Paper AS17

Crafting a recipe for success : Aligning Target Product Profiles and study design to ensure statistical significance translates to clinical relevance and role of DGF

Soumitra Kar, Novo Nordisk GBS, Bangalore, India

ABSTRACT

Target Product Profile (TPP) serves as a strategic planning tool that outlines the desired characteristics and specifications of a drug candidate, defining the claim the sponsor aims to include on product label. Aligning TPP with study design is essential to ensure that statistical significance leads to clinical relevance.

Role of TPP, in conjunction with decision-guiding framework (DGF) is explored to facilitate actionable decision-making beyond the traditional focus on statistical significance.

Utilising examples from cardiovascular domain, particularly time-to-event endpoints it is demonstrated how smaller, shorter studies can expedite decision-making processes.

Additionally, the concept of probability of success is examined, highlighting the necessity of precisely defining the intended effect - whether it is the minimal detectable difference or the target difference.
Insights gained from this emphasizes the importance of optimizing study designs that prioritize clinically relevant results, ultimately enhancing the likelihood of successful drug development.

Keywords : Target product profile (TPP), decision guiding framework (DGF), statistical significance, clinical relevance, minimal detectable difference (MDD), minimum value (MV), target value (TV).

INTRODUCTION

Clinical research aims not only to detect whether an investigational therapy produces a statistically significant effect, but also to establish that the effect is meaningful for patients, clinicians, payers, and commercial stakeholders. However, statistical significance and clinical relevance are distinct concepts: a result can be unlikely due to chance yet still be too small to change clinical practice or justify the costs and risks of bringing a product to market. This tension is especially important across the drug development process, where Phase 2 readouts are used to decide whether to progress to larger and more costly Phase 3 studies. Without explicit alignment between the study design and the intended Target Product Profile (TPP), sponsors might risk advancing candidates that will be statistically positive but clinically marginal, or conversely terminating promising programs because the study was not designed to detect effects that matter.

To address this gap, integrating a pre-specified Target Product Profile with quantitative decision rules provides a transparent, operational pathway from early signals to later-phase commitments. The TPP defines the product attributes required at launch to ensure clinical utility and commercial viability covering indication, dosing, efficacy, safety, use in specific populations, and health-economic considerations and thereby sets the criteria for what constitutes a clinically meaningful effect. Embedding these clinically driven thresholds into study planning ensures that sample size calculations, choice of primary and secondary endpoints, and statistical power are all focused on detecting effects that meet the TPP, not merely the smallest detectable difference.

A Decision Guiding Framework (DGF) operationalizes this linkage by translating TPP-derived targets into probabilistic decision rules at interim or pre-lock readouts. The DGF uses key inputs, a Target Value (TV) representing the lowest effect size aligned with an attractive business case, a Minimum Value (MV representing the smallest effect with clinical or commercial value), acceptable false-stop and false-go risks, observed variability, and the planned sample size to produce stop/go/consider boundaries. For example, in a Phase 2 setting using a biomarker such as Biomarker XX, pre-specified treatment ratios relative to placebo (e.g., TV = 0.70, MV = 0.90), together with defined false-stop and false-go probabilities, yield quantitative decision thresholds that guide whether a program should progress, be halted, or require further deliberation supported by secondary endpoints (e.g., GLS, 6MWT, hsCRP, imaging parameters) and commercial input. This structured approach also clarifies how uncertainty and variability (e.g., high coefficient of variation) affect decision confidence and the risk profile of advancing a candidate.

For event-driven Phase 3 endpoints such as time-to-first major adverse cardiovascular events (MACE), alignment between study power and the TPP is equally critical. Calculations of the minimal detectable difference (MDD), required number of events, and the hazard ratio that a study is powered to detect must be set with the TPP's clinically

meaningful threshold in mind. Otherwise, studies may achieve statistical significance for hazard ratios that fall short of the pre-specified minimal clinically important difference (MCID), producing results that are difficult to interpret from a clinical or commercial perspective.

In summary, combining TPP-driven clinical thresholds with rigorous statistical design and an explicit Decision Guiding Framework enhances that statistically significant findings will be clinically relevant and actionable. The scenarios will be illustrated through an application that integrates biomarker-based Phase 2 readouts with event-driven Phase 3 planning, accompanied by practical guidance for managing "consider" zones in which go/no-go decisions are informed by secondary endpoints, safety data, and competitive intelligence.

BACKGROUND

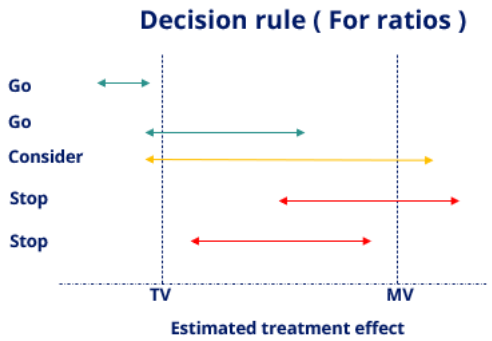
Cardiomyopathy remains a leading cause of morbidity and mortality worldwide, and development of novel therapeutics requires robust demonstration of both safety and clinically meaningful efficacy across the drug development continuum. Lets imagine a scenario where Company X has advanced a small-molecule candidate for cardiomyopathy through a Phase 1 clinical study in which four dose levels (10–100 mg/kg) were well tolerated and no safety signals were observed. Pharmacodynamic assessment using global longitudinal strain (GLS), a sensitive echocardiographic marker of myocardial function, suggested a potential beneficial effect on cardiac performance.

Given the promising safety and exploratory efficacy signals, progression to a randomized Phase 2 study has been planned using Biomarker XX as the primary biomarker endpoint. Biomarker XX is an established biomarker for diagnosis and prognosis in heart failure and has been used widely as an intermediate marker to inform later-phase development, since changes in Biomarker XX are often associated with clinical outcomes. Phase 3 planning will be contingent on the Phase 2 results; therefore, a structured decision-making approach is required to translate Phase 2 evidence into an informed go/no-go decision for large-scale outcome studies.

To support this decision process prior to database lock, a Decision Guiding Framework (DGF) is proposed. The DGF formalizes pre-specified, clinically relevant thresholds derived in-line with the Target Product Profile (TPP) , including a Target Value (TV) representing an effect size aligned with a commercially and clinically attractive outcome and a Minimum Value (MV) representing the smallest effect with potential clinical or commercial value and integrates these with acceptable false-stop and false-go risk tolerances, observed biomarker variability, and planned sample size. The framework produces probabilistic stop/go/consider boundaries for the Phase 2 “Biomarker XX” readout, enabling transparent, quantitative guidance on whether to advance to Phase 3, halt development, or convene a multidisciplinary review when results fall in an intermediate “consider” zone where supplementary evidence (for example, GLS, 6-minute walk distance, hsCRP, imaging endpoints, safety findings, and competitive intelligence) should be evaluated.

DECISION GUIDING FRAMEWORK (DGF) and Target product profile (TPP) :

Input parameter	Description
Target value (TV)	Lowest treatment effect according to guided decision where the business case is attractive in-line with TPP. This shows the goal sponsor would want to achieve to maximise the vlaue and provide transformational benefit.
Minimum value (MV)	At least the lowest treatment effect with clinical/commercial value. Minimally viable benefit that is expected to address an unmet need and change clinical practice. This refers to the effect clinicians would like to observe to switch to the drug.
False stop risk	Risk of stop if true treatment effect is at TV (e.g. 10%)
False go risk	Risk of go if true treatment effect is at MV (e.g. 15%)



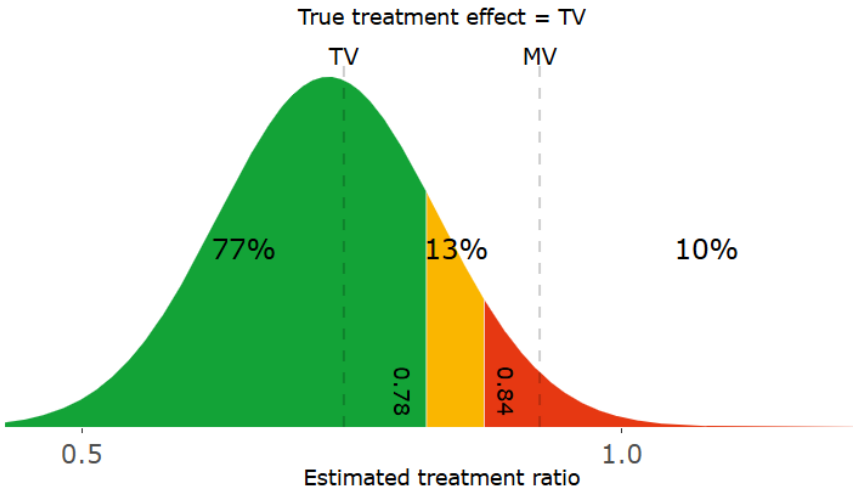
Interpretation / Decision rule as follows :
Stop, if there is ≥ 90 % probability that the true treatment effect is above TV
Else, Go, if there is ≥ 85 % probability that the true treatment effect is at least lesser than MV
Else, Consider.

Let us take an example of a DGF to understand in more details.

"Biomarker XX" **- DGF guiding Phase 3 decision based on phase 2 results**

	Justification/Assumptions
Parameter: **Ratio of** "Biomarker XX" **from baseline to 64 weeks for 100mg/kg vs placebo**	Well established biomarker for diagnosis and prognosis of heart failure. Competitors demonstrating an effect on survival have also demonstrated an effect on change in "Biomarker XX" versus placebo A consistent correlation between change in "Biomarker XX" and hard outcomes has not yet been established in internal or external resources.
Target Value: **0.7**	30% is considered progression threshold in disease area *(Based on literature, competitive intelligence and Commercial perspective)*
Minimum Value: **0.9**	SME input: • Anticipate laboratory preceding clinical/functional changes hence should demonstrate 'early' improvement
Risk setting: **False Stop: 10 %** **False Go: 15 %**	
Variation: **<Specify observed variation>**	CV = 80% (Based on variation at week 64 observed data (For around 25% data))
Sample size: **<Specify actual number of participants>**	N = 50 per arm

Probability of Go/Consider/Stop zone :



R code : For calculating cutoff points

```
> ## Calculation for DGF Zones for Phuse APAC
> (sigma_square <- log(0.8**2+1)) # Transforming CV
[1] 0.4946962
>
> # False stop (10%) : Cutoff calculation [ Under TV ]
> exp(qnorm(0.9)*sqrt(sigma_square*(2/50))+log(0.7)) # Assuming mean ratio=0.7 (equal to TV), Each arm=50 participants, under which value 90% of values lie ?
[1] 0.8382826
> round(exp(qnorm(0.9)*sqrt(sigma_square*(2/50))+log(0.7)),2)
[1] 0.84
>
> # False go (15%) : Cutoff calculation [ Under MV ]
> exp(-qnorm(0.85)*sqrt(sigma_square*(2/50))+log(0.9)) # Assuming mean ratio=0.9 (equal to MV), Each arm=50 participants, under which value 85% of values lie ?
[1] 0.7779019
> round(exp(-qnorm(0.85)*sqrt(sigma_square*(2/50))+log(0.9)),2)
[1] 0.78
```

Probability of recommended decisions

True treatment effect	Scenario	Go (%)	Consider (%)	Stop (%)
0.7	Target value	77.3	12.7	10.0
0.9	Minimum value	15.0	15.7	69.3

R code : For calculating probability of each zone under TV and MV

```
> # Under TV
> round(plnorm(0.7779019,log(0.7),sqrt(sigma_square*2/50)) * 100,1) # Prob of Go Zone
[1] 77.3
> round((plnorm(  0.8382826,log(0.7),sqrt(sigma_square*2/50))-plnorm(0.7779019,log(0.7),sqrt(sigma_square*2/50)))*100,1)  # Prob of Consider Zone
[1] 12.7
> round((1-plnorm( 0.8382826,log(0.7),sqrt(sigma_square*2/50)))*100,1) # Prob of Stop Zone
[1] 10
> # Under MV
> round(plnorm(0.7779019,log(0.9),sqrt(sigma_square*2/50)) * 100,1) # Prob of Go Zone
[1] 15
> round((plnorm(  0.8382826,log(0.9),sqrt(sigma_square*2/50))-plnorm(0.7779019,log(0.9),sqrt(sigma_square*2/50)))*100,1)  # Prob of Consider Zone
[1] 15.7
> round((1-plnorm( 0.8382826,log(0.9),sqrt(sigma_square*2/50)))*100,1) # Prob of Stop Zone
[1] 69.3
```

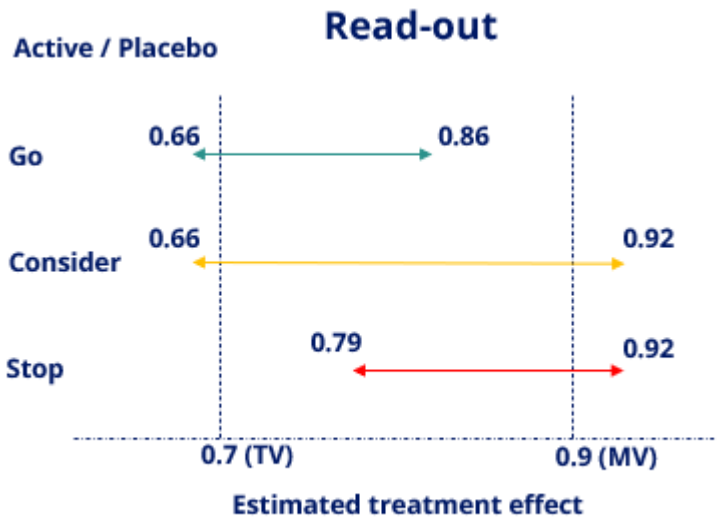
Hence, decisions would be as follows.

Stop, if treatment effect is < 16 (% decrease), corresponding to treatment ratio of 0.84

Go, if treatment effect is ≥ 22 (% decrease), corresponding to treatment ratio of 0.78.

Suppose, the analysis uses the log-transformed ratio from baseline to week 64, with treatment and the stratification factor included as categorical factors and the baseline log-transformed “Biomarker XX” value included as a covariate. Using the final imputed dataset, the confidence limits can be derived as follows to define the decision boundaries.

Example of readout based on DGF rules :



As, false stop probability has been set at 10 % and false go probability has been set at 15%, the limits can be found as one-sided lower 10% limit for 90% CI (same as lower confidence limit for 2 sided 80% CI) and one sided upper 15% limit for 85% CI (same as upper confidence limit for 2 sided 70% CI)

Sample R code to find limits for read-out :

```
# summarise based on final model ------------------------------------------
# Use mice for treatment ratio and difference

dif_sum <- map(final,
               ~lm(LOGCHG ~ TRTP + V200 + STRATAR, data = .x)) %>%
  pool(rule = "rubin1987", dfcom = ndof)

# get one-sided lower 10% limit for 90% CI ~ lower confidence limit for 2 sided 80% CI

dif_ll <- dif_sum %>%
  tidy(conf.int = TRUE, conf.level = 0.8) %>%
  filter(grepl("Treatment", term)) %>%
  mutate(contrast = paste(str_remove(term, "TRTP"), "/", "Placebo"),
         ratio = exp(estimate),
         SE = std.error*exp(estimate),
         "{lcl_nm}" := exp(conf.low),
         "{ucl_nm}" := exp(conf.high)) %>%
  select(contrast, all_of(c(lcl_nm)))

# get one-sided upper 15% limit for 85% CI ~ upper confidence limit for 2 sided 70% CI

dif_ul <- dif_sum %>%
  tidy(conf.int = TRUE, conf.level = 0.7) %>%
  filter(grepl("Treatment", term)) %>%
  mutate(contrast = paste(str_remove(term, "TRTP"), "/", "Placebo"),
         ratio = exp(estimate),
         SE = std.error*exp(estimate),
         "{lcl_nm}" := exp(conf.low),
         "{ucl_nm}" := exp(conf.high)) %>%
  select(contrast, ratio, SE, df, all_of(c(ucl_nm)), p.value) %>%
  left_join(dif_ll, by = 'contrast')
```

If the result falls in consider zone, then

- Stakeholders such as Portfolio manager, M&S, Safety, Commercial, Competitive intelligence should take informed decision based on pre-set decision criteria on other parameters/secondary endpoints. For example, hsCRP, 6MWT, MRI/Echo parameters (ECV/GLS etc) and take a decision based on historical data.
- This approach can also inform portfolio level decisions. For example, when resources are limited, a company may opt not to assume the additional risk of advancing a compound in the Consider zone and instead allocate resources to candidates with a clear positive decision.

Target product profile :

A set of specifications that define the key product attributes required at launch to ensure commercial viability and market appeal. Target product profile sets the direction for clinical programme. It acts as a guiding document for regulatory activities, CMC and product supply planning as well as patient access initiatives.

TPP contains information on Indications and usage, Dosing and administration, Dosage forms and strengths, Clinical efficacy/pharmacology, Safety and tolerability, Storage, Use in specific populations, Health economics and PROs

Suppose, in a Phase 3 clinical study the primary endpoint is time to first occurrence of a 3-component MACE (major adverse cardiovascular events) composite: cardiovascular death, non-fatal myocardial infarction, or non-fatal stroke. The study is planned using an unstratified Cox proportional hazards model, with 90% power to demonstrate superiority at a one-sided alpha of 2.5% for the primary endpoint. Randomisation is 1:1 and, assuming a true hazard ratio (HR) of 0.70, the required number of primary endpoint events is 331 (see accompanying R code for sample size calculations).

$$Number\ of\ events = E = \frac{(Z_{1-\beta} + Z_{1-\frac{\alpha}{2}})^2}{p_A * p_B * log(HR)^2}$$
$$p_A = p_B = 1/2$$

R code for sample size :

```
> Z_alpha <- qnorm(1-0.05/2)  # ~1.96
> Z_beta <- qnorm(0.90)    # ~1.28
> HR = 0.7
> p1 = 1/2  # 1:1 randomization
> p2=1-p1
> event_required <- (((Z_alpha + Z_beta)/log(HR) )**2)*(1/(p1*p2))
> event_required # 331 events
[1] 330.3779
```

At read-out the study met the statistical significance criterion (two-sided p < 0.05), but the observed HR was 0.806. This observed HR represents the most extreme value that would still produce statistical significance in this study design; we refer to it here as the minimal detectable difference (MDD). Because the MDD is generally not reported in the protocol, clinicians and patients reading the protocol might incorrectly assume that success implies observing the originally assumed HR of 0.70. That assumption can be misleading.

To actually observe an HR of 0.70 given the observed data would require much stronger evidence (for example, a p-value around 0.0012), so it can be disappointing if the observed HR of 0.806 falls short of what is considered clinically meaningful or competitive. In other words, an observed HR between 0.70 and 0.806 could be statistically significant but of marginal clinical importance — and in extreme cases, not clinically relevant at all.

This highlights two important considerations when designing a study :

- Are the HR assumptions used for powering the study aligned with the minimally clinically important difference (MCID)?
- Should we power the study to have 90% probability of detecting effect sizes that are clinically relevant, rather than simply powering to detect a potentially optimistic target HR?

Our ultimate goal is to design studies that reliably detect effects that are both statistically significant and clinically meaningful.

How can we calculate the MDD ?

Let, H_0 : HR ≥1 *ag.* H_1 *: HR < 1*

$\theta = log(HR)$

$\hat{\theta} = log(HR)$ *follows approximately* $N(\theta, \frac{4}{d})$

SE $(\hat{\theta}) = \sqrt{\frac{4}{d}}$

$$\frac{\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}}{\boldsymbol{SE}(\hat{\boldsymbol{\theta}})} \sim \boldsymbol{N}(\mathbf{0}, \mathbf{1})$$

Hence, the minimal detectable difference = - $Z_{1-\frac{\alpha}{2}}$ * SE($\hat{\theta}$)

R code for calculating MDD :

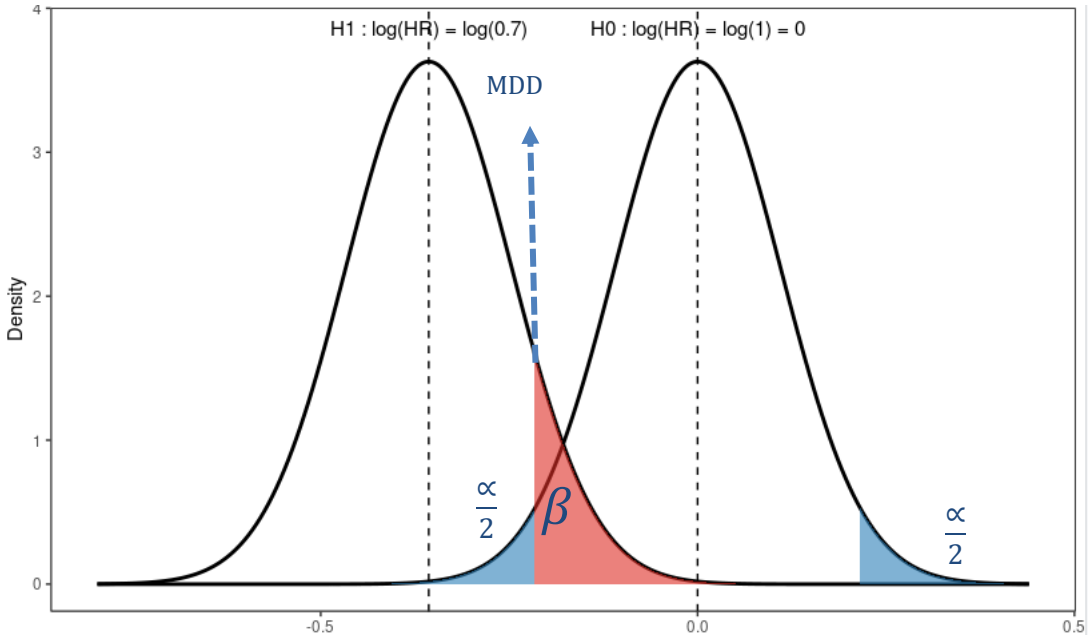
```
> round(exp(-qnorm(1 - 0.05 / 2) * sqrt(4 / 331)),3)
[1] 0.806
```

R code for calculating MDD using Rpact :

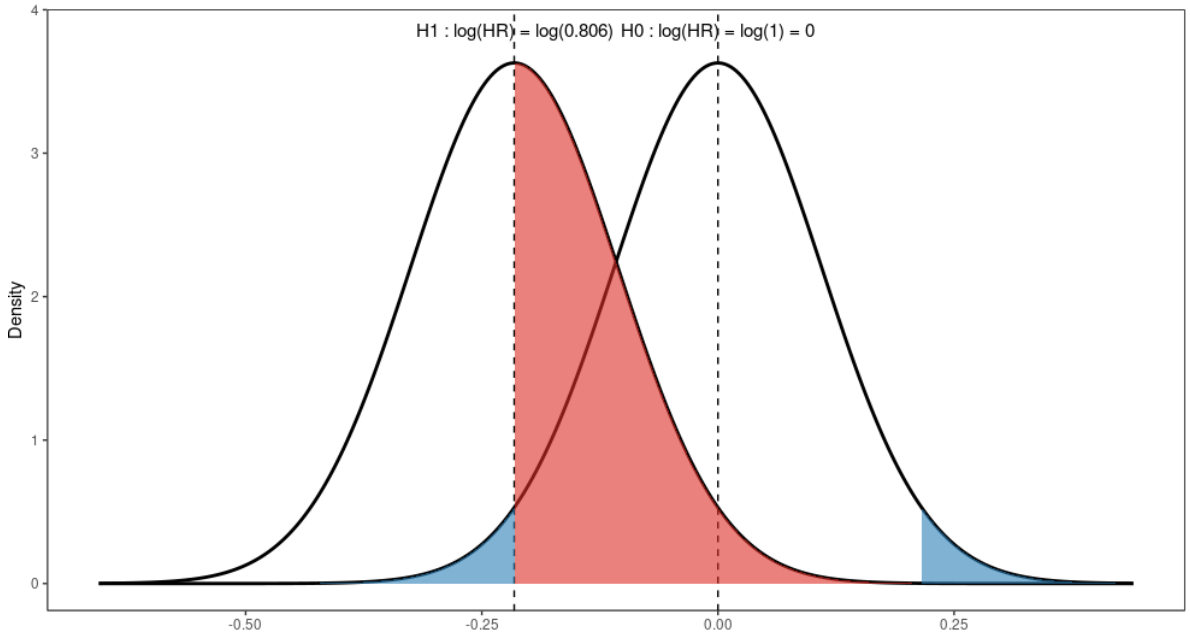
```
> # Sample size
> library(rpact)
> samplesize <- getSampleSizeSurvival(sided = 2, alpha = 0.05, beta = 1-0.9, hazardRatio = 0.7)
> samplesize$eventsPerStage
         [,1]
[1,] 330.3779
>
> # MDD from rpact
> mdd <- as.vector(samplesize$criticalValuesEffectScaleLower)
> mdd
[1] 0.8060081
```

Matching study design with Target product profile :

Below figure illustrates the sampling distributions under the null and alternative hypotheses and marks the minimal detectable difference (MDD)



If we centre the alternative distribution at the MDD, the shaded area beyond the statistical cut-off corresponds to 50% power, by definition the effect size at which the study has a 50% chance of producing a statistically significant result.



For this example, we set an aspirational TPP of HR = 0.70 and a base (minimal acceptable) TPP of HR = 0.80. We therefore take the base TPP (HR = 0.80) as the MDD for the design and determine the event count required to achieve 50% power for that effect size. Using standard survival sample-size calculations in R, 309 events are required to obtain 50% power when the true HR = 0.80.

Sample size calculation using R :

```
> samplesize_new <- getSampleSizeSurvival(sided = 2, alpha = 0.05, beta = 1-0.5, hazardRatio = 0.8)
> samplesize_new$maxNumberOfEvents
[1] 308.594
```

Holding the event count fixed at 309, we then ask: what hazard ratio would give 90% power with this number of events? Using the same sample-size routine in R but solving for the hazard ratio that attains 90% power at 309 events yields a hazard ratio close to the aspirational TPP (HR ≈ 0.70). Thus, with 309 events the study has approximately 50% power to detect the minimally acceptable effect (HR = 0.80) and approximately 90% power to detect the aspirational effect (HR ≈ 0.70).

R code :

```
> exp(-2 * (qnorm(1-0.05/2) + qnorm(1-0.1)) / sqrt (309))
[1] 0.691559
```

$$HR = \exp\left(\frac{-2\left(Z_{1-\beta} + Z_{1-\frac{\alpha}{2}}\right)}{\sqrt{E}}\right)$$

CONCLUSION

Define clinical relevance before statistical success: Pre-specify the magnitude of benefit that would be considered meaningful for patients and the business (TPP-derived Target and Minimum Values). This anchors all downstream decisions and prevents over-reliance on p-values alone.

Use a Decision Guiding Framework (DGF): A probabilistic DGF that incorporates Target Value, Minimum Value, acceptable false-stop/false-go risks, observed variability, and sample size provides transparent, quantitative guidance at interim or pre-lock readouts to decide Go / Stop / Consider.

Power studies for clinically meaningful effects: Design studies so they have adequate power to detect effects that meet the TPP (not merely the smallest statistically detectable difference). Align event counts, hazard-ratio assumptions, and MDD calculations with the TPP to avoid statistically significant but clinically irrelevant outcomes.

Plan for the “consider” zone: Predefine secondary endpoints and decision criteria (biomarkers, functional measures, imaging, safety, commercial context) to inform decisions when primary results fall between Target and Minimum values.

Interpret results in context: Combine effect size, confidence intervals, safety, cost, and real-world applicability. Even statistically significant results require interpretation against the TPP, clinical meaningfulness, and competitive landscape to be actionable.

Better portfolio decisions and patient outcomes: Applying these principles reduces the chance of advancing marginal candidates, improves resource allocation across programs, and increases the likelihood that positive studies translate into meaningful clinical benefit.

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.
- Kaspar Rufibach (2025, May 6), Statisticians in Finnish Pharmaceutical Industry webinar: “Probability of success”. YouTube : https://www.youtube.com/watch?v=6Rmd1nQnLb4
- Sharma H. Statistical significance or clinical significance? A researcher's dilemma for appropriate interpretation of research results. Saudi J Anaesth. 2021 Oct-Dec;15(4):431-434. doi: 10.4103/sja.sja_158_21. Epub 2021 Sep 2. PMID: 34658732; PMCID: PMC8477766.
- Carroll, K. J. (2009). Back to basics: explaining sample size in outcome trials, are statisticians doing a thorough job? 8(4), 333–345
- U.S. Food and Drug Administration (2007). Guidance for Industry and Review Staff Target Product Profile - A Strategic Development Process Tool. Draft guidance

DISCLAIMER :

Views expressed in the presentation by the presenter are the views, thoughts, and opinions belonging solely to the presenter, and not necessarily to the presenter's employer, organization or other group or individual.
Any data if used is from readily available package in CRAN or simulated. No data or confidential information belonging to Novo Nordisk has been used in this presentation.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Soumitra Kar
Novo Nordisk GBS
DD-16,Narayantala East, Baguiati
Kolkata / Postcode : 700159
Work Phone: +91-7980623725
Email: zsuk@novonordisk.com