Aligning the target product profile with your trial design - or avoiding statistically significant but clinically irrelevant effects

Kaspar Rufibach
Advanced Biostatistical Sciences, Merck KGaA, Darmstadt
Statisticians in Finnish Pharmaceutical Industry, 6th May 2025

Two definitions:

Target product profile (TPP):
Ideal version of what sponsor would like to claim in label.

Probability of success (POS):
P(beat effect of interest in clinical trial | averaged over prior evidence on effect).

Define TPP and keep it fixed.

Minimal TPP drives trial design.

Do not focus on effect at which we have 80% power when planning sample size.

Power trials such that statistically significant = clinically relevant.

Make clear what "success" is for overpowered trials.

If you want to compute probability of success - define success!

P(beat MDD) $\neq$ P(beat target).

AI Gobbledygook POS predictions: Agnostic to Phase 3 trial design and internal prior evidence.

You design trial with 80% power to detect HR = 0.75 and 2-sided $\alpha = 0.05$.

Trial reads out positive with $p = 0.05$.

What is your observed HR?

a) HR = 0.75,

b) HR = 0.818,

c) HR = 0.682?

You design trial with 80% power to detect HR = 0.75 and 2-sided $\alpha = 0.05$.

Trial reads out positive with $p = 0.05$.

What is your observed HR?

a) HR = 0.75,

b) HR = 0.818,

c) HR = 0.682?

Million dollar question:

Is HR = 0.818 clinically relevant?

Did you discuss this quantity during trial design?

Billion dollar question:

Is HR = 0.818 clinically relevant?

Did you discuss this quantity during trial design?

Hazard ratios and their p-values:

a) HR = 0.75 $\Rightarrow$ $p = 0.0051$,

b) HR = 0.818 $\Rightarrow$ $p = 0.05$,

c) HR = 0.682 $\Rightarrow$ $p = 0.0002$.

How do we compute $p = 0.05$?

Reject H_0 if $|\text{test statistic}| \geq z_{1-\alpha/2}$.

$$|\widehat{\Delta}/\mathrm{SE}(\widehat{\Delta})| \geq z_{1-\alpha/2} \Rightarrow \widehat{\Delta} = \pm z_{1-\alpha/2} \cdot \mathrm{SE}(\widehat{\Delta}).$$

Critical value $z_{1-\alpha/2}$ of hypothesis test on scale of interest (hazard ratio).

Minimal detectable difference.

Defines sample size!

Carroll (2009); Brock et al. (2015); Duquesne et al. (2020)


Give me a
break!

Target product profile (TPP)

Target product profile

- **Planning** and **decision-making** tool for therapeutic candidates.
- Ideal version of what sponsor would like to claim in label: What trial results will make a good drug in the marketplace?
- **Independent** of any trial design!
- But guides design, conduct, and analysis of clinical trials.
- Updated over time to reflect key changes in available treatments, health authority guidelines, payer policies, biomarker subgroups, etc.
- FDA draft guidance, U.S. Food and Drug Administration (2007):

  A TPP can be prepared by a sponsor and then shared with the appropriate FDA review staff to facilitate communication regarding a particular drug development program.

Target product profile

Target product profile

HR = 0.75

HR = 0.818

Match trial design to TPP

Minimal TPP drives ideal sample size!

Pick Phase 3 sample size such that

Statistically significant $\Leftrightarrow$ clinically relevant.

Does anyone care at which effect size you have 80% power?

How much power do we have to detect an effect equal to the MDD?

Derivation of sample size formula

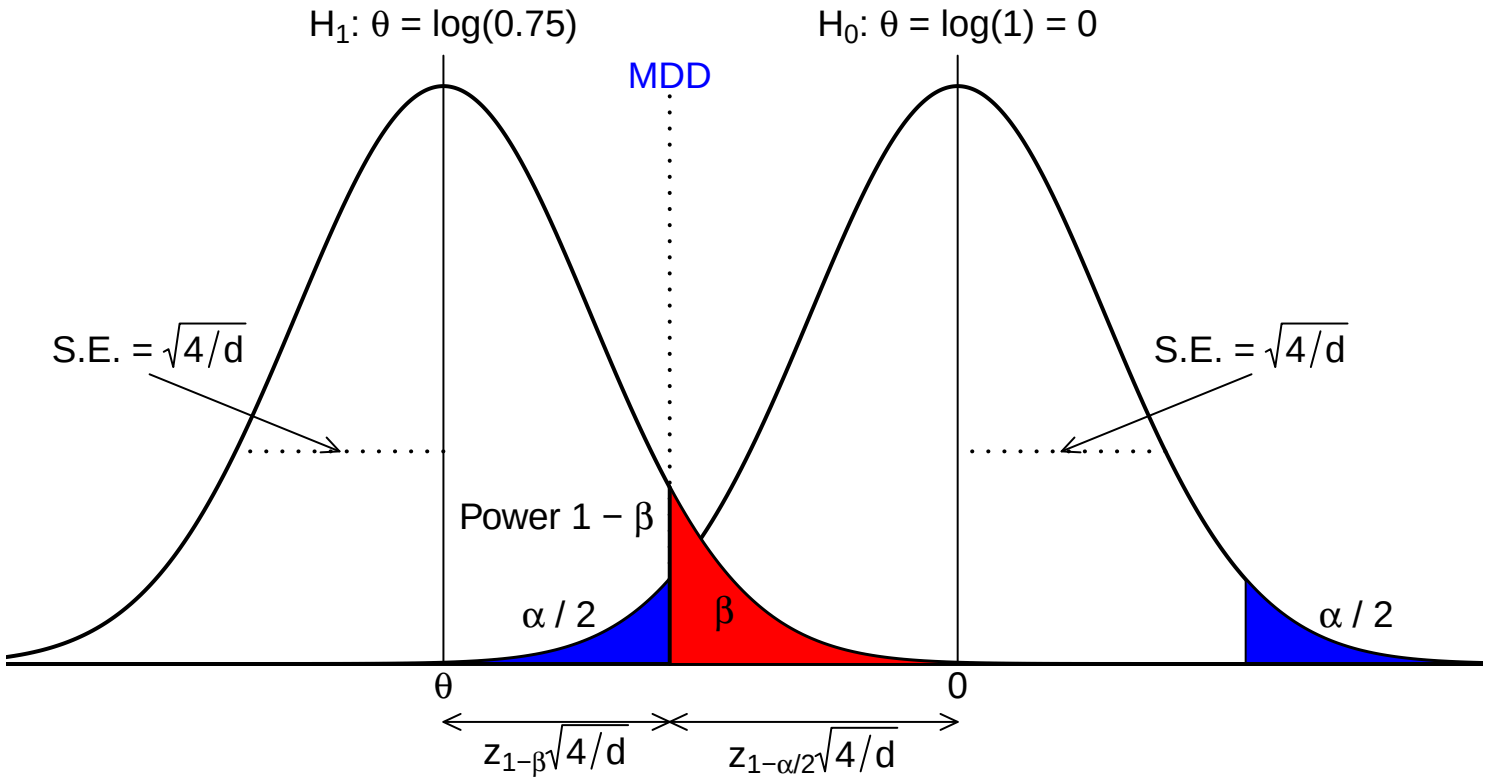

Illustration for time-to-event endpoint

```
> library(rpact)
```

```
> # necessary number of events to have 80% power
> ss <- getSampleSizeSurvival(sided = 2, alpha = 0.05, beta = 0.2, hazardRatio = 0.75)
> d <- ss$eventsPerStage
> d
          [,1]
[1,] 379.3517
```

```
> # MDD from rpact
> mdd <- as.vector(ss$criticalValuesEffectScaleLower)
> mdd
[1] 0.8177
```

```
> # MDD manually
> se <- sqrt(4 / d)
> exp(- qnorm(1 - ss$criticalValuesPValueScale / 2) * se)
        [,1]
[1,] 0.8177
```

What if my trial is overpowered?

What if my trial is overpowered?

For some reason, need **500** events instead of 380.

How does **MDD** change?

Number of events	MDD	Effect for 80% power
$d = 380$	0.818	0.750
$d = 500$	0.839	0.778

We are able to detect smaller effects.

Trial success as in TPP $\neq$ statistically significant!

How about interim analyses?

MDD smaller or larger than at final?

MDD $= \pm z_{1-\alpha/2} \cdot \mathrm{SE}(\widehat{\Delta})$**.**

α **and** $\mathrm{SE}(\widehat{\Delta})$ **change.**

Unless for very extreme scenarios MDD at interim larger than at final.

Illustration for every endpoint type

Global parameters:

- 1:1 randomization,
- 2-sided $\alpha = 0.05$,
- power = 80%,
- interim analysis after 70% of information using O'Brien-Fleming boundary.

	Binary	Continuous	Time-to-event
Effect size for 80% power	0.15	10	0.75
Endpoint-specific parameter	Baseline proportion = 0.45	Standard deviation = 24	
Number of patients / events	176 per arm	92 per arm	386 in total
MDD at interim	0.16	10.32	0.74
MDD at final analysis	0.11	7.09	0.82

Compare effect we power at to MDD at interim.

Match trial design to TPP

Target product profile

HR = 0.6

HR = 0.7

Plan a trial for this TPP

Minimal hazard ratio: 0.7 – match to MDD of trial $\Rightarrow$ **sample size computation with 50% power**.

```
> library(rpact)
> ss_tpp <- getSampleSizeSurvival(sided = 2, alpha = 0.05, beta = 0.5, hazardRatio = 0.7)
> d_min <- ceiling(ss_tpp$maxNumberOfEvents)
> d_min
[1] 121
```

Compute hazard ratio that corresponds to $d = 121$ events with 80% power. Invert sample size formula to get:

$$\text{hazard ratio} \quad = \quad \exp\Big(\frac{-2(z_{1-\alpha/2} + z_{1-\beta})}{\sqrt{d}}\Big).$$

```
> hr_tpp <- exp( - 2 * (qnorm(1 - 0.05 / 2) + qnorm(1 - 0.2)) / sqrt(d_min))
> hr_tpp
[1] 0.6008685
```

0.601 should then $\approx$ match target hazard ratio 0.6.

Want to compute probability of success?

First define success!

P(beat MDD)?

P(beat target)?

Assurance computation

```
> library(bpp)
> # assume design prior
> hr0 <- 0.5
> sd0 <- sqrt(4 / 25)
```

```
> # assurance to beat MDD
> ass1mdd <- bpp_t2e(prior = "normal", successHR = 0.7, d = d_min,
+                    priorHR = hr0, priorsigma = sd0)
> ass1mdd
[1] 0.7780972
```

```
> # assurance to beat target
> ass1target <- bpp_t2e(prior = "normal", successHR = hr_tpp, d = d_min,
+                       priorHR = hr0, priorsigma = sd0)
> ass1target
[1] 0.6621144
```

```
> # difference
> ass1mdd - ass1target
[1] 0.1159827
```

Define TPP and keep it fixed.

Minimal TPP drives trial design.

Do not focus on effect at which we have 80% power when planning sample size.

Power trials such that statistically significant = clinically relevant.

Make clear what "success" is for overpowered trials.

If you want to compute probability of success - define success!

P(beat MDD) $\neq$ P(beat target).

AI Gobbledygook POS predictions: Agnostic to Phase 3 trial design and internal prior evidence.

References

Brock, T. C. M., Hammers-Wirtz, M., Hommen, U., Preuss, T. G., Ratte, H.-T., Roessink, I., Strauss, T., and Van den Brink, P. J. (2015). The minimum detectable difference (mdd) and the interpretation of treatment-related effects of pesticides in experimental ecosystems. *Environmental Science and Pollution Research*, **22**(2), 1160–1174.

Carroll, K. J. (2009). Back to basics: explaining sample size in outcome trials, are statisticians doing a thorough job? **8**(4), 333–345.

Duquesne, S., Alalouni, U., Gräff, T., Frische, T., Pieper, S., Egerer, S., Gergs, R., and Wogram, J. (2020). Better define beta–optimizing mdd (minimum detectable difference) when interpreting treatment-related effects of pesticides in semi-field and field studies. *Environmental Science and Pollution Research*, **27**(8), 8814–8821.

U.S. Food and Drug Administration (2007). *Guidance for Industry and Review Staff Target Product Profile - A Strategic Development Process Tool*. Draft guidance.

Thank you for your attention.

kaspar.rufibach@merckgroup.com

Slides can be downloaded on

www.kasparrufibach.ch

R version and packages used to generate these slides:

R version: R version 4.4.3 (2025-02-28 ucrt)

Base packages: stats / graphics / grDevices / utils / datasets / methods / base

Other packages: bpp / mvtnorm / rpact / reporttools / xtable

This document was generated on 2025-05-04 at 14:44:10.