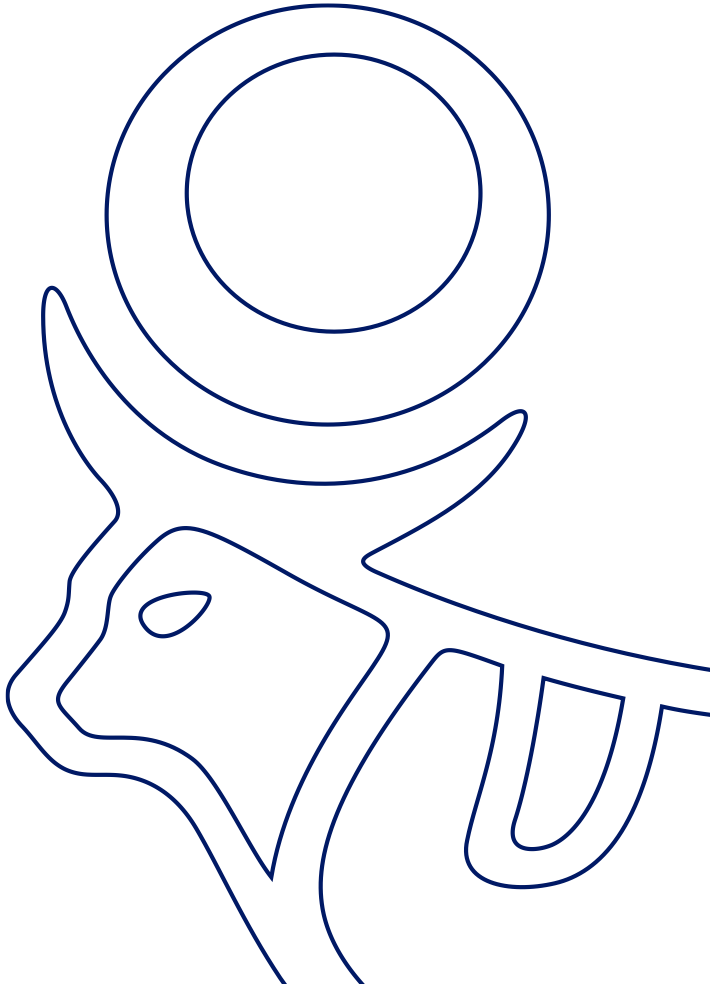
phuse·connect
connect·share·advance
APAC 2026
The Clinical Data Science Conference
19–21 February
Novotel Hyderabad International Convention Centre

novo nordisk®

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

Statistical significance ≠ Clinical significance (not always)

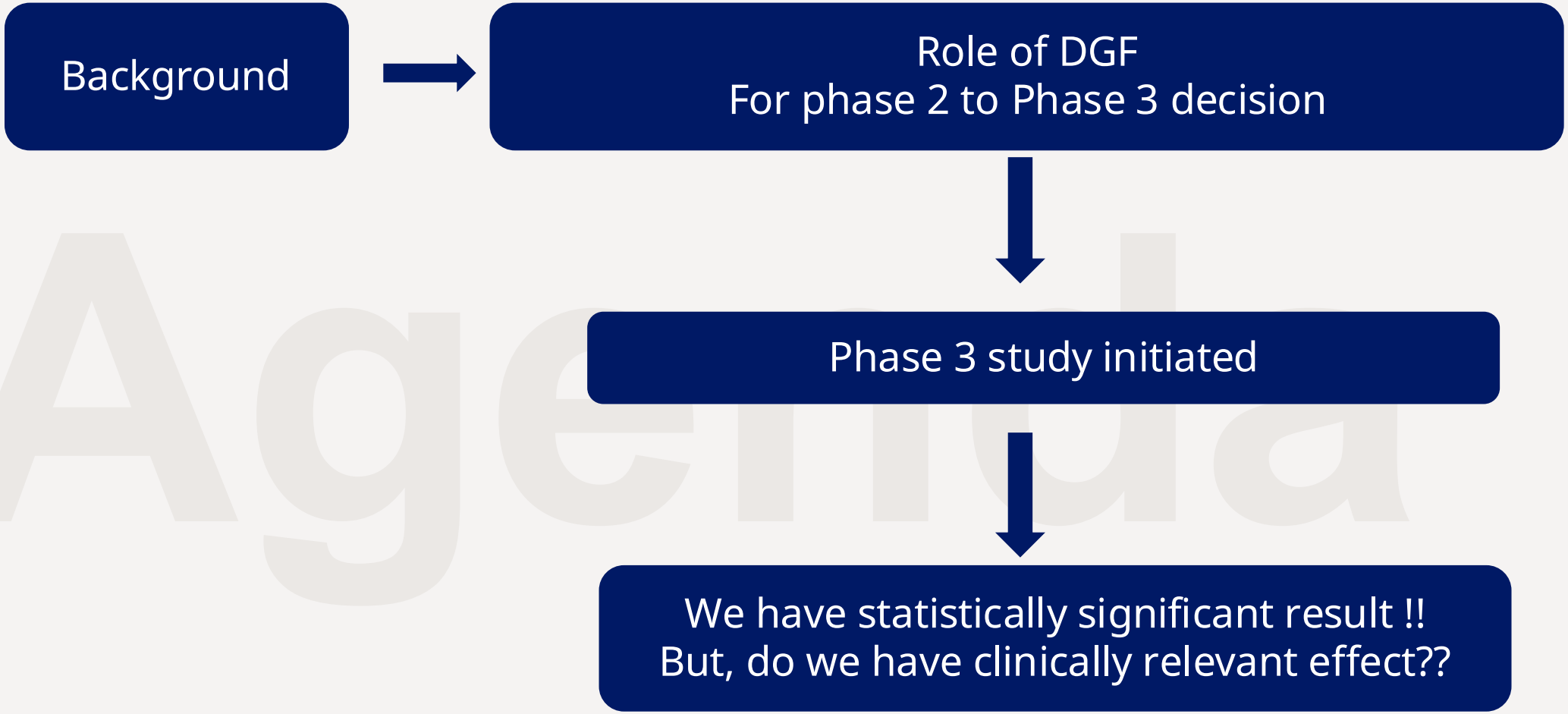


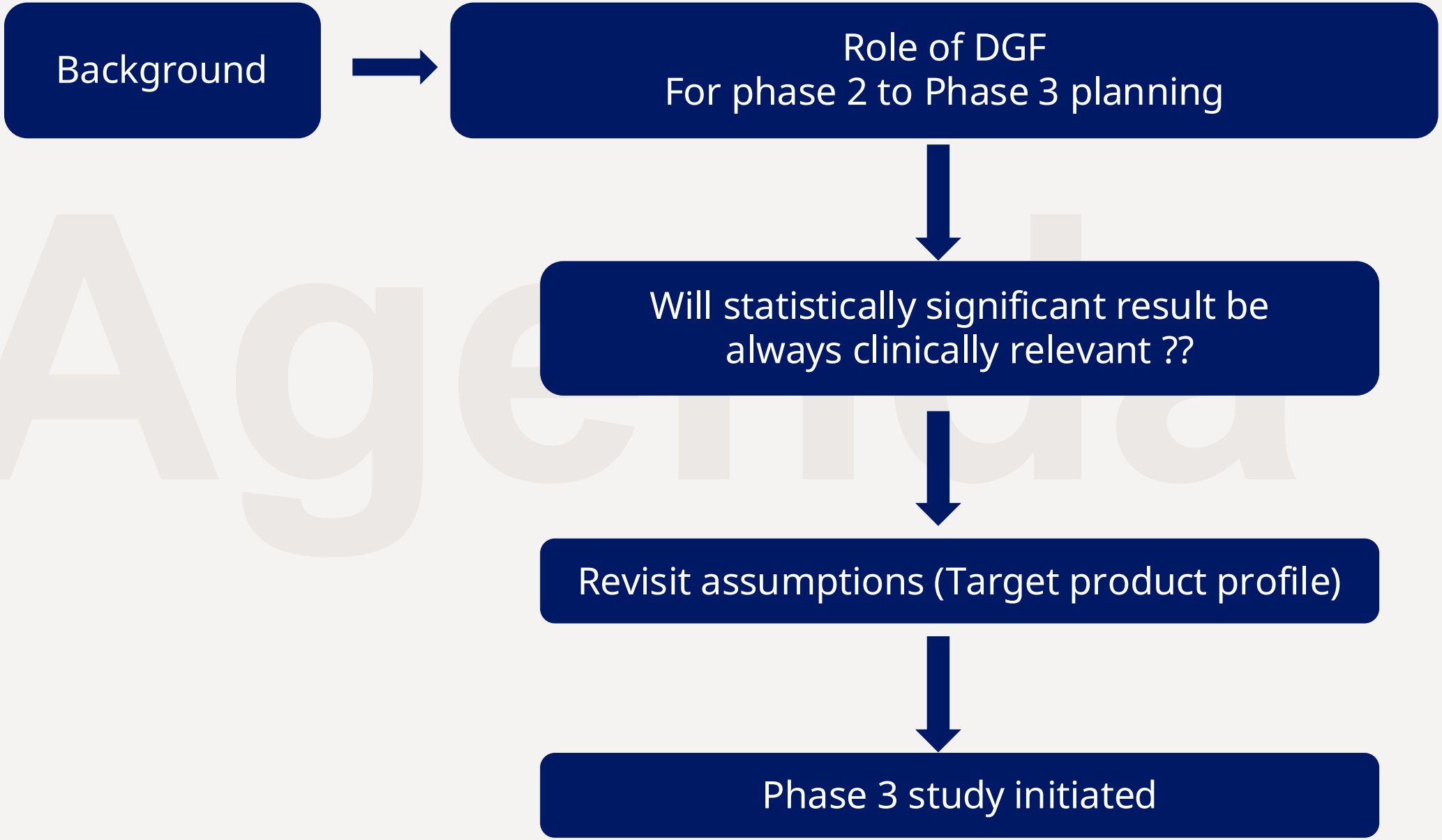
Soumitra Kar, Statistical specialist
Novo Nordisk GBS
Email : zsuk@novonordisk.com
21-FEB-2026

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.


Background
Role of DGF
For phase 2 to Phase 3 decision
Phase 3 study initiated
We have statistically significant result !!
But, do we have clinically relevant effect??


Background
Role of DGF
For phase 2 to Phase 3 planning
Will statistically significant result be always clinically relevant ??
Revisit assumptions (Target product profile)
Phase 3 study initiated

Background

- Company X has a drug for cardiomyopathy that completed a Phase 1 study.
- Four dose levels (10–100 mg/kg) were well tolerated with no safety signals. Global longitudinal strain (GLS), a cardiac function marker, indicated a potential benefit.
- The company plans to proceed to a Phase 2 study using “Biomarker XX", a well-established biomarker for diagnosis and prognosis in heart failure. Phase 3 planning will be based on the Phase 2 results.
- During Phase 2 (before Database lock), the team requires a tool to guide the Phase 3 decision.
- What could be a suitable tool ?

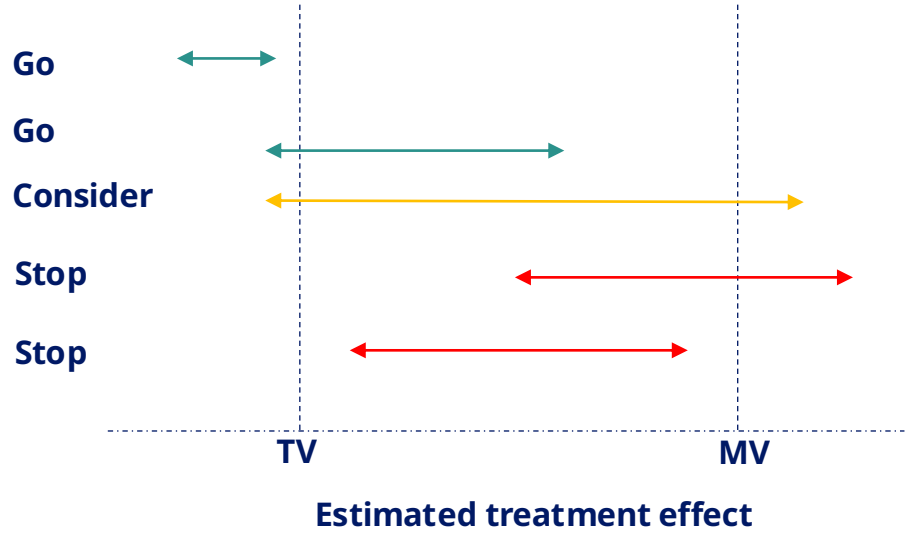
Decision Guiding Framework (DGF)

Construction of DGF and interpretation

Input parameter	Description
Target value (TV)	Highest treatment effect according to guided decision where the business case is attractive in-line with TPP (Also inline with results of competitor)
Minimum value (MV)	At least the lowest treatment effect with clinical/commercial value
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%)

* Frewer et al 2016 DOI:10.1002/pst.1746
** Lower values are better; higher values are worse.

Decision rule (For ratios**)



Interpretation :

- **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**

Will come back to this in few moment ..

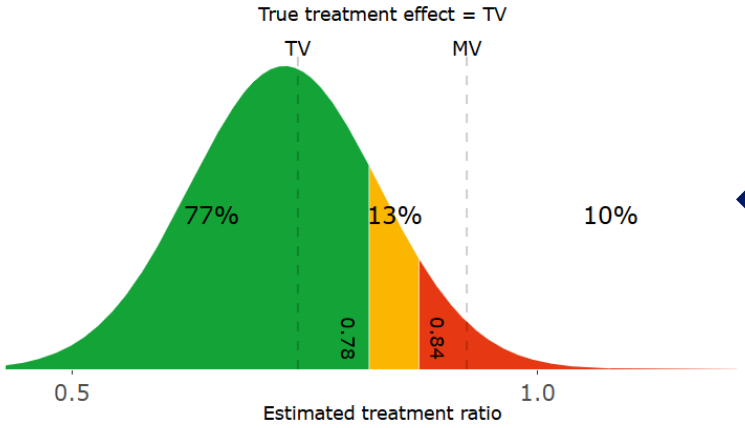
Construction of DGF and interpretation

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 particular 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 subjects>**	N = 50 per arm

Construction of DGF and interpretation

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




```
> ## 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

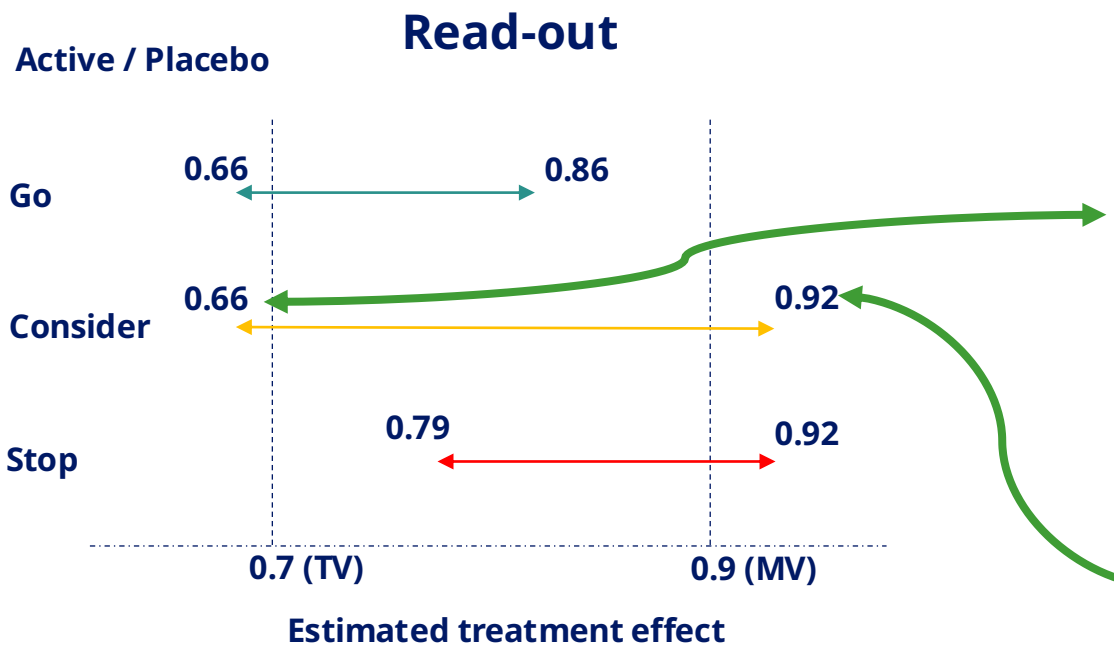
True treatment effect	Scenario	Go (%)	Consider (%)	Stop (%)
0.7	Target value	77.3	12.7	10.0
0.9	Minimum value	15.0	15.7	69.3

```
> # 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
```

- 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**

Construction of DGF and interpretation

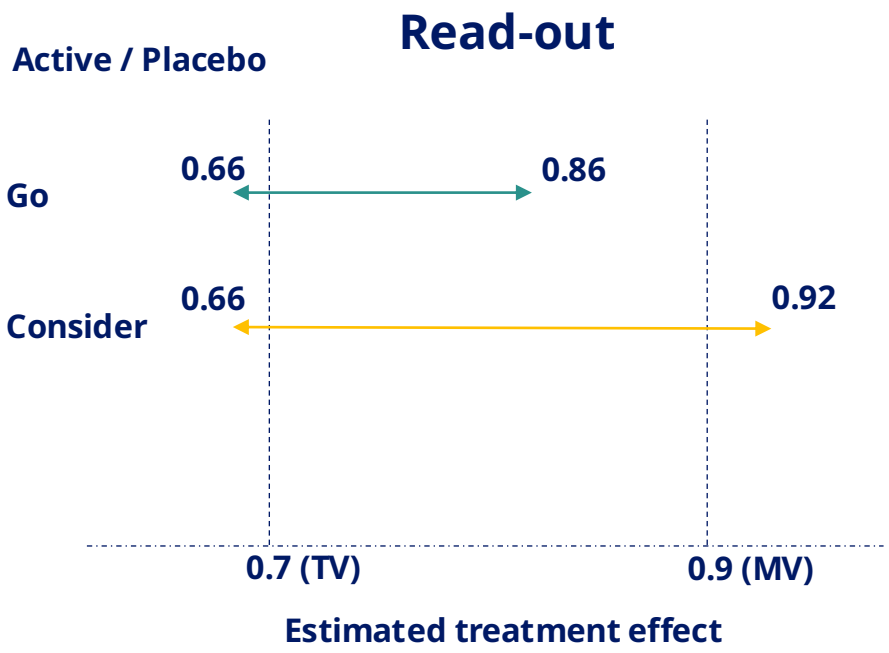
***NT-proBNP** - DGF guiding Phase 3 decision based on phase 2 results*




```r
# 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')
```

Construction of DGF and interpretation

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



Consider zone :

- Stakeholders to discuss with : Portfolio manager, M&S, Safety, Commercial, Competitive intelligence.
- Develop decision criteria based on other parameters/ secondary endpoints.

  For example, hsCRP, 6MWT, MRI/Echo parameters (ECV/GLS

  etc) and take a decision based on historical data.

(This should be planned earlier before database lock as a backup plan)

- 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.
- Sets the direction for clinical programme.
- Guiding document for regulatory activities, CMC and product supply planning as well as patient access initiatives.

TPP comprises of :

- 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

Figure source : AI generated

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

Phase-3 study read-out

Primary endpoint :

Time to first occurrence of a 3-component MACE endpoint comprising CV death, Non-fatal MI, Non-fatal stroke

The study is designed using an unstratified Cox model and a **power of 90%** to confirm superiority using a study-wise type I error rate of **2.5%** (one-sided) for the primary endpoint. Using a **randomisation ratio of 1:1** and **assuming a true HR of 0.7** a total of 331 primary endpoint events are required.

```
> 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
```

$$\text{Number of events} = \text{E} = \frac{(\text{Z}_{1-\beta} + \text{Z}_{1-\frac{\alpha}{2}})^2}{\text{p}_\text{A} * \text{p}_\text{B} * \log(\text{HR})^2}$$

$$\text{p}_\text{A} = \text{p}_\text{B} = 1/2$$

Study read-out

Study reads out positive with p-value*<0.05

Obs(HR) = 0.806

Is this HR clinically relevant?

* : 2-sided p-value

Questions around study read-out

- Note that, to actually observe the HR=0.7, we would need very large effect (p-value = 0.0012)
- If clinicians/commercial teams expect a hazard ratio (HR) of 0.7 when the study reaches statistical significance and treat 0.7 as the boundary of clinical relevance given the competitive landscape — there is a risk of disappointment.
- Specifically, if the final estimate falls between 0.70 and 0.806, the result may be statistically significant but clinically marginal; in the extreme, the trial could be "**statistically significant but clinically as well as commercially irrelevant**."

1. Is the observed hazard ratio consistent with the minimal clinically important difference (MCID)?
2. Do we "really" care at which effect we have 90% power?

Ultimately, we want studies that are powered in such a way that are both statistically significant for all effects which are also clinically relevant.

$$P(\text{Surpassing the MDD*}) \neq P(\text{Surpassing the Target})$$

*MDD as the minimal value we need to have to get a statistically significant difference.

Minimal detectable difference

How to calculate?

$$H_0 : \text{HR} \geq 1 \text{ ag. } H_1 : \text{HR} < 1$$

Let, $\theta = \log(\text{HR})$

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

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

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

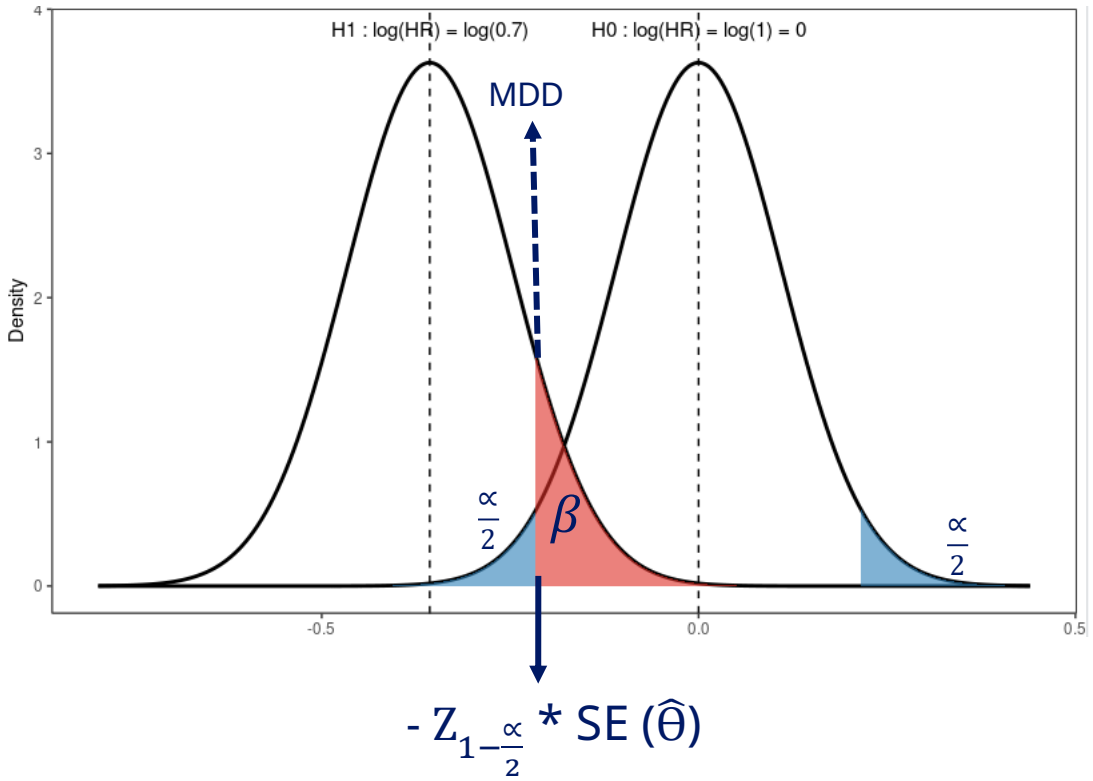
So, Minimal detectable difference = $- Z_{1-\frac{\propto}{2}} * SE(\hat{\theta})$

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

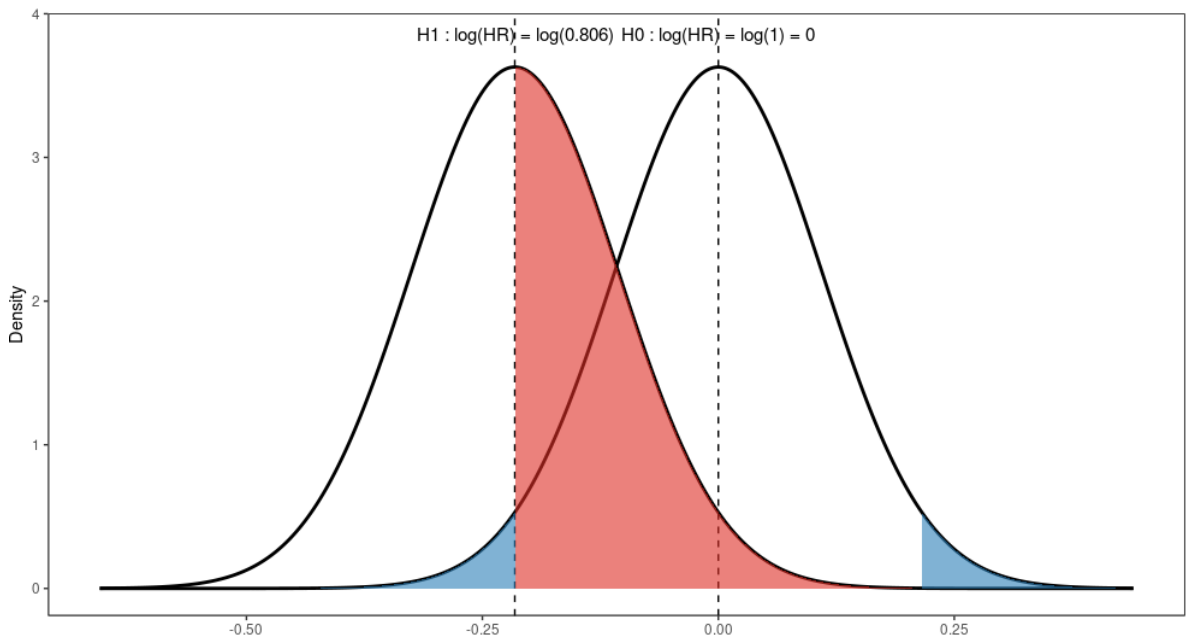
```
> # Sample size
> library(rpact)
> samplesize <- getSampleSizeSurvival(sided = 2, alpha = 0.05, beta = 1-0.9, hazardRatio = 0.7)
> samplesize$eventsPerStage
         [,1]
[1,] 330.3779
>
> # MDD from rpact
> mdd <- as.vector(samplesize$criticalValuesEffectScaleLower)
> mdd
[1] 0.8060081
```

How can we match our study design with Target product profile ?

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



If we shift the distribution of the alternative at MDD then we would see that the red area under the curve would be 50% - the effect at which we have 50% power.



How can we match our study design with Target product profile ?

Upside scenario/Aspirational TPP

HR=0.7

Base TPP

HR=0.8

Figure source : AI generated

How can we match our study design with Target product profile ?

The minimal hazard ratio we need to achieve is 0.8, so we set our MDD=0.8
So, we start our sample size calculation with 50% power

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

Now we would find the hazard ratio at which with **309 events we can achieve 90% power.**

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

$$p_A = p_B = 1/2$$

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

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

Approximately we are in-line with the aspirational TPP with 309 events to achieve 90% power.

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 trials 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 trials 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 : Probability of Success webinar by SSL Statisticians in Finnish Pharmaceutical Industry 20250506 13
- 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

Thank you