EFSPI Webinar Series:
Statistical Considerations in
Oncology Study Design

# Safety-Driven Response Adaptive Randomisation

## An Application in Non-inferiority Oncology Trials


**Maria Vittoria Chiaruttini, PhD**
University of Padua, Italy

**Sofía S. Villar, PhD**
MRC Biostatistics Unit, University of Cambridge, UK

**Lukas Pin, PhD**
MRC Biostatistics Unit, University of Cambridge, UK


February 26th , 2026s


**Contact:** mariavittoria.chiaruttini@studenti.unipd.it

BACKGROUND

# The Challenge in Modern Oncology

## FDA Project Optimus

**A paradigm shift toward patient-centered dosing** —seeking the optimal balance between drug efficacy and safety to maximize patient benefit while minimizing toxicity. This initiative challenges traditional dose-selection approaches.

## Modern Treatment Landscape

**Targeted therapies, monoclonal antibodies, and immune checkpoint inhibitors** typically span prolonged treatment cycles with chronic, low-grade adverse events (fatigue, GI disturbances, dermatologic reactions) that significantly impact adherence and quality of life.

## Plateauing Efficacy

**Novel agents often exhibit plateauing efficacy at doses well below MTD** , underscoring the need for a revised perspective on the interplay between safety and efficacy outcomes.

## Case Study: Sotorasib

### Initial Selection

960 mg dose selected for pivotal trials in KRAS G12C inhibitor

### Subsequent Analysis

240 mg dose offered similar efficacy with better safety profile

### Regulatory Action

**FDA mandated post-marketing dose optimization study**

Singh et al., Journal of Clinical Oncology 2025

## Key Insight

Traditional early-phase trials prioritize immediate, severe safety data, while later phases focus primarily on efficacy.

**Safety information is typically not formally integrated into late-phase design, despite the massive amount of safety data that become rapidly available from the very early follow-up period.**

RESEARCH MOTIVATION

# Research Question & Motivation

## Core Research Question

How can a **late-phase adaptive design** help identify **alternative dosing strategies** (lower or fractionated doses, different schedules) with **comparable or superior efficacy** and **fewer side effects** ?

## The Challenge with Traditional RAR

### Delayed Efficacy Outcomes

Most RAR designs rely on efficacy data to guide allocation, but in oncology trials, primary efficacy outcomes (e.g., overall survival, PFS) are observed with random delay.

### Limited Applicability

When efficacy outcomes are unavailable during enrollment, traditional RAR becomes challenging due to lack of timely, informative data for adaptation.

### Surrogate Endpoint Limitations

While surrogate endpoints have been proposed, they require validation and may not always be available or appropriate.

## Our Solution

### 1 Leverage Early Safety Data

Use early-emerging safety data (e.g., time to dose-reduction/discontinuation) to inform treatment allocation decisions.

### 2 Non-inferiority Framework

Particularly relevant for non-inferiority trials to demonstrate experimental treatment is not inferior to standard of care while offering improved tolerability.

### 3 Address Dosing Decisions

Address the potential impact on Phase II/III studies from dosing decisions in early trials often based on immediate and severe safety data.

> **"Agents with delayed efficacy signals and cumulative toxicity profiles that influence long-term adherence and outcomes."**

FUNDAMENTAL CHALLENGE

# The Statistical Trade-off

## Balanced Allocation: The Gold Standard

In non-inferiority trials, **a balanced 1:1 allocation ratio is typically optimal** due to the mathematical structure of the variance.

**Mathematical Rationale:**

- Under $H_1$, treatment effects are clinically equivalent within a narrow margin
- Outcome variances in both arms are expected to be similar

**Neyman Allocation Insight:**
Strategies minimizing variance of estimated treatment effect (Neyman allocation) tend to **approximate or coincide with balanced allocation** in trials with Exponentially distributed outcomes (e.g., time to progression/death).

## The Unavoidable Trade-off

### Statistical Efficiency

Preserve inferential efficiency for the primary non-inferiority outcome. Maintain adequate power to demonstrate that experimental treatment is not inferior to control.

### Safety Considerations

Integrate safety considerations into the randomization process through RAR. Preferentially allocate patients to arms with superior safety profiles.

## The Safety-driven RAR Challenge

Response-adaptive randomization driven by safety data can dynamically skews allocation favoring the SAFER arm, which may **deviate from the optimal variance-reducing strategy** for the efficacy endpoint test.

**Power Loss**
Unequal allocation can reduce statistical power for detecting non-inferiority.

**Patient Benefit**
More patients assigned to better-tolerated treatment arms.

## Key Considerations

1. **Clinical Risk:** Assigning more patients to better-tolerated arms is not always optimal if that arm is less effective.
2. **Statistical Risk:** RAR designs may introduce challenges in preserving statistical power under certain conditions.
3. **Need for Balance:** A design that can dynamically adjust based on the observed association between endpoints.

SAFEGUARD MECHANISM

# Efficacy-Safety Association

## The Safeguard Mechanism

If the allocation is primarily driven by safety, there's an **inherent risk of preferentially assigning patients to a treatment that, although better tolerated, may be less effective**.

### Solution:

The design should incorporate a **safeguard mechanism** that exploits the estimated relationship between efficacy and safety outcomes. It can detect situations where an apparent safety advantage is offset by insufficient efficacy.

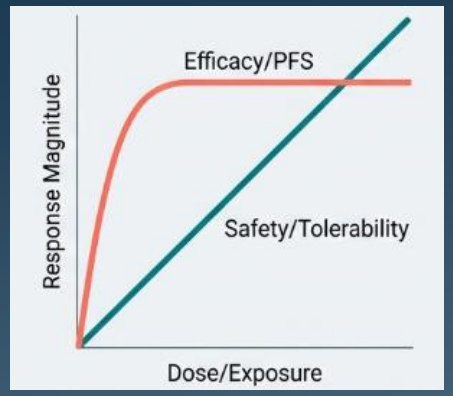
**Without Safeguard Mechanism**

Risk of allocating to safer but potentially futile or inadequately active treatment arm.

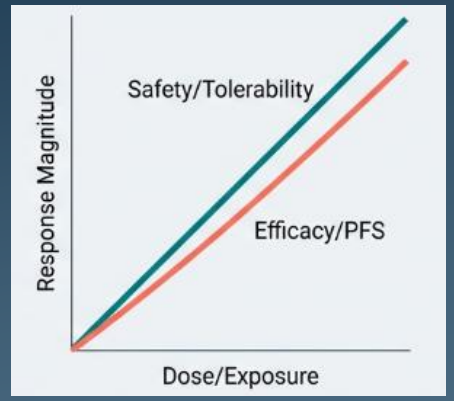
**With Safeguard Mechanism**

Allocation reverts toward 1:1 if efficacy data fail to demonstrate benefit.

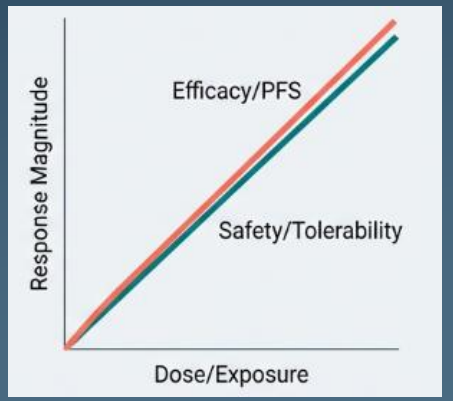
The plateau target




The trade-off risk




The cytotoxic paradigm



PROPOSED SOLUTION

# Introducing SAFER: Design Overview

## SAFER

### Safety-Aware Flexible Elastic Randomization

A Novel RAR Design

SAFER dynamically adjusts patient allocation proportions based on accumulating safety endpoint data, with the aim of **leveraging the association between efficacy and safety outcomes** to balance clinical benefit and statistical validity.

## Key Innovations

### Safety-Driven Adaptation

Uses **early-emerging safety data** (time to dose-reduction/discontinuation) rather than delayed efficacy outcomes to inform allocation decisions.

### Endpoint Association

Explicitly evaluates and incorporates **the interplay between safety and efficacy** endpoints through the **Φ weight parameter**.

### Flexible Elasticity

The **η parameter controls adaptation aggressiveness**, allowing customization based on clinical context and trial objectives.

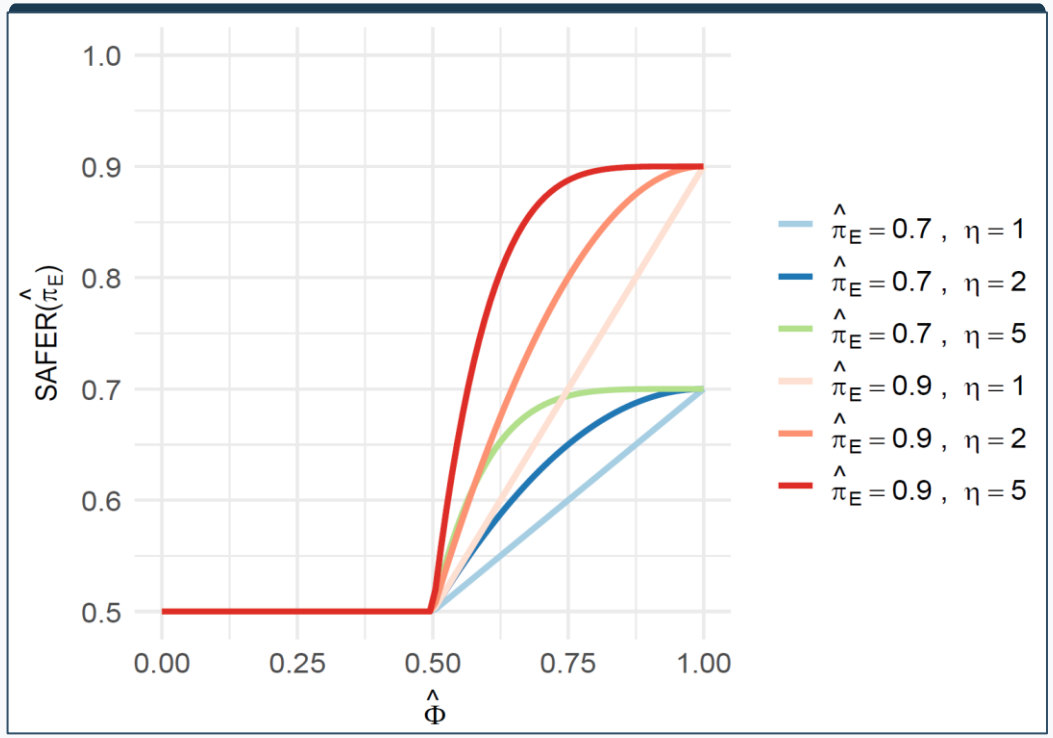
## SAFER Design Workflow

1. **Burn-in Period**
   Equal 1:1 randomization for initial patients (e.g., first 3 months)
2. **Safety Data Collection**
   Accumulate safety data from enrolled patients
3. **Estimate Target Allocation**
   Compute target allocation  based on safety data to the Experimental a
4. **Evaluate Efficacy Evidence**
   Calculate Φ weight from available efficacy data
5. **Apply SAFER Function**
   Compute SAFER() incorporating both safety and efficacy information
6. **Update & Continue**
   Use updated allocation for next enrollment period, repeat process

CORE INNOVATION

# The SAFER Function

## SAFER Function Visualization




## Mathematical definition

**SAFER($\widehat{\pi_E}$) =**

if $\widehat{\Phi}$ OR $\widehat{\pi_E} \le 0.5$: 0.5

if $0.5 < \widehat{\Phi} < 1$: $0.5 + (\widehat{\pi_E} - 0.5)*(1-(1-\frac{\widehat{\Phi}-0.5}{0.5})^{\eta})$

if $\widehat{\Phi} = 1$: $\widehat{\pi_E}$

**Weight Parameter (Φ):**

Larger values = stronger efficacy evidence.

**Elastic Parameter (η):**

Controls adaptation aggressiveness. η = 1 (linear), η > 1 (concave, conservative early).

## Function Behavior

**$\widehat{\Phi} \le 0.5$ (Weak Efficacy)**
Allocation reverts to **0.5** (equal randomization). Protects against assigning patients to **safer but potentially inferior arm.**

**$0.5 < \widehat{\Phi} < 1$ (Moderate Efficacy)**
Gradual increase in allocation based on η. Higher η = more conservative early, aggressive late.

**$\widehat{\Phi} = 1$ (Strong Efficacy)**
Allocation reaches **$\hat{\pi}_E$** (full Neyman allocation). Maximum benefit from safety advantage.

## Key Insight

The SAFER function provides **regularization** that prevents premature and extreme imbalances in allocation. This is especially critical in non-inferiority settings where the experimental arm could ultimately prove inferior.

MOTIVATING EXAMPLE

# The CAPP-IT Study & Re-design framework

## Original Study Design

**Study Type**
Multicentre, randomized, double-blind, placebo-controlled Phase III trial (10 UK sites)

**Population**
Patients with colorectal or breast cancer receiving a treatment with single-agent capecitabine

**Intervention**
Concomitant pyridoxine vs. matching placebo to reduce capecitabine dose modifications

**Primary Objective**
Assess whether pyridoxine could reduce need for capecitabine dose reduction

## Original Study Results

**Recruitment Challenge**
Study closed prematurely after enrolling **106 patients** (target: 270)

**Objective Response Rate**
Pyridoxine did not demonstrate statistically significant effect (OR: 1.37; 95% CI: 0.475–3.96)

**PFS Result (Concerning)**
Median PFS: 7.4 months (pyridoxine) vs. 9.9 months (placebo)
HR: 1.62 (95% CI: 0.91–2.88)

## SAFER Re-design Framework

### Primary Endpoint

The design accommodates typical Phase II/III oncology trial endpoints—**time-to-event measures such as PFS**. Sample size and primary analysis are expressed in terms of this endpoint.

### Non-inferiority Hypothesis

**Control Arm PFS monthly rate**
$\lambda_C = 0.069$
Median ≈ 10 months

**Non-inferiority Margin**
$HR_0 = 1.25$
Median > 8 months

Under regularity conditions, construct **Wald statistic** for hypothesis testing.

### Safety Endpoint Definition

Time to dose-reduction or drug discontinuation due to low drug tolerability or adverse events (AEs).

#### Clinical Rationale

This endpoint captures treatment tolerability and directly impacts patient adherence and quality of life—critical in modern oncology with prolonged treatment cycles.

DESIGN COMPONENTS

# Target Allocation

## Target Allocation

Goal: The larger is the mean in the experimental arm ($\hat{\theta}_{E(s)}$), the higher is the proportion assigned to the Experimental Arm

**Estimated Target Allocation:**

$\hat{\theta}^2_{E(s)}$: Variance of time to drug discontinuation in arm E (experimental)

$\hat{\theta}^2_{C(s)}$: Variance of time to drug discontinuation in arm C(control)

Neyman proportion = $\widehat{\pi_E} = \hat{\theta}^2_{E(s)} / (\hat{\theta}^2_{E(s)} + \hat{\theta}^2_{C(s)})$

## Sequential Update Mechanism

0 **Burn-in Period (u = 0)**
$\hat{\pi}E$ = 0.5 (equal allocation) for patients before first update

1 **First Update (u = 1)**
$\hat{\pi}E$ estimated using all available safety data up to first update time

2 **Subsequent Updates (u = 2, ..., U)**
Pattern continues with increasingly refined estimates

**Implementation Details**

U = 15 updates every 3 months (months 3-45), 48-month enrollment period. Updates independent of interim analyses.

## Why Neyman Allocation?

**1. Ethical Alignment (Exponential)**

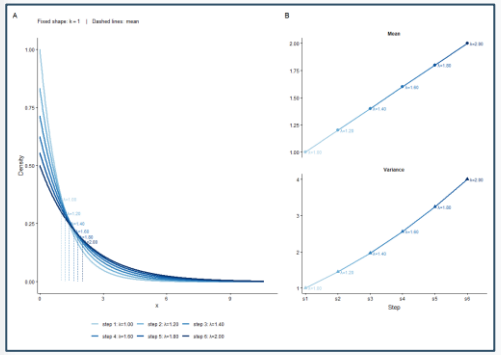
Under **exponential distribution**, Neyman allocation consistently assigns majority to treatment with longer expected time to dose reduction (better tolerated).

**2 Statistical Efficiency**

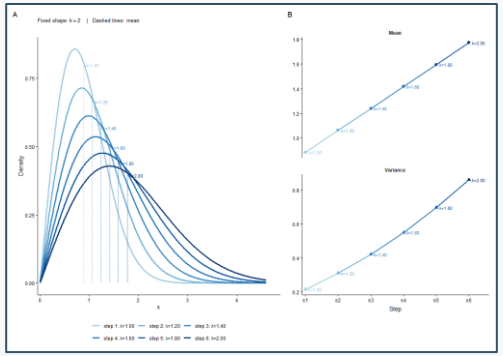
Maximizes power of Wald test to detect true difference between arms by minimizing variance of treatment effect estimator.

**Note:** This alignment is specific to exponential distribution and may not hold for binary endpoints. However, it is robust in many cases when the Weibull is assumed as underlying distribution , since variability and mean increase /decrease on the same direction!!!!

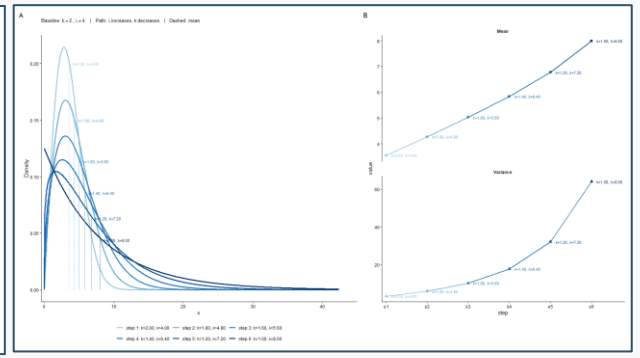
**K=1, λ= from 1 to 2**



**K=2, λ= from 1 to 2**

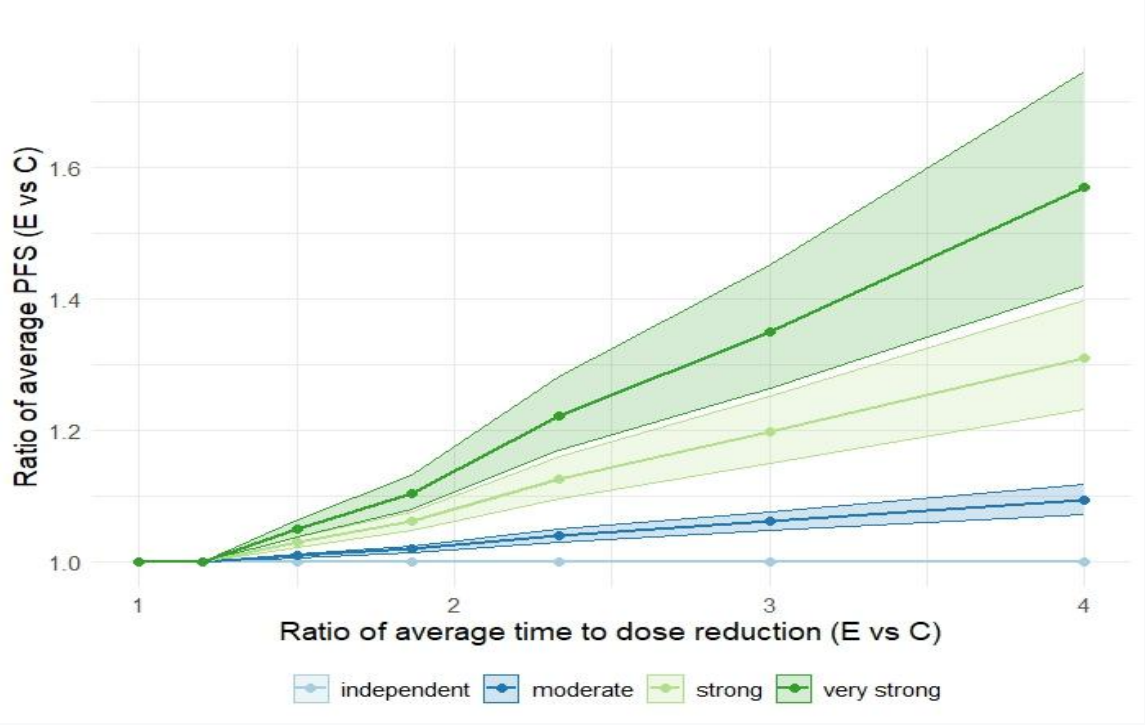


**K=from 1 to 2, λ= from 4 to 8**



SIMULATION MODEL

# Endpoints Linear Association




## How The Φ Weight Parameter is Estimated?

**Definition:**

$\hat{\phi}$ is calculated as the **cumulative distribution function (CDF)** of the Wald-statistic (Z-scale) from the efficacy analysis (Cox model).

**Transformation:**

For Cox model (negative coefficient = risk reduction), **invert z-value sign** so larger $\hat{\phi}$ consistently corresponds to stronger efficacy evidence.

**Range:** $\hat{\phi} \in [0, 1]$, with default threshold at 0.5 (z-statistic = 0)

## Why Not Simple Correlation to detect their association?

A simple correlation metric between safety and efficacy is **not appropriate** in this context because it fails to account for the specific characteristics of survival endpoints.

1. **Administrative Censoring:** Both endpoints subject to truncation from study end.
2. **Structural Truncation:** Safety endpoint truncated by PFS events (follow-up ends at progression).
3. **Misleading Estimates:** Conventional correlation coefficients may yield misleading association estimates.

EVALUATION FRAMEWORK

# Simulation Study Design

## Seven Simulation Scenarios

**Scenario 0: Independence Baseline**
Safety-driven RAR without SAFER under independence between endpoints (Type I Error evaluation under the null hypothesis)

**Scenario 1: Complete Randomization**
1:1 allocation with varying association levels (baseline for comparison)

**Scenario 2: Safety-driven RAR**
Pure safety-driven RAR without SAFER under varying association

**Scenario 3a/b: SAFER Design**
SAFER with $\eta = 1$ (3a) and $\eta = 5$ (3b) under varying association

**Scenario 4: Endpoint Timing**
Impact of differential timing in endpoint observation (PFS: 3-24 months)

**Scenario 5: Informative Drop-out**
Impact of informative drop-out (5% to 25%) with **composite strategy**

**Scenario 6: Under-reporting**
Impact of under-reported safety events on operating characteristics

## Endpoint Association Levels

| | | |
|---|---|---|
| **Independent** | No relationship between safety and efficacy | $E[\gamma_1] = 0.001$ |
| **Weak** | Minimal positive association | $E[\gamma_1] = 0.005$ |
| **Moderate** | Meaningful positive association | $E[\gamma_1] = 0.01$ |
| **Strong** | Substantial positive association | $E[\gamma_1] = 0.03$ |
| **Very Strong** | Strong positive association | $E[\gamma_1] = 0.05$ |

## Target Allocations

| | |
|---|---|
| $\pi_E = 0.5$ Equal allocation | $\pi_E = 0.6$ Mild preference |
| $\pi_E = 0.7$ Moderate preference | $\pi_E = 0.8$ Strong preference |

## Performance Metrics

### Statistical Power

Overall power (interim + final) and power at interim ($power_1$) under $H_1$

### Type-I Error Rate

Proportion of trials rejecting $H_0$ under null hypothesis (target: 0.05)

### Allocation Proportion

Average proportion of patients assigned to experimental arm ($N_E/N$)

### Adverse Event Rate

Total AEs divided by total person-years (rate per patient-year)

## Simulation Setup

**Iterations:** 10,000 per scenario

**Monte Carlo Error:** < 0.5% for power and type-I error

**Software:** R with survival package

**Validation:** Theoretical power calculations vs. empirical results

SIMULATION RESULTS

# Scenario 0: Independence Between Endpoints

## Allocation Proportion & Power Summary

| πE | NE/N | Metric | IF=0.2 | IF=0.3 | IF=0.4 | IF=0.5 |
|---|---|---|---|---|---|---|
| **0.5** | 0.50 | Power | **0.80** | **0.80** | **0.80** | **0.80** |
| | | Type-I Error | 0.048 | 0.050 | 0.048 | 0.053 |
| **0.6** | 0.59 | Power | 0.78 | 0.79 | 0.78 | 0.78 |
| | | Type-I Error | 0.047 | 0.050 | 0.055 | 0.053 |
| **0.7** | 0.69 | Power | 0.74 | 0.74 | 0.73 | 0.73 |
| | | Type-I Error | 0.049 | 0.051 | 0.051 | 0.053 |
| **0.8** | 0.78 | Power | **0.66** | **0.66** | **0.65** | **0.66** |
| | | Type-I Error | 0.050 | 0.055 | 0.056 | 0.051 |

## Key Validation Results

**Allocation Accuracy**

NE/N closely aligns with πE across all information fractions

**Power Target Achievement**

0.80 maintained at πE=0.5; expected decline at higher allocations

**Type-I Error Control**

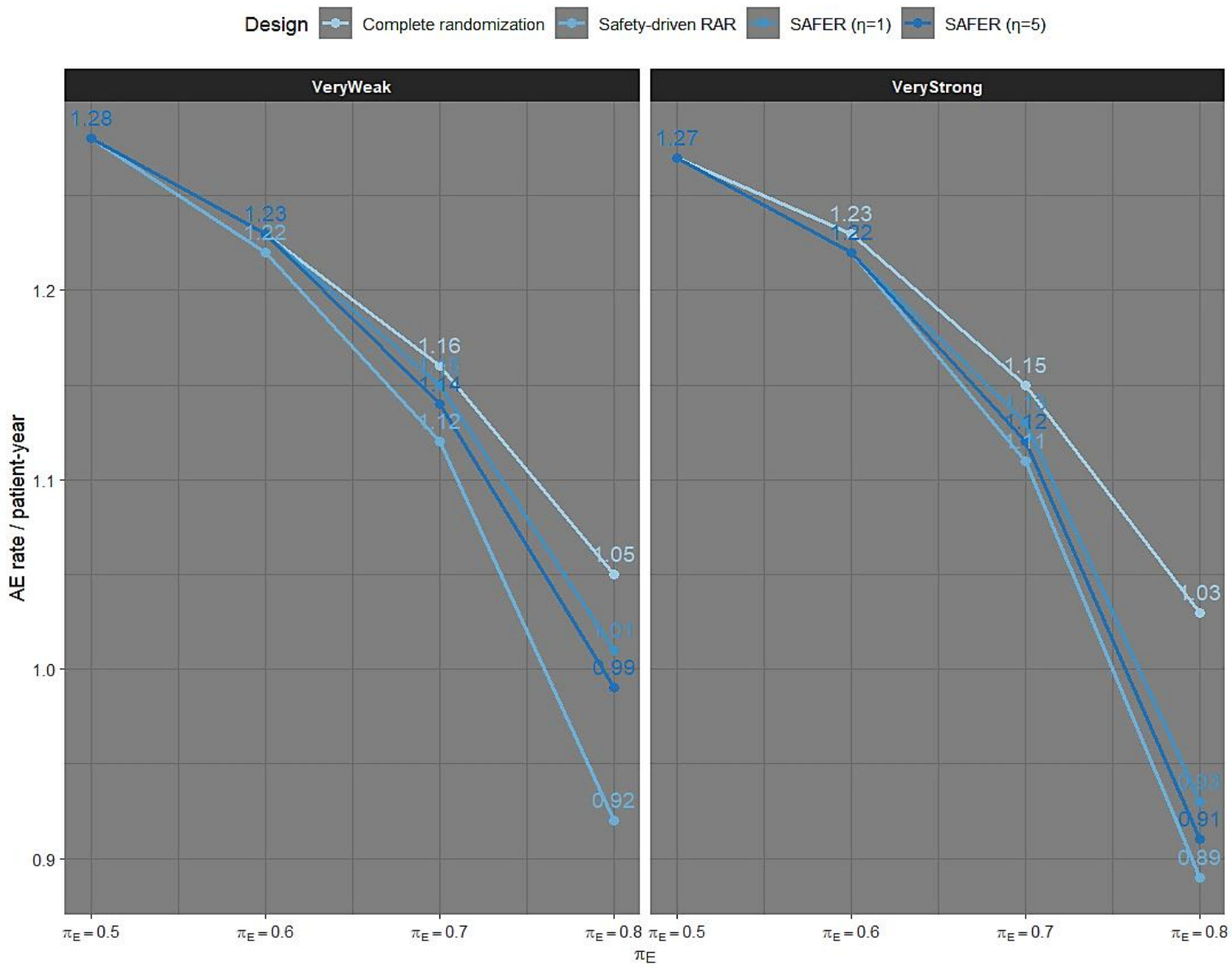
Remains close to nominal 0.05; no consistent inflation pattern

## Power vs. Allocation Trade-off


πE = True target allocation; NE/N = Estimated allocation; IF: information fraction at interim.

SIMULATION RESULTS

# Adverse Event Rate Reduction




## Key Findings

**Complete Randomization** 1.03

**SAFER (η=1)** 0.93

**10%** reduction vs. complete RAR

**SAFER (η=5)** 0.91

**12%** reduction vs. complete RAR

## Clinical Significance

SAFER designs achieve **AE rates below 1.0** for high allocation proportion values with both weak and strong endpoint associations. This represents substantial patient benefit through preferential assignment to better-tolerated arms while maintaining statistical power.

SIMULATION RESULTS

# Comparative Design Performance

## Performance Summary by Design Approach

| πE | Association | Complete RAR | | | Safety-driven RAR | | | SAFER η=1 | | | SAFER η=5 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | NE/N | $p_1$ | Power | NE/N | $p_1$ | Power | NE/N | $p_1$ | Power | NE/N | $p_1$ | Power |
| 0.8 | Very Weak | 0.50 | 0.23 | 0.81 | 0.78 | 0.16 | 0.66 | **0.57** | 0.22 | **0.81** | 0.62 | 0.20 | 0.81 |
| | Weak | 0.50 | 0.28 | 0.87 | 0.78 | 0.19 | 0.73 | **0.59** | 0.27 | **0.87** | 0.64 | 0.25 | 0.86 |
| | Moderate | 0.50 | 0.34 | 0.92 | 0.78 | 0.24 | 0.79 | **0.60** | 0.33 | **0.92** | 0.66 | 0.31 | 0.91 |
| | Strong | 0.50 | 0.62 | 0.99 | 0.79 | 0.46 | 0.95 | **0.66** | 0.59 | **0.99** | 0.71 | 0.56 | 0.99 |
| | Very Strong | 0.50 | 0.82 | 1.00 | 0.79 | 0.66 | 0.99 | **0.70** | 0.80 | **1.00** | 0.74 | 0.76 | 1.00 |

$p_1$ = Power at interim; Power = Overall power; Bold gold = SAFER optimal performance

Red = Power loss >10% from target

### Complete RAR

Power gains from endpoint association. power reaches 0.82 at πE=0.8 with very strong association

### Safety RAR

Power decreases at weak associations, gains at strong. Risk of substantial power loss without SAFER

### SAFER η=1

Shrinks allocation toward 0.5 at weak associations. Restores power while favoring better-tolerated arm

### SAFER η=5

More aggressive approach reaching target allocation. Slight power reduction vs. η=1 but greater patient benefit

CONCLUSIONS

# Key Takeaways & Future Directions

Safety-Driven Response Adaptive Randomisation in Non-inferiority Oncology Trials

## Main Contributions

1. **Power Preservation:** SAFER preserves statistical power while reducing adverse event rates through intelligent allocation adjustment.
2. **Flexible Adaptation:** The design offers flexible adaptation speed based on temporal alignment of endpoints—faster when safety and efficacy are temporally aligned.
3. **Safeguard Mechanism:** The Φ weight parameter protects against allocating to safer but potentially inferior arms when efficacy evidence is insufficient.
4. **Practical Customization:** Comprehensive 9-step guidance enables adaptation across different trial settings, endpoint types, and monitoring schemes.

## Clinical Impact

**Patient-Centered Design**
Enhances patient treatment experience by preferentially allocating to better-tolerated arms while maintaining statistical rigor.

**Regulatory Alignment**
Supports FDA Project Optimus objectives for dose optimization and benefit-risk balance in oncology drug development.

**Late-Phase Applicability**
Particularly suitable for Phase IIb/III trials where event accumulation is predictable and treatment effect estimates stabilize.

## Future Directions

**How to manage informative drop-outs**
Using a composite strategy may inflate type-I error when safety & efficacy diverge; interpret as combined measure only

**Multi-arm Extensions**
Extend SAFER framework to trials with multiple experimental arms or dose levels.

**Digital Endpoints**
Leverage digital health technologies for more frequent adaptation and real-time monitoring.

**Non-monotonic Relationships**
Explore impact of non-monotonic safety-efficacy relationships

**Alternative Distributions**
Evaluate performance under Weibull, log-normal, and other survival distributions.

**Empirical Validation**
Validate design performance using real-world oncology trial data and case studies.

> "SAFER offers a significant contribution in terms of practical tools to deliver patient-centred trials by adaptive learning from quickly observable safety data to proactively mitigate the assignment to poorly tolerated treatments without negatively affecting the power of the primary trial's objective."