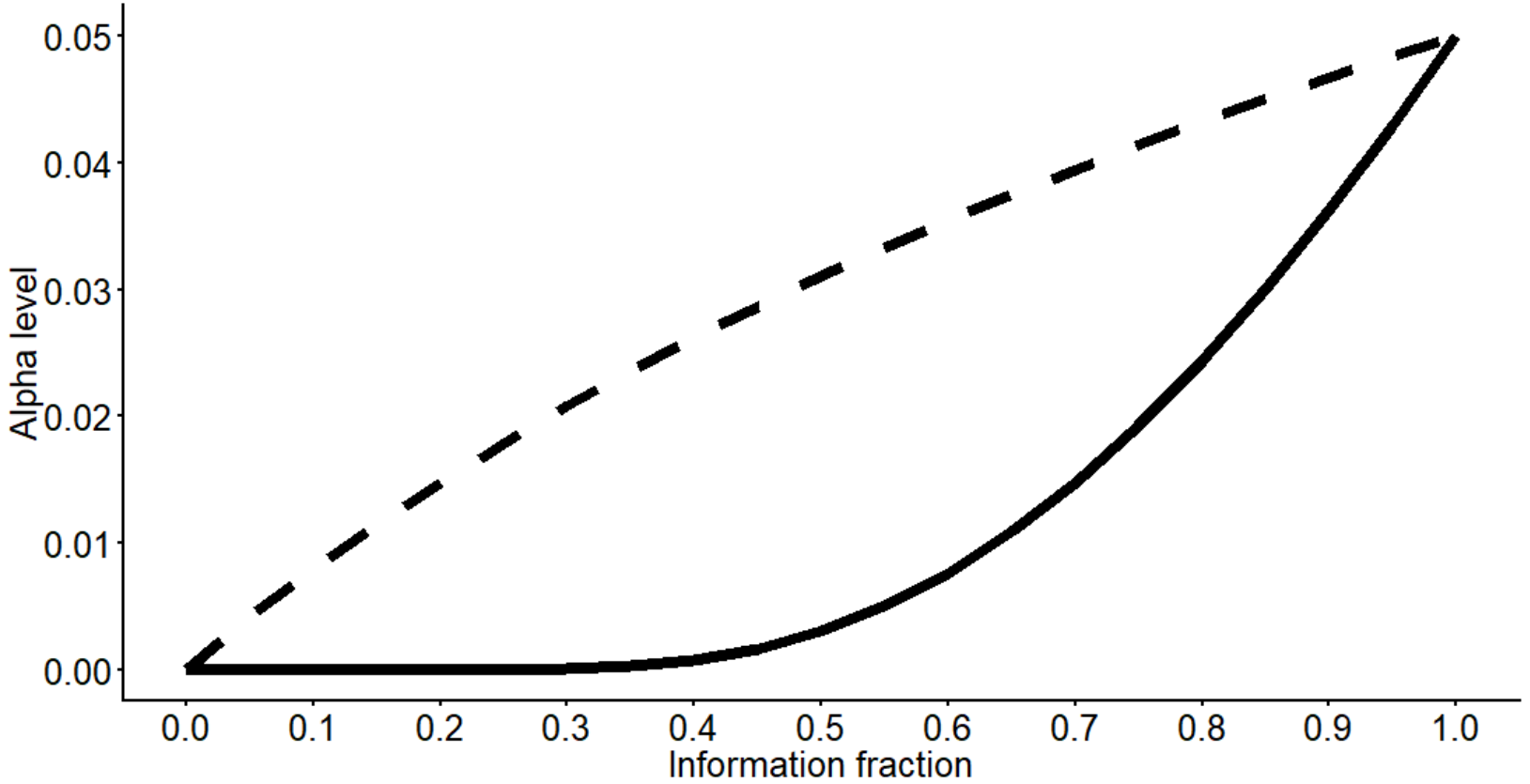
- Given the interim results, what is the chance of success at the final analysis?
- Conditional power (CP) allows to quantify this probability
- Continue the trial only if CP is high enough, e.g. CP = 0.8 or CP = 0.9

- Also possible to increase the sample size if CP not high enough
- Mehta and Pocock (2011) proposed the *promising zone* approach

α-spending function approach

- Lan & DeMets (1994)
- Allow flexibility in the number of interim analyses to be conducted and at what times
- α–spending function $\alpha(t_k)$ is an increasing function depending on t_k satisfying $\alpha(0) = 0$ and $\alpha(1) = \alpha$
- $t_k = n_k/N_{max}$ = information rate
- α considered as a budgeted quantity to spend

Pocock and OBF spending functions



Procedure α is spent

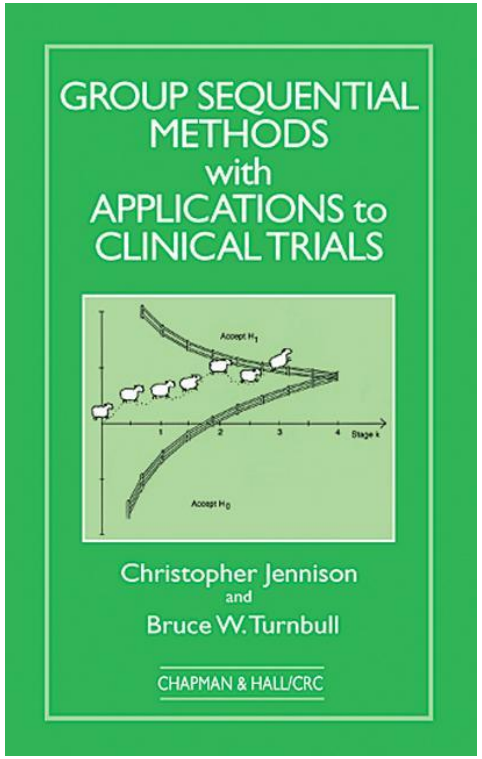
- $P_1 = P_{H_0}(|Z_1| \geq u_1) = \ \alpha(t_1)$
- $P_2 = P_{H_0}(|Z_1| < u_1, |Z_2| \geq u_2) = \ \alpha(t_2) - \alpha(t_1)$
- …
- $P_k = \alpha(t_k) - \alpha(t_{k-1})$

- $P_1 + \cdots + P_K = \alpha$

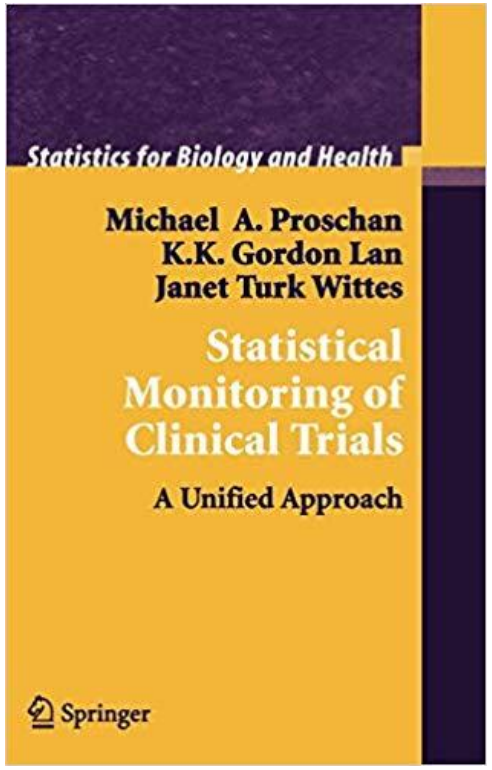
- Z_k depends only on the information rates t_1, … t_k,not on the unobserved ones $t_{k+1}, \ldots t_K$

References

(1999)



(2006)



(2016)