Part 2. Choosing Design Parameters and Stopping Boundaries

Contents

2.1 Objectives 1

2.2 What needs to be fixed before you randomize anyone? 2

2.3 Group-sequential logic in a nutshell 4

2.4 Example 1 – continuous endpoint MAMS design (TAILoR-inspired) 5

2.5 Example 2 – survival MAMS design (STAMPEDE-inspired) 9

2.6 Example 3 – linking boundaries to real survival data (ovarian cancer) 11

2.7 Putting it all together: A practical design workflow 13

2.8 Take-Aways 15

In Part 1 we saw *why* multi-arm multi-stage (MAMS) and platform trials are attractive, using examples like STAMPEDE, I-SPY 2 and REMAP-CAP.

In this part we switch from "why" to **"how do I actually design one?"** – in particular:

- how to choose **design parameters** (number of arms, stages, effect sizes, error rates, allocation ratios),
- how to construct **efficacy and futility boundaries**, and
- how to interpret them using **real-world trial scenarios** in R.

2.1 Objectives

- List the main **design choices** in a MAMS trial.
- Describe how **group-sequential boundaries** (Pocock, O'Brien–Fleming, triangular) behave over time.
- Use the R packages `MAMS` and `gsMAMS` to:
  - design continuous-endpoint and survival MAMS trials,
  - obtain **stopping boundaries** and sample sizes, and
  - explore operating characteristics (FWER, power, expected sample size).
- Relate those boundaries to **actual trial data**, using real datasets such as the ovarian cancer trial from the `survival` package.

2.2 What needs to be fixed before you randomize anyone?

2.2.1 Number of experimental arms (K)

You need to decide how many treatments (or doses, or combinations) you will test **concurrently** against a shared control:

- **Clinical input:**
  - How many hypotheses are truly plausible?
  - Are multiple doses of the same drug needed (dose-finding) or multiple different regimens?
- **Statistical implications:**
  - More arms → more complex multiplicity / correlation structure.
  - But, as we saw in Part 1, often **fewer total patients** than running separate trials.

Packages like `MAMS` and `gsMAMS` treat this as a simple integer: `K` (or `k`), the number of treatment arms.

2.2.2 Number of stages (J) and information times

A MAMS trial has **J stages** (or "looks") at the data. At each stage you can:

- stop an arm (or the whole trial) for **efficacy**,
- stop an arm for **futility**, or
- continue to the next stage.

Crucially, what matters statistically is not the calendar time but the **information time**, usually defined by:

- number of patients accrued (continuous/binary endpoints), or
- number of events (survival endpoints).

For example, in a two-stage design you might plan interims at information fractions $(0.5, 1)$, i.e., after 50% and 100% of the target information has accrued.

2.2.3 Endpoints: intermediate vs. final

Most MAMS and platform trials:

- Use an **intermediate endpoint** to drive early decisions (e.g., failure-free survival in STAMPEDE, pathologic complete response in I-SPY 2).
- Use a **final, usually more definitive endpoint** for confirmatory conclusions (e.g., overall survival).

In this part we'll focus on **continuous** and **time-to-event** endpoints, since these cover a large proportion of real platform trials and are well supported in `MAMS` and `gsMAMS`.

2.2.4 Error rates and power

For confirmatory platforms, the usual target is:

- **One-sided familywise type I error (FWER)**: α, e.g. 2.5% or 5%.
- **Power**: 1-β, e.g. 80% or 90%, under some pre-specified configuration of true treatment effects.

The `MAMS::mams()` function explicitly controls the one-sided FWER across all treatment arms via a generalised Dunnett test, and finds sample sizes to guarantee power under a **least favourable configuration** (LFC).

`gsMAMS` similarly targets one-sided FWER and power when constructing sequential conditional probability ratio tests (SCPRT) boundaries.

2.2.5 Clinically relevant effect sizes

For each arm you need to distinguish:

- an **"interesting"** effect size δ (or hazard ratio (HR)) that you would like to detect with high power;
- an **"uninteresting"** or minimal effect δ_0 that would not justify further development. An effect that is "too small to care about".
- Note: δ_0 is not used in the interim analyses, but only in sample size calculation and solving for efficacy boundaries!

In `MAMS`, these are:

- `delta` / `delta0` for a continuous outcome (e.g., mean difference), or
- `p` / `p0` on the probability scale $P(X_k > X_0)$.

In `gsMAMS`, these appear as:

- `delta1` / `delta0` for continuous endpoints,
- `HR1` / `HR0` for survival endpoints.

2.2.6 Allocation ratios

You can randomize:

- **equally** to all arms (including the controls/placebo arm), or
- **unequally**, e.g., more patients to control or to particularly promising arms.

`MAMS` lets you specify per-stage allocation ratios to control and treatment arms (`r`, `r0`), while `gsMAMS` assumes balanced allocation by default.

From a clinical perspective, unbalanced allocation can be used to:

- limit exposure to a risky experimental treatment, or
- stabilize the control group estimates when many arms come and go.

2.3 Group-sequential logic in a nutshell

Before going fully multi-arm, it's worth recalling the **group-sequential** idea for a single treatment vs. control.

2.3.1 Interim test statistics

At each interim look/stage $j = 1, \ldots, J$ for treatment k, we compute a test statistic:

- continuous endpoint: $Z_{k,j} \approx \frac{\hat{\delta}_{k,j}}{\mathrm{SE}(\hat{\delta}_{k,j})}$
- survival endpoint:$Z_{k,j} \approx \frac{\log(\widehat{HR}_{k,j})}{\mathrm{SE}(\log(\widehat{HR}_{k,j}))}$

Under the null hypothesis (no difference), these statistics are asymptotically standard normal, but they are **correlated across stages** because they are based on overlapping data.

2.3.2 Efficacy and futility boundaries

At each stage j, we pre-define:

- a **lower (futility) boundary** a_j, and
- an **upper (efficacy) boundary** b_j, with $a_j < b_j$.

The basic decision rules for each treatment arm k:

- If $Z_{k,j} \leq a_j$: stop for **futility** (lack of benefit).
- If $Z_{k,j} \geq b_j$: stop for **efficacy** and claim success for that arm.
- If $a_j < Z_{k,j} < b_j$: continue recruitment to that arm (and control) into the next stage.

Different **boundary shapes** (Pocock, O'Brien–Fleming, triangular, Haybittle-Peto, fixed) specify how a_j and b_j behave over time:

- **Pocock:** relatively constant boundaries at each analysis → easier to stop early but slightly higher information needed overall.
- **O'Brien–Fleming:** very stringent early stopping, converging to the usual critical value (~1.96 one-sided) at the final analysis → conservative but popular.

`MAMS::mams()` lets you choose `ushape` and `lshape` from `"pocock"`, `"obf"`, `"triangular"` or `"fixed"`. `gsMAMS` uses SCPRT-based boundaries but the behaviour is similar: strong control of FWER with early stopping opportunities.

2.3.3 Extending to multiple arms

With multiple experimental arms vs. a shared control:

- each arm k has its own sequence $Z_{k,1}, \ldots, Z_{k,J}$, and
- these statistics are **correlated across arms**, because each comparison uses the same control patients.

This means we can't just treat arms as independent tests. The `MAMS` package deals with this by using a **generalised Dunnett test** and closed-testing arguments to ensure that the **familywise type I error rate** (probability of any false positive among arms) stays at your chosen α.

`gsMAMS` uses a related SCPRT-based approach to derive sample sizes and boundaries that respect FWER across arms and stages.

2.4 Example 1 – continuous endpoint MAMS design (TAILoR-inspired)

Our first hands-on example uses a **continuous outcome** and is based on the **TAILoR trial**, a phase II MAMS study which compared three doses of telmisartan (20, 40, 80 mg) to control for improvement in insulin resistance (HOMA-IR) in HIV-positive patients.

2.4.1 Trial context

- Patients: HIV-positive adults on stable combination antiretroviral therapy, at risk of insulin resistance.
- Arms:
  - Control: no telmisartan;
  - Three experimental doses: 20 mg, 40 mg, 80 mg daily.
  - Participants were initially randomised 1:1:1:1 to these four arms.
  - Primary outcome: change in **HOMA-IR** (continuous).
  - Objective: **identify the most promising dose**, if any.

The adaptive design had **two stages**:

1. **Stage I (dose-finding / activity stage)**

- Recruitment to all four arms continues until we have 24-week outcome data for approximately half of the maximum planned sample size, at which point the interim analysis is performed.
- An interim analysis compares each dose with control on the primary outcome.
- Depending on these results, one or more arms can be dropped; promising dose(s) continue to Stage II with control.

2. **Stage II**

- Only the selected promising dose(s) plus control continue recruitment and follow-up.
- The primary endpoint (change in HOMA-IR at 24 weeks) is re-assessed in this second stage, while patients are followed for up to 48 weeks for secondary outcomes.

This two-stage structure is exactly what our MAMS design functions are built to handle: several arms share a control, and the trial can **drop unpromising doses at the interim** while preserving error-rate control.

From a statistical viewpoint, this is a classic **4-arm, 2-stage MAMS design** with a continuous endpoint.

2.4.2 Effect sizes and error rates

In the `gsMAMS` vignette, the TAILoR example adopts standardized effect sizes:

- Clinically "interesting" effect (effective dose): $\delta^{(1)} = 0.545$,
- "Uninteresting/futile" effect (ineffective dose): $\delta^{(0)} = 0.178$,

with:

- one-sided FWER $\alpha = 0.05$,
- power $1 - \beta = 0.90$,
- 3 treatment arms $k = 3$,
- 2 stages with information fractions $0.5, 1$.

We'll mimic this specification with **two different R tools**.

2.4.3 Designing the trial with `MAMS`

First, we start with the `MAMS` package:

```
# install.packages("MAMS")

library(MAMS)

# TAILoR-like 3-arm, 2-stage design, continuous endpoint
tailor_mams <- mams(
  K      = 3,         # 3 doses / treatment arms
  J      = 2,         # 2 stages
  alpha  = 0.05,      # one-sided FWER
  power  = 0.90,      # power under least favourable configuration
  r      = 1:2,       # allocation ratios (stage 1 and 2)
  r0     = 1:2,       # ratios for control
  delta  = 0.545,     # interesting effect size
  delta0 = 0.178,     # uninteresting effect size
  sd     = 1,         # standardized SD
  ushape = "obf",     # O'Brien-Fleming-type efficacy boundary
  lshape = "fixed",   # fixed futility boundary
  lfix   = 0,         # Z <= 0 → drop arm for futility
  nstart = 30         # starting guess for stage 1 control size
)

tailor_mams
```

What this code tells `MAMS`:

- we want a 3-arm, 2-stage trial,
- controlling one-sided FWER at 5% and achieving 90% power to detect at least one arm with effect size 0.545,
- using O'Brien–Fleming-like efficacy boundaries and a fixed futility boundary at $Z = 0$.

The returned object has components:

```
tailor_mams$u       # Upper (efficacy) boundaries at stages 1 and 2
tailor_mams$l       # Lower (futility) boundaries
tailor_mams$n       # Stage 1 sample size in control
tailor_mams$N       # Maximum total sample size
tailor_mams$alpha   # FWER
```

From a **clinician's perspective**, you can read:

- "At stage 1 we recruit around `tailor_mams$n` patients to control and the same per treatment arm.
- We compute a Z-statistic for each dose vs. control.
  - If the Z statistics for a dose is $\leq$ lower boundary, that dose is dropped as futile.
  - If the Z statistics for any dose is $\geq$ upper boundary, we declare that dose promising enough to stop the trial for efficacy (and usually recommend it for phase III).
  - Otherwise, we continue to stage 2 with remaining arms."

2.4.4 Designing the same trial with `gsMAMS`

Now we repeat the design using the `gsMAMS` package, which directly implements **SCPRT-based group-sequential MAMS designs** and provides built-in tools for operating characteristics.

```
# install.packages("gsMAMS")

library(gsMAMS)

tailor_gs <- design_cont(
  delta0 = 0.178,    # uninteresting effect size
  delta1 = 0.545,    # interesting effect size
  alpha  = 0.05,     # one-sided FWER
  beta   = 0.10,     # 1 - power
  k      = 3,        # number of experimental arms
  frac   = c(0.5, 1) # information times for 2 stages
)

tailor_gs
```

`tailor_gs` is a list containing:

- cumulative **sample size per arm** at each stage,
- **maximum total sample size** for the trial,
- SCPRT **lower and upper boundaries**.

You can pull out the key bits as:

```r
# Sample sizes
tailor_gs[["Sample size"]]

# Boundary values
tailor_bounds <- tailor_gs[["Boundary values"]]
tailor_bounds
```

For these parameters, stage-wise boundaries are approximately:

- Stage 1: lower ≈ 0.006, upper ≈ 2.91
- Stage 2: lower ≈ 2.06, upper ≈ 2.06

Interpretation for clinicians:

- **First interim (after ~40 patients per arm):**
  - if an arm's $Z < 0.006 \rightarrow$ drop that dose for futility;
  - if an arm's $Z > 2.91 \rightarrow$ trial stops early, that dose is recommended;
  - otherwise $\rightarrow$ continue to stage 2 with remaining arms.
- **Second stage (full information):**
  - if any remaining arm's $Z > 2.06 \rightarrow$ declare success;
  - if all $Z < 2.06 \rightarrow$ no dose has sufficient evidence; trial declares no success.

2.4.5 A quick hands-on Z statistic example

Suppose at stage 1 we analyse HOMA-IR for the 80 mg arm vs. control and obtain:

- estimated mean difference (treatment – control) $\hat{\delta} = 0.50$ (higher is better),
- standard error $\mathrm{SE}(\hat{\delta}) = 0.18$.

The stage-1 Z statistic is:

```r
delta_hat <- 0.50
se_delta  <- 0.18

Z_stage1 <- delta_hat / se_delta
Z_stage1
```

We get $Z \approx 2.78$.

- Under the **MAMS** design, check whether 2.78 exceeds the stage-1 upper boundary `tailor_mams$u[1]`.
- Under the **gsMAMS** design, compare 2.78 against stage-1 upper bound (~2.91).

In this toy calculation, we'd be **very close** to an early stopping for efficacy under gsMAMS and would almost certainly continue the arm, while strongly considering it to be promising.

2.5 Example 2 – survival MAMS design (STAMPEDE-inspired)

Next we move to a **time-to-event** endpoint, which is most relevant for large oncology platforms like **STAMPEDE**. STAMPEDE uses **failure-free survival (FFS)** as an intermediate endpoint and overall survival as a final endpoint, and has evaluated combinations of androgen deprivation therapy (ADT) with docetaxel, abiraterone, radiotherapy and other agents.

We'll use a generic STAMPEDE-like scenario:

- Control median survival: 20 months.
- 4 treatment arms vs. control (K = 4).
- We want to detect a hazard ratio HR = 0.67 for at least one arm.
- Two looks at the data when 50% and 100% of the events have occurred.
- One-sided FWER 5%, power 90%.

2.5.1 Designing the survival MAMS trial with `gsMAMS`

```
library(gsMAMS)

stampede_like <- design_surv(
  m0    = 20,          # median survival in control (months)
  hr0   = 1,           # null HR (no effect)
  hr1   = 0.67,        # interesting HR for treatment arms
  ta    = 40,          # accrual time (months)
  tf    = 20,          # additional follow-up (months)
  alpha = 0.05,        # one-sided FWER
  beta  = 0.10,        # type II error
  k     = 4,           # number of treatment arms
  kappa = 1,           # shape parameter (1 = exponential)
  eta   = 0,           # loss-to-follow-up rate (0 = none)
  frac  = c(0.5, 1)    # information fractions for the 2 looks
)

stampede_like
```

This returns:

- **Cumulative number of events** at each stage (across all arms combined),
- **Maximum total number of events** and total sample size,
- **Lower and upper boundaries** at each stage.

```
# Events and total subjects
stampede_like[["Sample size"]]
stampede_like[["Total number of subjects required for the trial"]]

# SCPRT boundaries
stampede_surv_bounds <- stampede_like[["Boundary values"]]
stampede_surv_bounds
```

The boundaries are roughly:

- Stage 1: lower ≈ 0.075, upper ≈ 2.98

- Stage 2: lower ≈ 2.16, upper ≈ 2.16

Interpretation:

- Conduct the **first interim** after ~164 events.
- For each arm k, compute the log-rank Z statistic $Z_{k,1}$:
  - If $Z_{k,1} < 0.075$: drop arm k for futility.
  - If $Z_{k,1} > 2.98$: stop the trial and recommend the best arm.
  - Otherwise: continue to wait for more events (stage 2).

At **final analysis** (around 328 events):

- For remaining arms, compute $Z_{k,2}$:
  - If any $Z_{k,2} > 2.16$: declare that arm efficacious vs. control;
  - If all $Z_{k,2} \leq 2.16$: no arm shows sufficient evidence.

From a **platform perspective**, this is exactly the type of rule STAMPEDE uses on FFS to decide whether to continue, drop, or "graduate" regimens for further study (e.g., abiraterone + ADT showed a very large FFS effect with HR ≈ 0.29).

2.5.2 Operating characteristics under null and alternative

A nice feature of `gsMAMS` is that we can easily check the **operating characteristics** of our design.

```
# FWER and behaviour under global null
op_null <- op_fwer_surv(
  m0    = 20,
  alpha = 0.05,
  beta  = 0.10,
  p     = 4,           # number of treatment arms
  frac  = c(0.5, 1),
  hr0   = 1,           # null HR
  hr1   = 0.67,        # interesting HR
  nsim  = 10000,
  ta    = 40,
  tf    = 20,
  kappa = 1,
  eta   = 0,
  seed  = 1000
)

op_null

# Power and behaviour under alternative
# Usually the least-favourable configuration: only one arm has HR=0.67
op_alt <- op_power_surv(
  m0    = 20,
  alpha = 0.05,
  beta  = 0.10,
  p     = 4,
```

```
  frac  = c(0.5, 1),
  hr0   = 1,
  hr1   = 0.67,
  nsim  = 10000,
  ta    = 40,
  tf    = 20,
  kappa = 1,
  eta   = 0,
  seed  = 1000
)

op_alt
```

You'll see outputs like:

- overall FWER ≈ 0.05 with stage-wise contributions spread across looks,
- overall power ≈ 0.90,
- stopping probabilities at interim vs. final under null and alternative,
- average number of events and average trial duration under null and alternative.

These give both statisticians and clinicians a transparent view of how often the trial is expected to stop early and how many patients/events are typically required.

2.6 Example 3 – linking boundaries to real survival data (ovarian cancer)

So far, we've used real trial contexts but not actual patient-level data, because most MAMS platforms don't publish full datasets. To get a more tactile feel, we'll use the `ovarian` dataset from the **`survival`** package.

This dataset comes from a randomized trial of two treatments for ovarian cancer and includes:

- `futime`: survival or censoring time
- `fustat`: event indicator (1 = death, 0 = censored)
- `rx`: treatment group
- `age`, `resid.ds`, `ecog.ps`: covariates

2.6.1 Fitting a Cox model to estimate HR and Z

```
library(survival)

# Load the data
data(ovarian)
head(ovarian)
```

```
##   futime fustat     age resid.ds rx ecog.ps
## 1     59      1 72.3315        2  1       1
## 2    115      1 74.4932        2  1       1
## 3    156      1 66.4658        2  1       2
## 4    421      0 53.3644        2  2       1
## 5    431      1 50.3397        2  1       1
## 6    448      0 56.4301        1  1       2
```

```r
# Simple two-arm comparison using Cox PH
fit_ov <- coxph(Surv(futime, fustat) ~ rx, data = ovarian)
summary(fit_ov)
```

```
## Call:
## coxph(formula = Surv(futime, fustat) ~ rx, data = ovarian)
##
##   n= 26, number of events= 12
##
##       coef exp(coef) se(coef)      z Pr(>|z|)
## rx -0.5964    0.5508   0.5870 -1.016     0.31
##
##    exp(coef) exp(-coef) lower .95 upper .95
## rx    0.5508      1.816    0.1743      1.74
##
## Concordance= 0.608  (se = 0.07 )
## Likelihood ratio test= 1.05  on 1 df,   p=0.3
## Wald test            = 1.03  on 1 df,   p=0.3
## Score (logrank) test = 1.06  on 1 df,   p=0.3
```

```r
# Extract log-HR and its standard error
logHR    <- coef(fit_ov)["rx"]
se_logHR <- sqrt(vcov(fit_ov)["rx", "rx"])

# Z statistic for H0: HR = 1
Z_final <- logHR / se_logHR
Z_final
```

```
##        rx
## -1.015998
```

`Z_final` is the standard log-rank-type statistic you would compare to the **final-stage boundary** in a group-sequential (or MAMS) design.

2.6.2 Comparing to a MAMS survival boundary

Let's (purely for illustration) imagine that this ovarian trial was part of a 4-arm, 2-stage MAMS survival design like the STAMPEDE-inspired one we designed above, with final upper boundary around 2.16.

You can compare:

```r
# Suppose we re-use the stampede_like design above
stampede_surv_bounds <- stampede_like[["Boundary values"]]

upper_stage2 <- stampede_surv_bounds["Upper bound", "Stage 2"]
upper_stage2

-Z_final
-Z_final > upper_stage2
```

The negative sign is present since a negative Z-score indicates the treatment is beneficial.

Interpretation for clinicians:

- If `-Z_final > upper_stage2`, this arm would have **crossed the final efficacy boundary** in the MAMS design and been declared beneficial.
- If `-Z_final <= upper_stage2`, the evidence would be **insufficient** at the pre-specified error rates to declare success.

2.7 Putting it all together: A practical design workflow

2.7.1 Step 1 – define arms, endpoints, and clinically relevant effects

1. **List candidate arms**
   - Experimental regimens or doses.
   - Decide which are worth including given costs and logistical burden.
2. **Choose endpoints**
   - Final-stage primary endpoint (e.g., Overall Survival).
   - Intermediate endpoint for early decisions.
3. **Specify effect sizes**
   - For each arm type (drug, dose level), specify:
     - "interesting" effect δ that warrants success.
     - "uninteresting" effect δ_0 that you're happy to reject.

Use real datasets (such as `ovarian`, `colon` in `survival`) to check that your chosen effects are **plausible** in scale and variance.

2.7.2 Step 2 – choose error rates and stages

4. **Set one-sided FWER** (typically 2.5% or 5%) and power (80–90%).
5. **Pick number of stages and information times**, based on:
   - how quickly the endpoint accumulates,
   - resource constraints (how many interims a DMC can realistically handle).

For early-phase "screening" designs you might accept higher type I error and more stages; for phase III confirmatory platforms, you'll want stricter FWER and fewer looks.

2.7.3 Step 3 – design with R and iterate

6. **Use `MAMS` or `gsMAMS` to find boundaries and sample sizes**

   Examples:

```r
# Continuous TAILoR-style design
library(MAMS)
cont_design <- mams(
  K = 3, J = 2, alpha = 0.05, power = 0.90,
  r = 1:2, r0 = 1:2,
  delta  = 0.545, delta0 = 0.178, sd = 1,
  ushape = "obf", lshape = "fixed", lfix = 0, nstart = 30
)

# Survival STAMPEDE-style design
library(gsMAMS)
surv_design <- design_surv(
  m0 = 20, hr0 = 1, hr1 = 0.67,
  ta = 40, tf = 20,
  alpha = 0.05, beta = 0.10,
  k = 4, kappa = 1, eta = 0,
  frac = c(0.5, 1)
)
```

7. **Check operating characteristics** (FWER, power, expected sample size):

```r
# gsMAMS survival example
op_null  <- op_fwer_surv(m0=20, alpha=0.05, beta=0.10,
                         p=4, frac=c(0.5,1),
                         hr0=1, hr1=0.67,
                         nsim=10000, ta=40, tf=20,
                         kappa=1, eta=0, seed=12)

op_alt   <- op_power_surv(m0=20, alpha=0.05, beta=0.10,
                          p=4, frac=c(0.5,1),
                          hr0=1, hr1=0.67,
                          nsim=10000, ta=40, tf=20,
                          kappa=1, eta=0, seed=12)

op_null
op_alt
```

8. **Iterate with the clinical team**

   - Is the **maximum sample size** acceptable?
   - Are the **stopping rules** clinically sensible (e.g., do you need stricter futility stopping rules to limit exposure to toxic therapies)?
   - Do you want different allocation ratios (e.g., more patients on control, or on the most promising arms)?

2.7.4 Step 4 – document boundaries for the protocol

Finally, summarize the design for the protocol:

- A **table of information times**, events/patients per stage.
- For each stage:

  - the **futility Z-cutoff** for each arm,
  - the **efficacy Z-cutoff**,
  - and a short text description of what happens if a boundary is crossed.

2.8 Take-Aways

- MAMS designs **formalize** what many platforms are doing informally: multiple arms, staged decisions, and explicit control of error rates across arms and time.
- The key design choices are:
  - number of arms and stages,
  - clinically meaningful effect sizes,
  - FWER/power,
  - information times and allocation ratios.
- R packages `MAMS` and `gsMAMS` let you:
  - turn these choices into **concrete boundaries and sample sizes**,
  - check operating characteristics via simulation, and
  - connect those boundaries back to familiar analyses (hazard ratios, mean differences, Z-statistics).