Interim Decision Making Using Parametric LM

ISBS Conference Kyoto 29th of August 2019

Tobias Mielke

janssen | PHARMACEUTICAL COMPANIES OF Johnson & Johnson

Objective

„Making the best possible decision at the earliest time-point in the most efficient manner“

Efficiency:

Depends on the underlying assumptions at the design stage, e.g.

- Too low effect assumed => too many patients enrolled
- Too high effect assumed => study underpowered

Adaptive designs:

Use interim data to inform on adjustments to the study design, maintaining efficient study, e.g.:

- Stop early for success / futility
- Drop treatment arms / subpopulations
- Increase study size

Adaptive Study designs

Early interim analyses

⇒ Few data available

⇒ Highly variable estimator

⇒ High chance of bad design change

Late interim analyses

⇒ Much data available

⇒ Less variable estimator

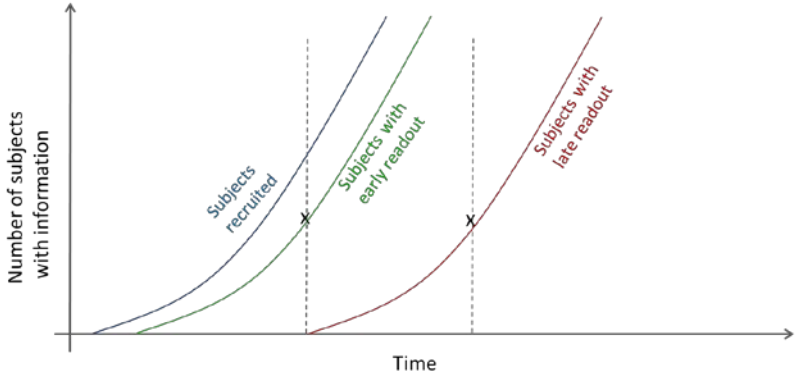
⇒ Less room left to improve study (end almost reached)

Early interim analyses require:

- Utilization of as much data as possible to support optimal decision making

Potential sources of additional data:

1. Don't use just "landmark"
2. Use data of non-completers
3. Use of Bayesian modelling



Problem / Scope

Potential sources of additional data:

1. Don't use just "landmark"
2. Use data of non-completers

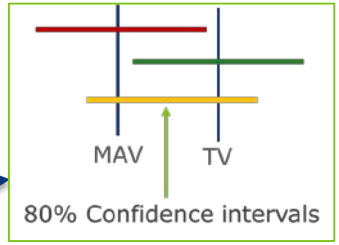
For simplicity, we're here only looking at:

- Early futility decision

Considered rules for decision making:

1. Is there any difference between control and test?
2. Do we reach a targeted effect size?
3. What is the conditional power?

H_0: $\cap_{i=1}^{T} \mu_i^1 \leq \mu_i^0$ at α_0.



Probability of study success given data <x

What is the potential impact of modelling on operating characteristics?

Motivating Example

Phase 3 with cont. “exercise” assessment at times (t=0, 2, 4, 8, 12, 16, 24):

$$Y_i^0 \sim N(\mu^0, \Sigma_0), \text{ with } \mu^0 = (\mu_0^0, \ldots, \mu_T^0),\ i = 1, \ldots, n$$

$$Y_i^1 \sim N(\mu^1, \Sigma_1), \text{ with } \mu^1 = (\mu_0^1, \ldots, \mu_T^1)\ ,\ i = 1, \ldots, n$$

Objective of study: Show superior change from baseline of active vs. placebo

↳ Reject hypothesis $H_0: \mu_{24}^1 \leq \mu_{24}^0$ at level α

Problem: Uncertainty on effect, small population and competing studies

↳ Terminate for futility if non-promising

Interim objectives:

1. Any difference between μ^0 and μ^1?
2. Relevant difference between μ_{24}^0 and μ_{24}^1?
3. Conditional power for H_0?

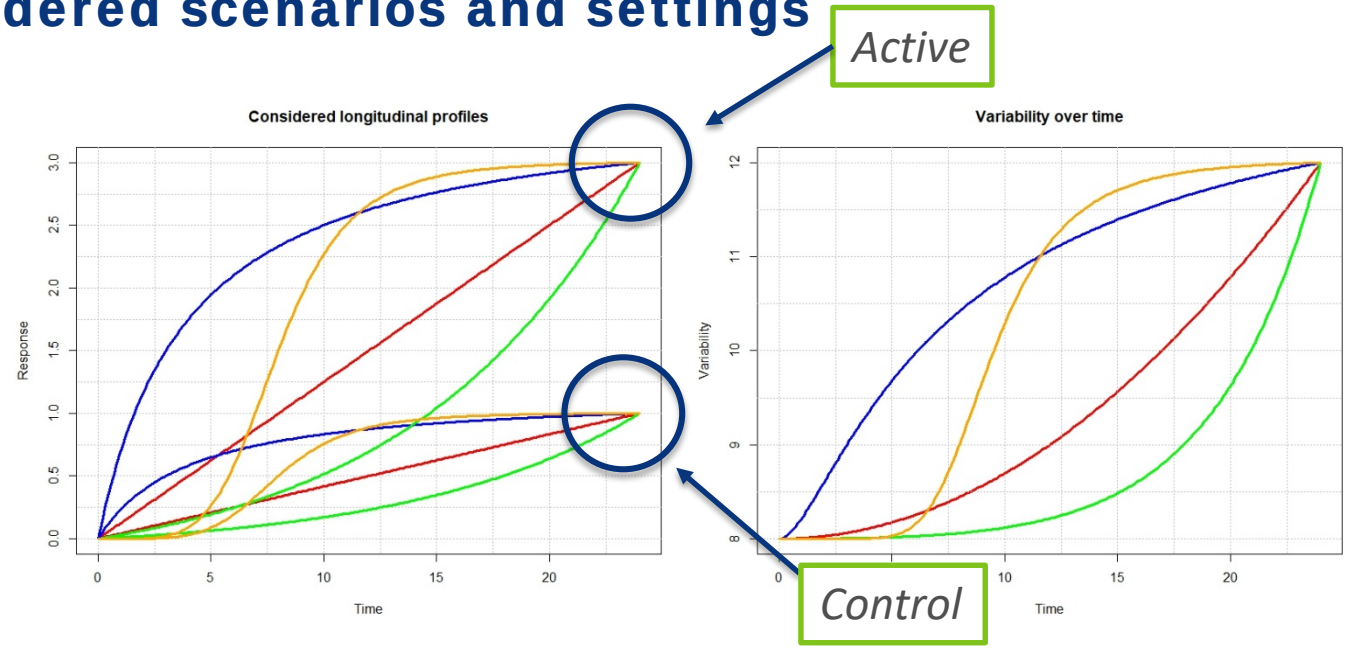
Standard approach:

$$\hat{\delta} := \overline{Y_{24}^1 - Y_0^1} - \overline{Y_{24}^0 - Y_0^0}$$

Variability:

$$\text{s.e.}(\hat{\delta}) = \sqrt{\frac{2\sigma_{T;pooled}^2}{n}}$$

Considered scenarios and settings



- Example settings: Interim analysis with N=25 completers per arm
- Effect: “2” at wk 24, intra-individual variability: 4
- Variance of change from baseline (wk 24): 12.

Model-based estimation of treatment effect

We might do better than "standard", by using modelling, e.g. LME:

- $Y_i = F_i\beta_i + \epsilon_i,\ \beta_i \sim N(\beta, D), \epsilon_i \sim N(0, \sigma^2 I_T) \Rightarrow\ \Sigma_i = \sigma^2 I_T + F_i D F_i^T$
- $\hat{\beta} = M^{-1} \sum_{i=1,\dots,N} F_i^T \Sigma_i^{-1} Y_i$, with $\mathrm{M} := \left(\sum_{i=1,\dots,N} F_i^T \Sigma_i^{-1} F_i\right)$
- $\hat{\delta} = c^T \hat{\beta}$

Problem: Estimator of treatment effect depends on:

- Considered model F_i
- Fitted variance D and σ^2

Example: 77% efficiency

Mean effect / Var	Linear	Emax
"Standard"	2.00 / 0.96	2.00 / 0.96
LME (1+t+trt x t)	2.00 / 0.74	1.87 / 0.65

**Example settings: True effect 2, Variance of change from baseline: 12, N=25/arm*

Estimator of treatment effect

Relative Bias / Coverage probability of 95%-CI

		Simulated scenario			
		Linear	EMax	Exponential	Sigmoidal
Analysis model	Linear	-1.3% / 93.5%	-8.2% / 93.1%	-11.0% / 93.0%	23.0% / 90.0%
	EMax	-25.0% / 87.0%	-0.5% / 93.3%	-43.0% / 69.5%	-1.9% / 93.1%
	Exponential	-4.0% / 93.0%	-29.8% / 87.2%	-0.5% / 93.6%	4.3% / 93.8%
	Sigmoidal	-30.0% / 81.0%	-31.0% / 80.0%	-43.3% / 63.6%	0.2% / 93.6%

Message: Wrong model assumed => Biased estimate and coverage probability off.

Does this mean: Don't use modelling?

Addressing Model Uncertainty

MCPMod for Dose-Finding*:

- Construct optimized test to reject: $H_0: \mu_0 = \cdots = \mu_G$
- Test for H_0 is optimized taking some dose-response assumptions into account.
- Based on rejection of H_0, full modelling is conducted

Translation to longitudinal modelling:

- Construct optimized test to reject: $H_0: \mu_1^1 - \mu_1^0 = \cdots = \mu_T^1 - \mu_T^0$
- Rejection of H_0: There is some difference in longitudinal profiles
- Test for H_0 can be optimized taking longitudinal profiles into account.
- Based on rejection of H_0, full modelling is conducted

Bretz et al. (2005) Biometrics, **61: 738-748*

Optimized test for objective 1: „Any difference“

Frison and Pocock (~22 years ago)*: Optimize c for maximum power:

$$P_\delta\left(\frac{c^T(\bar{Y}^1-\bar{Y}^0)}{\sqrt{2c^T\Sigma c}} > z_{1-\alpha}\right) = 1-\Phi\left(z_{1-\alpha}-\frac{c^T\delta}{\sqrt{2c^T\Sigma c}}\right)$$

For optimization required:

- Assumption on variability Σ and between group difference $\delta_i = \mu_i^1 - \mu_i^0, i = 1,\dots,T$.

Approach similar to MCPMod test:

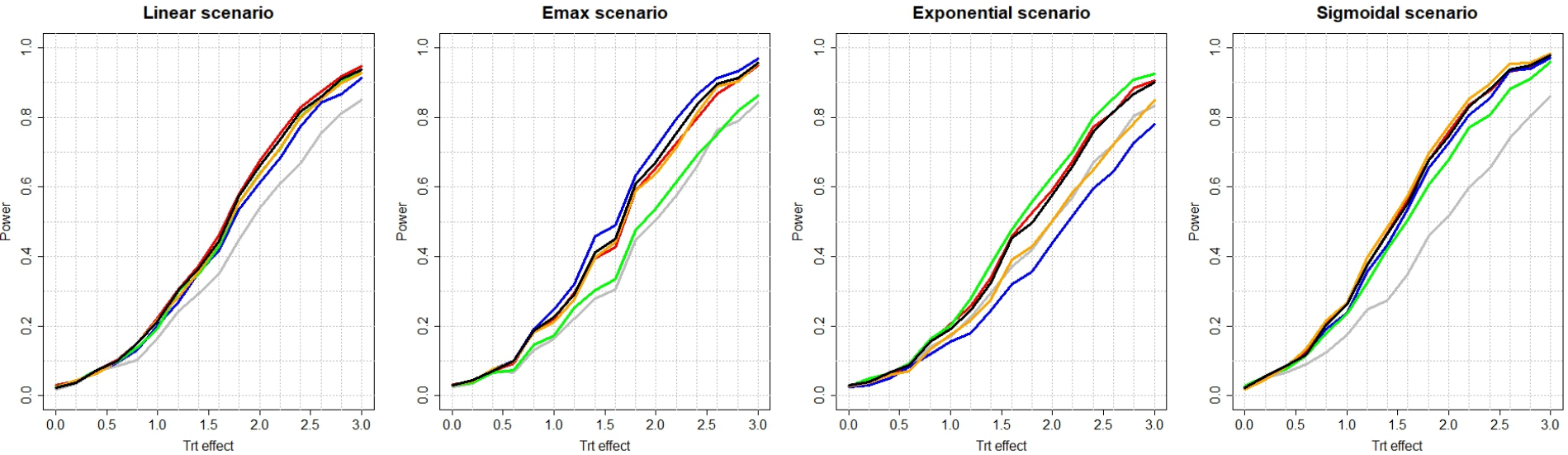
1. Assume parametric longitudinal model for both groups
2. Derive optimal coefficients c for these models based on above paper

e.g.: LME: $Y_i = F\beta_i^j + \epsilon_i$ with $\beta_i^j \sim N(\beta^j, D), \epsilon_i \sim N(0, \sigma^2 I_m), j = T, P$

Then: $\delta = F(\beta^T - \beta^P)$ and $\Sigma = \sigma^2 I_m + FDF^T$ and $\boldsymbol{c := \Sigma^{-1}\delta}$

*Frison & Pocock (1997), Statistics in Medicine 16, 2855-2872

Test for any difference in profiles



- Grey line: Change from baseline at wk 24. Substantial loss in power
- Blue/Green: Emax and exponential. Wrong will lead to decrease in power
- Red/Black: Linear and “all models”: similar high power in all scenarios
- Error control: Slight increase in type-1 error possible due to “non-normality”.

Taking Model Uncertainty into Account

Result:

- Efficient and robust test to establish activity.
- A test for “Linear” may increase efficiency as compared to simple test on change from baseline for all underlying scenarios
- Multiple models: May be robust, but comes at cost of multiplicity.

Missing:

Interpretability of the result.

- What is the effect?
- Does difference in longitudinal profiles translate into relevant benefit at any time?

Objective 2: Relevant difference between μ_T^0 and μ_T^1?

Dual "Go – No go" approach*:

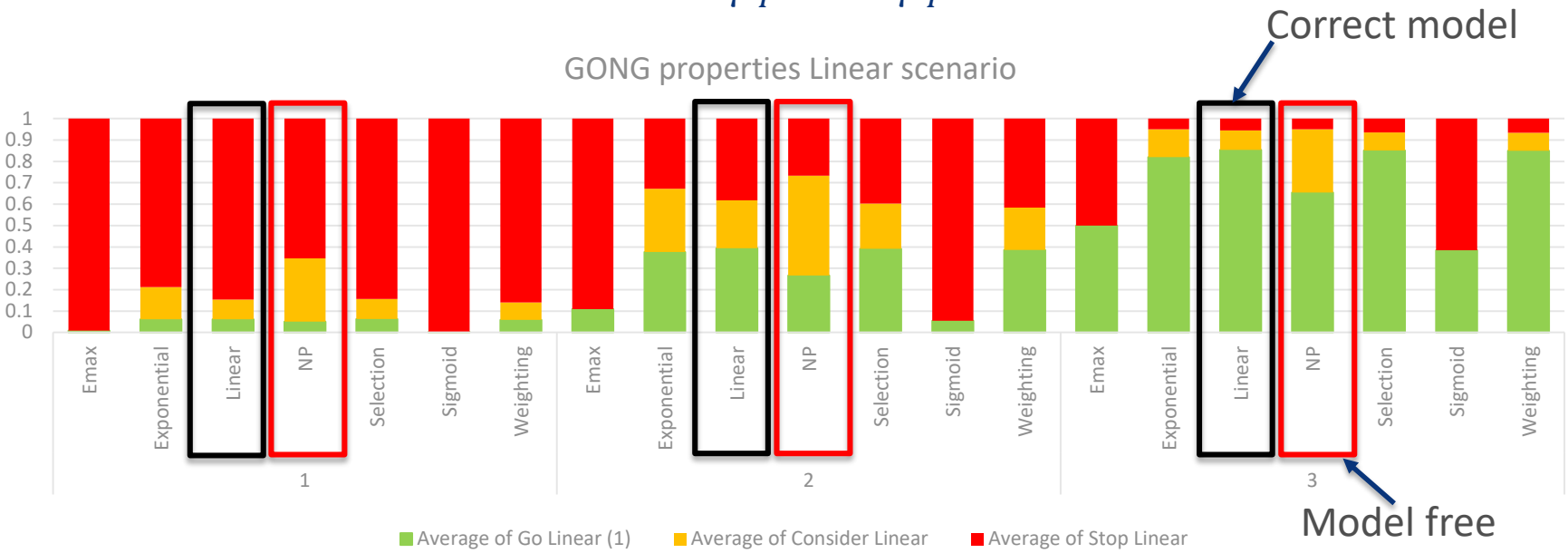
- Is the effect worth continued investment?
  - "Stop" if Δ significantly below some target effect (at some α_{TV})
  - Controls probability of false early termination for futility at α_{TV}
- Was a minimum relevant effect observed?
  - „Go" if Δ significantly above some maximum non-acceptable effect (at some α_{MAV})
  - Controls probability of false continuation at α_{MAV}

Problem:

- Control of false stopping / false continuation probability
- Model-based approach: data is drawn from model => error-control not guaranteed.
- Non-model based approach: significant amount of data not utilized.

*Lalonde et al. (2007), *Clin. Pharmacology & Therapeutics,* **82:** 21-32

Relevant difference between μ_T^0 and μ_T^1?



- Considered MAV=1 / TV=3 and required confidence =95% / acceptable risk = 5%.
- Wrong model may lead to false decisions (Emax and Sigmoid columns)
- Using no model leads to wider CI as less information utilized (NP columns)
- Weighting (by AIC) or selection (based on AIC) may provide robust alternative

Objective 2: Relevant difference between μ_T^0 and μ_T^1?

Result:

- Parametric modelling may substantially improve estimation of treatment effect
- Model selection / weighting can mitigate uncertainty on the correct model
- … but if appropriate models are not in the „candidate model set“, there is a chance for invalid tests.

Remaining problem:

- What is the probability that the study will show success in the final analysis with the final analysis methodology (i.e. „change from baseline at wk 24“, no LME)?

Objective 3: Conditional power

Conditional power: Probability of study success given observed data*

1. Patients with complete data at interim analysis. These provide statistic z_1
2. Patients with partial data (Z_2^*) at interim analysis. These will provide Z_2 (in future)
3. Patients with no data at interim analysis. These will provide Z_3 (in future)

Weighted combination test: $\sqrt{w_1}Z_1 + \sqrt{w_2}Z_2 + \sqrt{w_3}Z_3 > z_{1-\alpha}$, w_i proportional to size

Conditional power:

$$CP = P\left(\sqrt{w_1}z_1 + \sqrt{w_2}Z_2 + \sqrt{w_3}Z_3 > z_{1-\alpha} \middle| Z_1, Z_2^*\right) = 1 - P\left(\sqrt{w_2}Z_2 + \sqrt{w_3}Z_3 > z_{1-\alpha} - \sqrt{w_1}z_1 \middle| Z_1, Z_2^*\right)$$

For calculation required: assumption of distribution of $\sqrt{w_2}Z_2 + \sqrt{w_3}Z_3 | Z_1, Z_2^*$

*Similar to presentation of An Vandebosch today

Conditional power: Non-parametric with early data

Conditional power:

$$CP = P\left(\sqrt{w_1}Z_1 + \sqrt{w_2}Z_2 + \sqrt{w_3}Z_3 > z_{1-\alpha} \middle| Z_1, Z_2^*\right) = 1 - P\left(\sqrt{w_2}Z_2 + \sqrt{w_3}Z_3 > z_{1-\alpha} - \sqrt{w_1}z_1 \middle| Z_1, Z_2^*\right)$$

Need: distribution of $\sqrt{w_2}Z_2 + \sqrt{w_3}Z_3 | Z_1, Z_2^*$

- Z_3 independent of Z_1, Z_2^* => $Z_3 \sim N(\sqrt{n_3}\frac{\hat{\delta}}{\sigma\sqrt{2}}, 1)$, e.g. $\sqrt{n_3}\frac{\hat{\delta}}{\sigma\sqrt{2}} = \sqrt{n_3/n_1}z_1$
- Z_2: already observed something => variability <1.

Consider joint model for early and late test statistic:

$$Z_{2;joint} = \begin{pmatrix} Z_{2;early} \\ Z_{2;late} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{early} \\ \mu_{late} \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

Then: $Z_{2;late} | Z_{2;early} \sim N\left(\mu_{late} + \rho\left(Z_{2;early} - \mu_{early}\right), (1 - \rho^2)\right)$

- What to assume for μ_{late}, μ_{early} and ρ? Estimate this from data of completers.

**Galbraith & Marschner (2003), Statistics in Medicine, 22: 1787-1805*

Conditional power: Parametric approach with early data

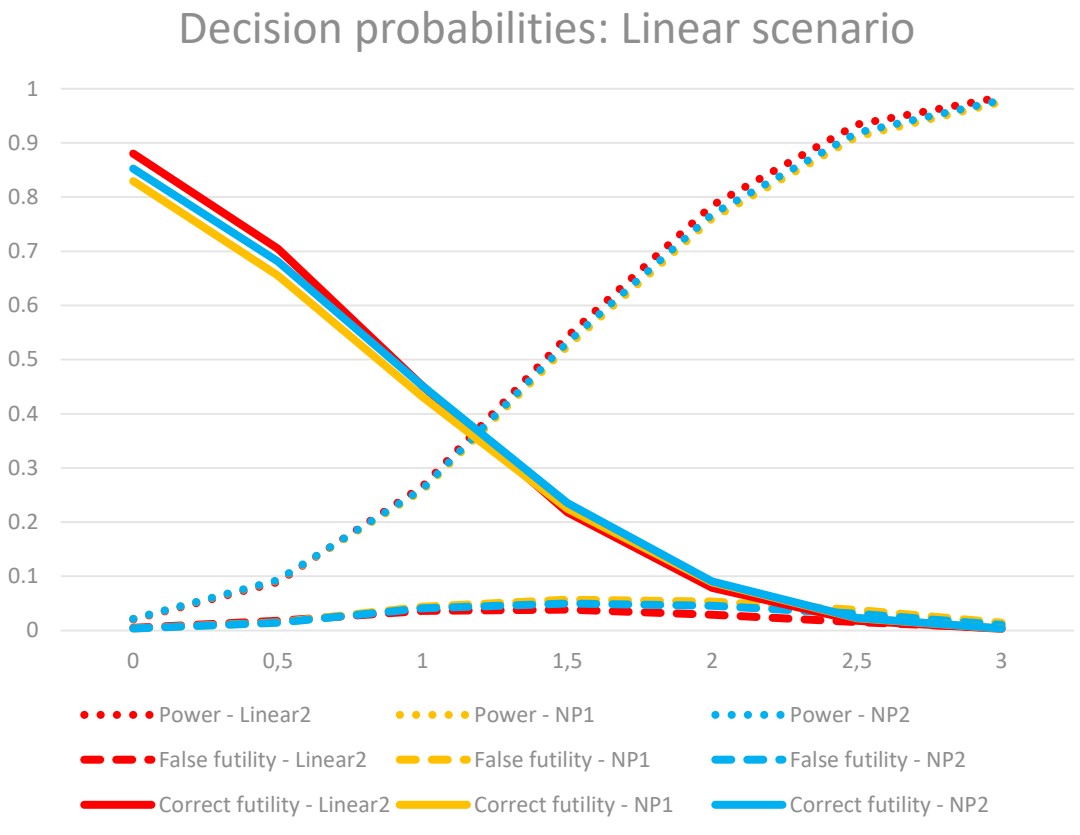
Conditional power:

$$CP = P\left(\sqrt{w_1}Z_1 + \sqrt{w_2}Z_2 + \sqrt{w_3}Z_3 > z_{1-\alpha} \middle| Z_1, Z_2^*\right) = 1 - P\left(\sqrt{w_2}Z_2 + \sqrt{w_3}Z_3 > z_{1-\alpha} - \sqrt{w_1}z_1 \middle| Z_1, Z_2^*\right)$$

- Get parameter estimates from LME
- For distribution of Z_3: Estimate standardized effect based on parameter estimates
- For conditional distribution of Z_2, use
  - Baseline observation (known)
  - Interim parameter estimates
  - Estimates of variability

  … to estimate conditional distribution of Z_2 given early data.

Implications on futility analysis (20% CP cutoff)



If correct model selected:

- Higher chance for correct futility in ineffective scenarios
- Lower chance of false futility in effective scenarios
- Superior in power

as compared to analysis without modelling

Conditional distribution:

- Utilizing short-term data also improves decision making against standard approach (NP1 vs. NP2)

What about model dependency?

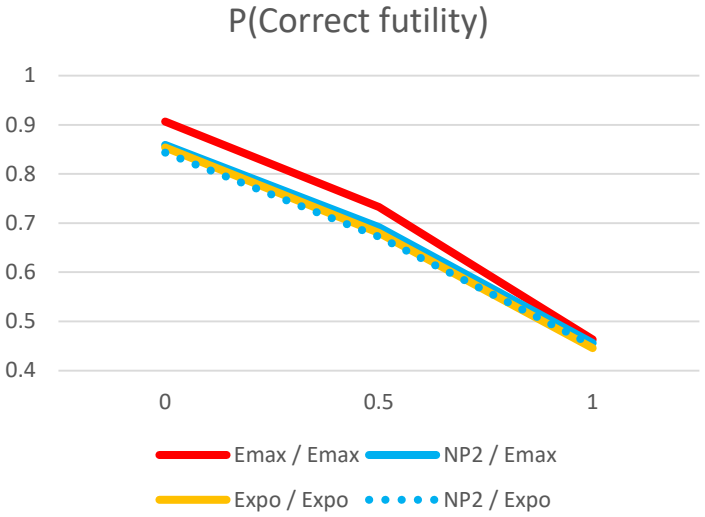
Shape of underlying model important:

- Largest benefit with EMax/Sigmoidal
- Smallest benefit with Exponential

Why this? Design theory:

- Early time in EMax/Sigmoidal highly informative
- Late time in Exponential highly informative

↳ **Less information available at analysis time**



Model misspecification:

- Wrong model can lead to bad results
- Model selection / weighting mitigates risk

False stops	Linear	Emax	Expo	Sigmoid
Min	3.3%	2.3%	4.5%	2.1%
Max	8.0%	5.9%	13.3%	6.2%
Selection	4.0%	3.4%	4.7%	3.4%
Weighting	4.0%	3.4%	4.7%	3.4%

Summary

Many ways how longitudinal data can inform interim decisions:

- Model-based but still non-parametric futility test for “any difference”
- Full modelling and model-based testing of primary endpoint
- Conditional power calculations utilizing model-based effect estimates

Not knowing the “true model” does not imply “don’t do it”

- Model uncertainty to be evaluated in design phase
- Use of robust models, model selection or weighting procedures can help

More general adaptations beyond futility stopping:

- Possible, e.g. based on conditional power
- … but cautious evaluation of conditional power required to be not mislead.

Thanks for your attention!

References

References:

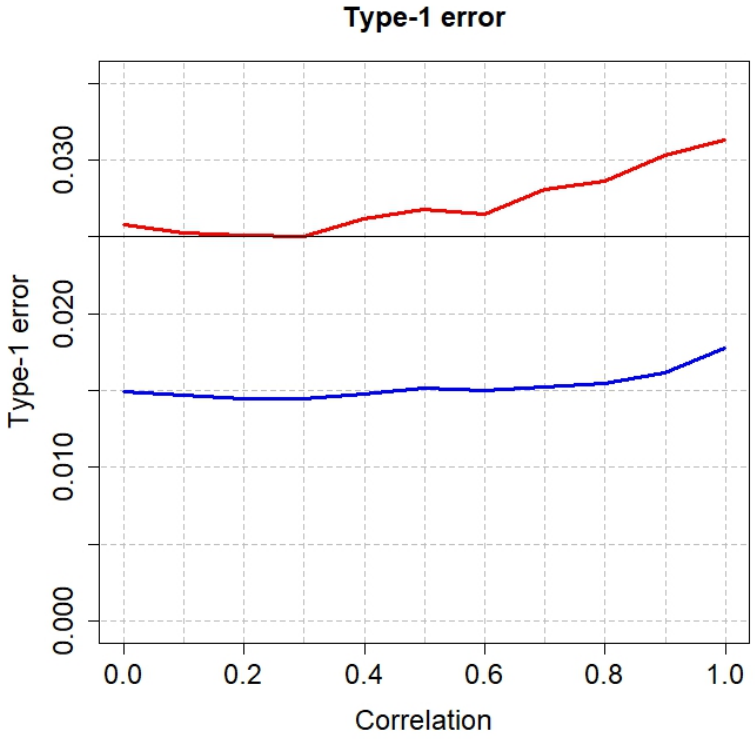
Bretz, F., Pinheiro, J.C., Branson, M. (2005), "Combining Multiple Comparisons and Modeling Techniques in Dose-Response Studies" *Biometrics*, **61**: 738-748

Frison, L.J., Pocock, S.J. (1997), "Linearly divergent treatment effects in clinical trials with repeated measures: efficient analysis using summary statistics", *Statistics in Medicine*, **16**, 2855-2872

Lalonde, R.L., Kowalski, K.G., Hutmacher, M.M., Ewy, W., Nichols, D.J., Milligan, P.A., Corrigan, B.W., Lockwood, P.A., Marshall, S.A., Benincose, L.J., Tensfeldt, T.G., Parivar, K., Amantea, M., Glue, P., Koide, H. And Miller, R., (2007), "Model-based Drug Development". *Clinical Pharmacology & Therapeutics,* **82:** 21-32

Galbraith, S., Marschner, I.C. (2003), "Interim analysis of continuous long-term endpoints in clinical trials with longitudinal outcomes" *Statistics in Medicine*, **22**: 1787-1805

Back-up 1: Sample size-reestimation



Problem:

- We don‘t control type-1 error any more if we don‘t adjust appropriately

Why?

- With large correlation, we know exactly, whether follow-up of the enrolled patients will lead to success.
- We only increase, in case original plan won‘t.
- It‘s some kind of a free shot for success.

Solving the problem:

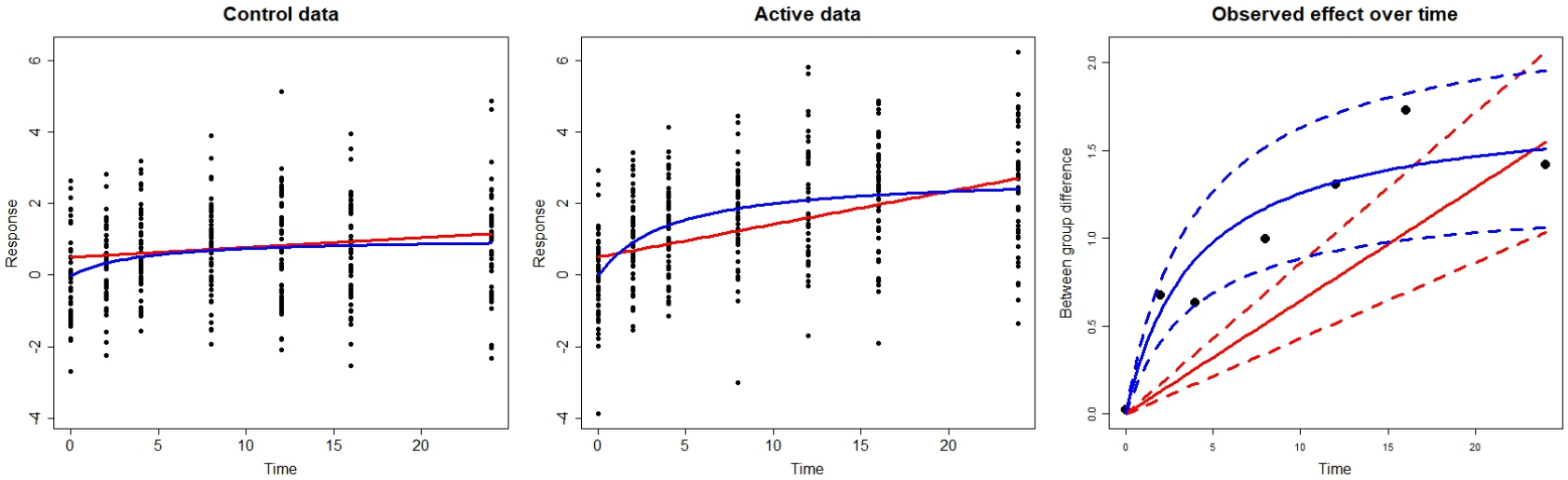
„Population-wise splitting“

- We may apply prespecified weights:

$$Z = \sqrt{w_1}Z_1 + \sqrt{w_2}Z_2 + \sqrt{1 - w_1 - w_2}Z_3$$

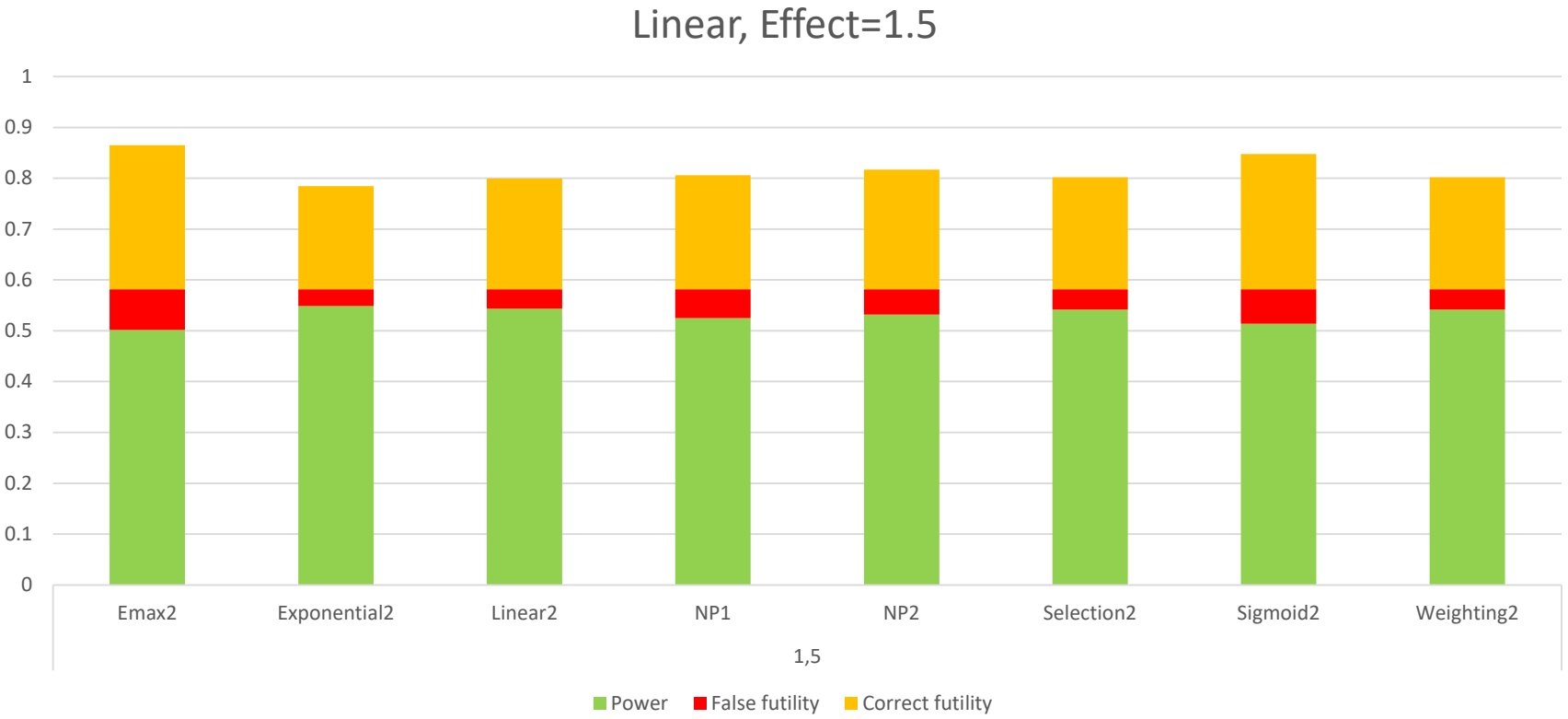
… but then we‘re conservative.

Estimator of treatment effect



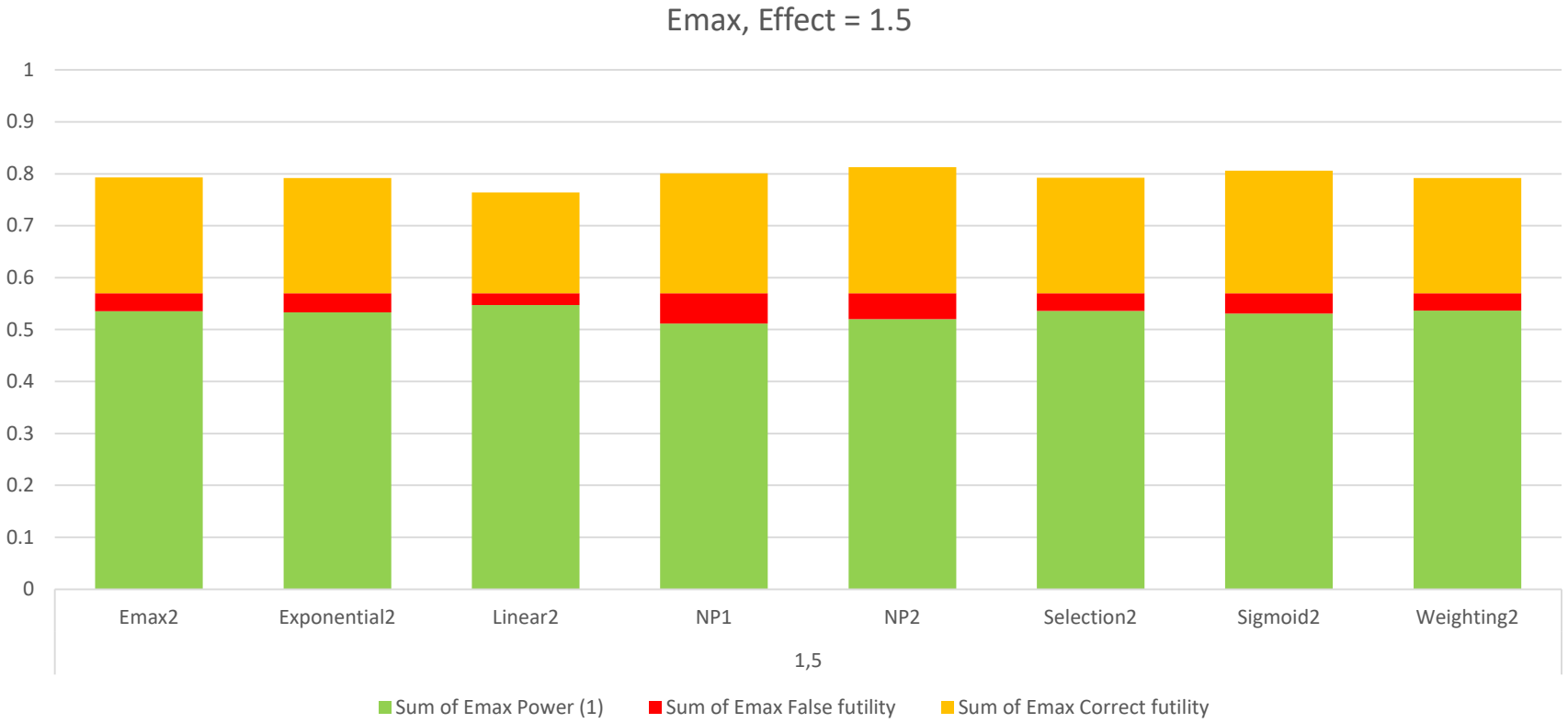
- True model: „Emax“ – but Linear may capture effect at last visit reasonably well
- Why? Extremes provide maximum information for linear model
- Model based confidence intervals?
  - Estimators follow normal distribution => $\hat{\delta} = c^T\hat{\beta}$ follows normal distribution
  - Variance matrix standard output of fitting procedure => $c^T\hat{\beta} \pm z_{1-\alpha}\sqrt{c^T M^{-1} c}$

Futility stopping - Linear scenario (CP)

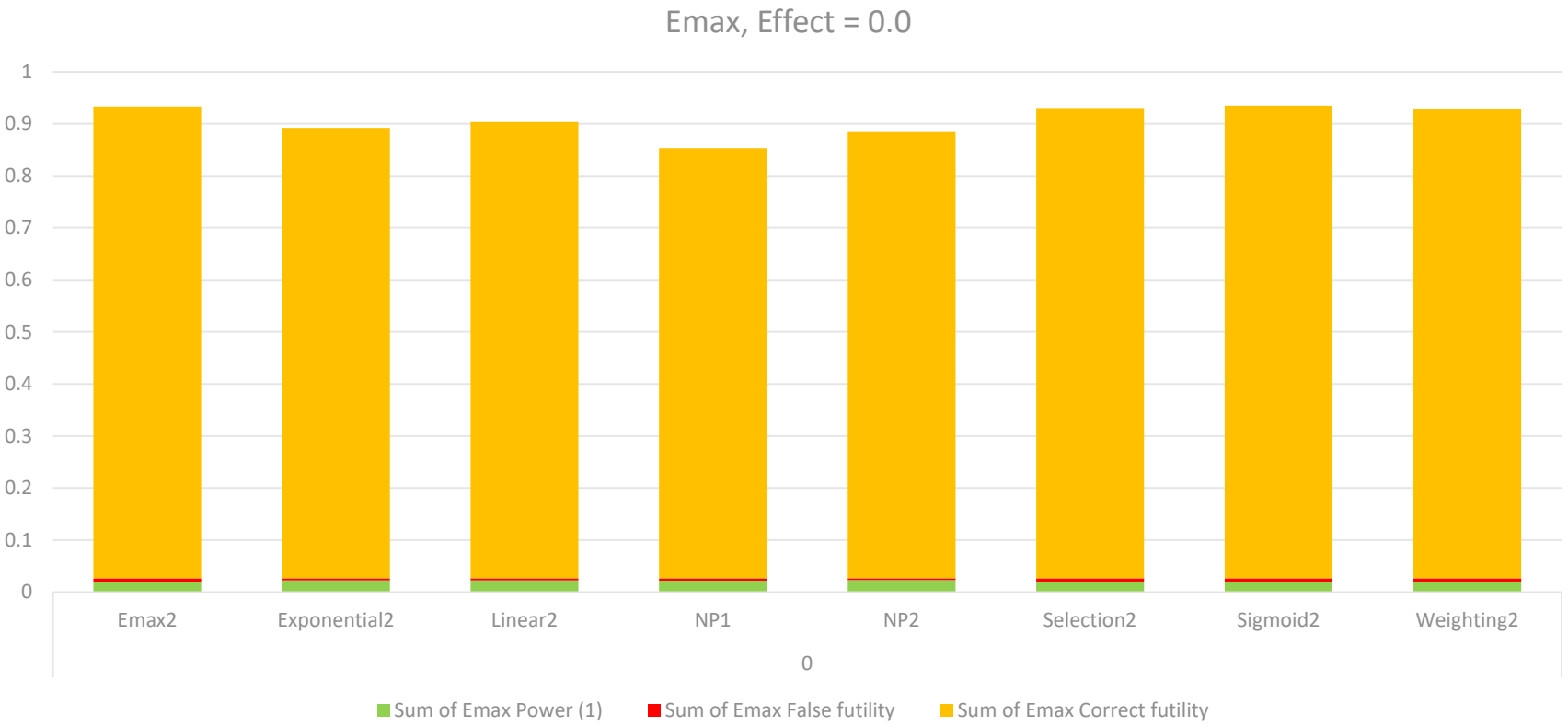


Futility stopping - Linear scenario (CP)

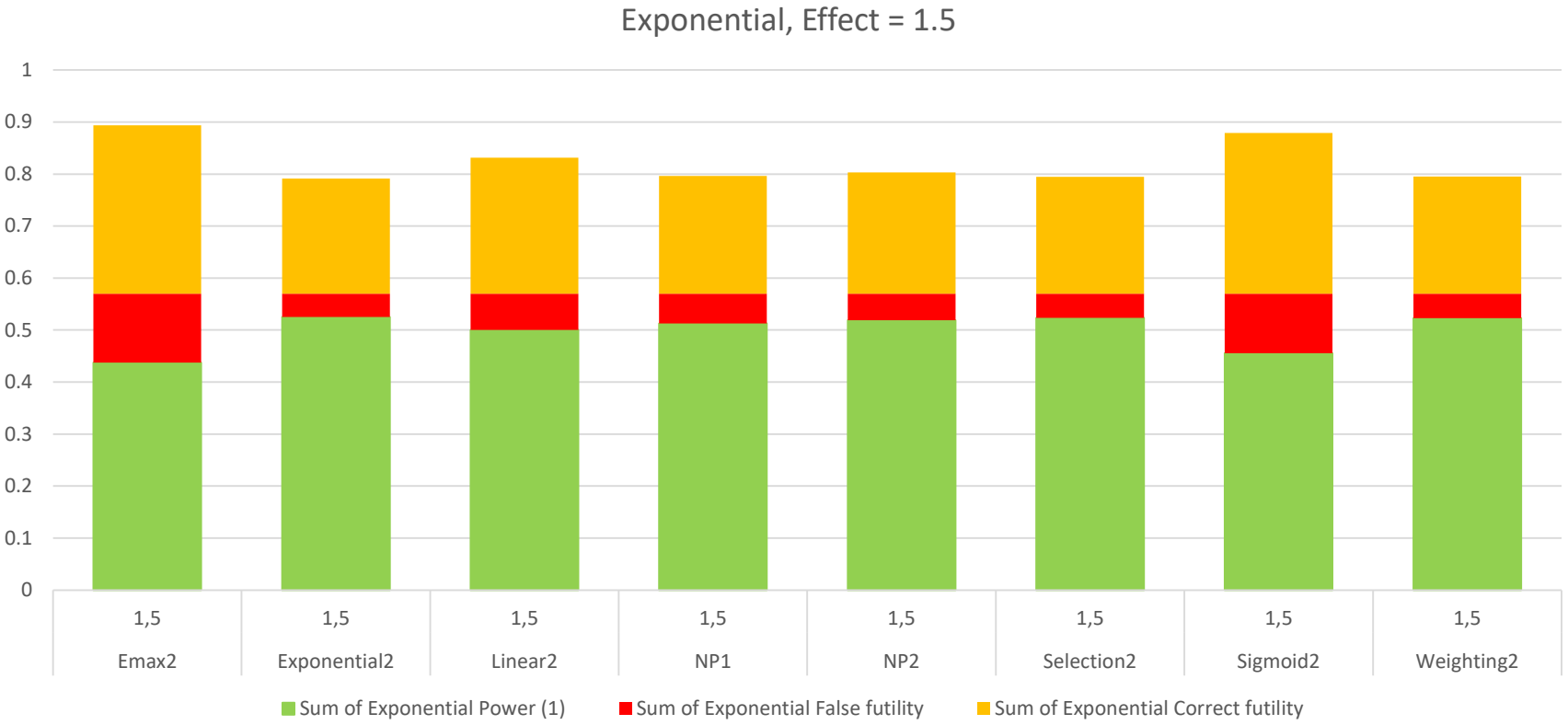
Futility stopping - Emax scenario (CP)



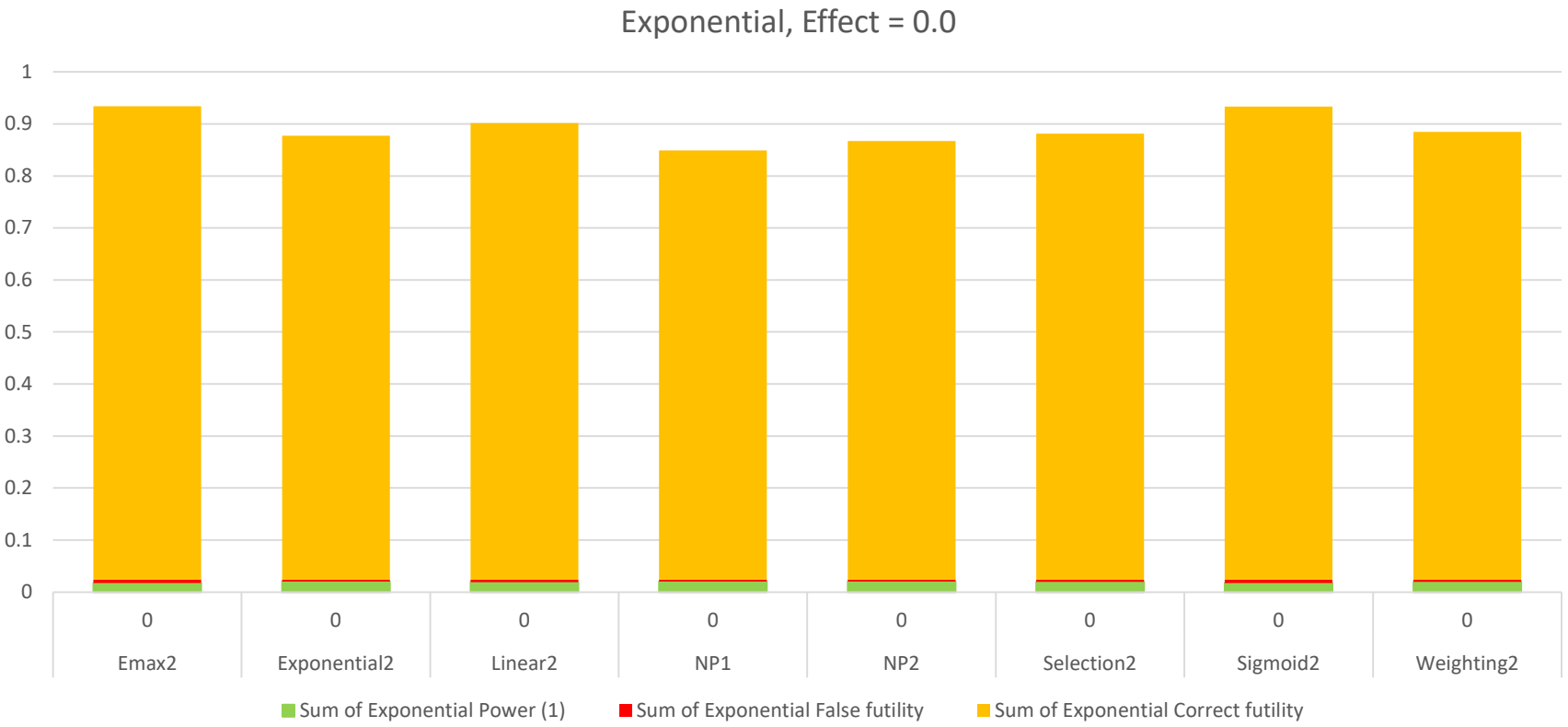
Futility stopping - Emax scenario (CP)



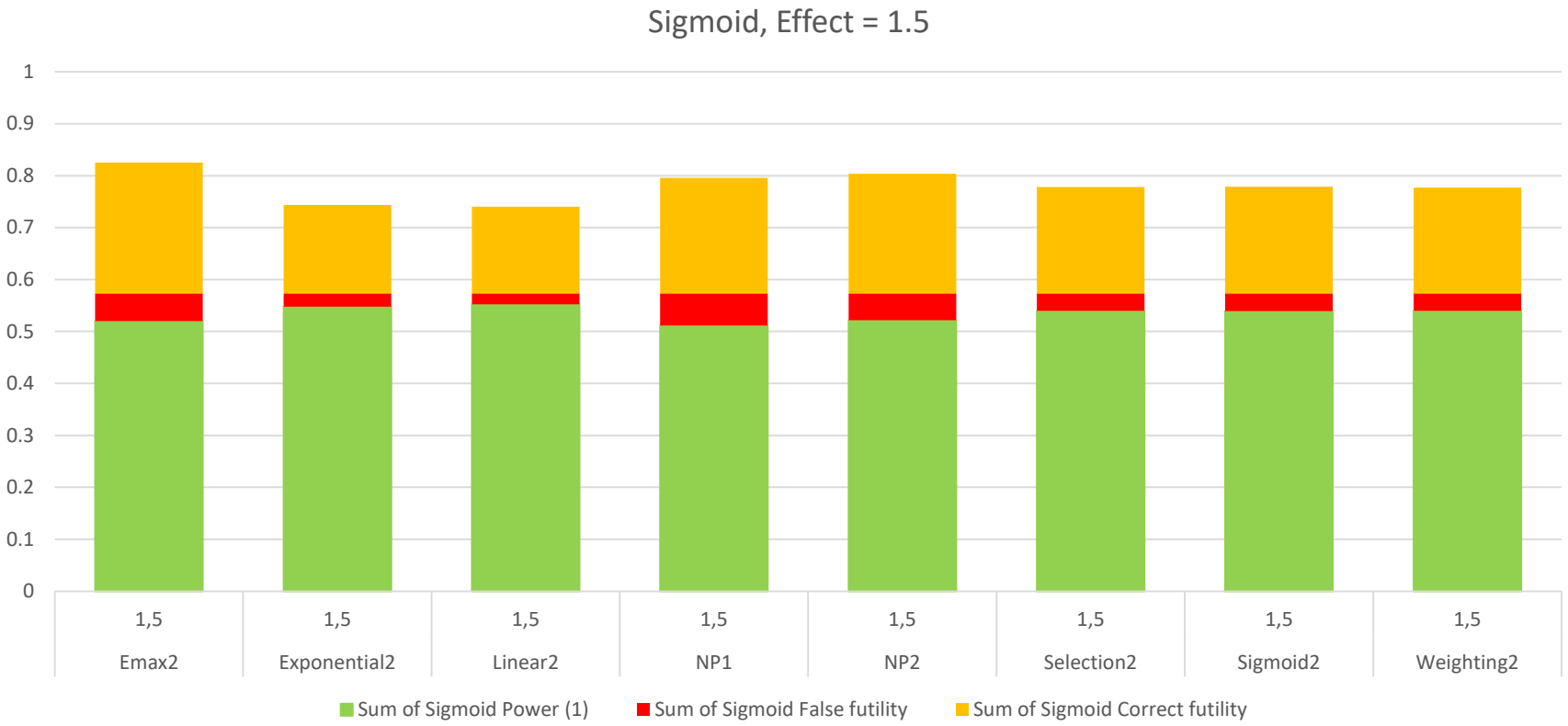
Futility stopping - Exponential scenario (CP)



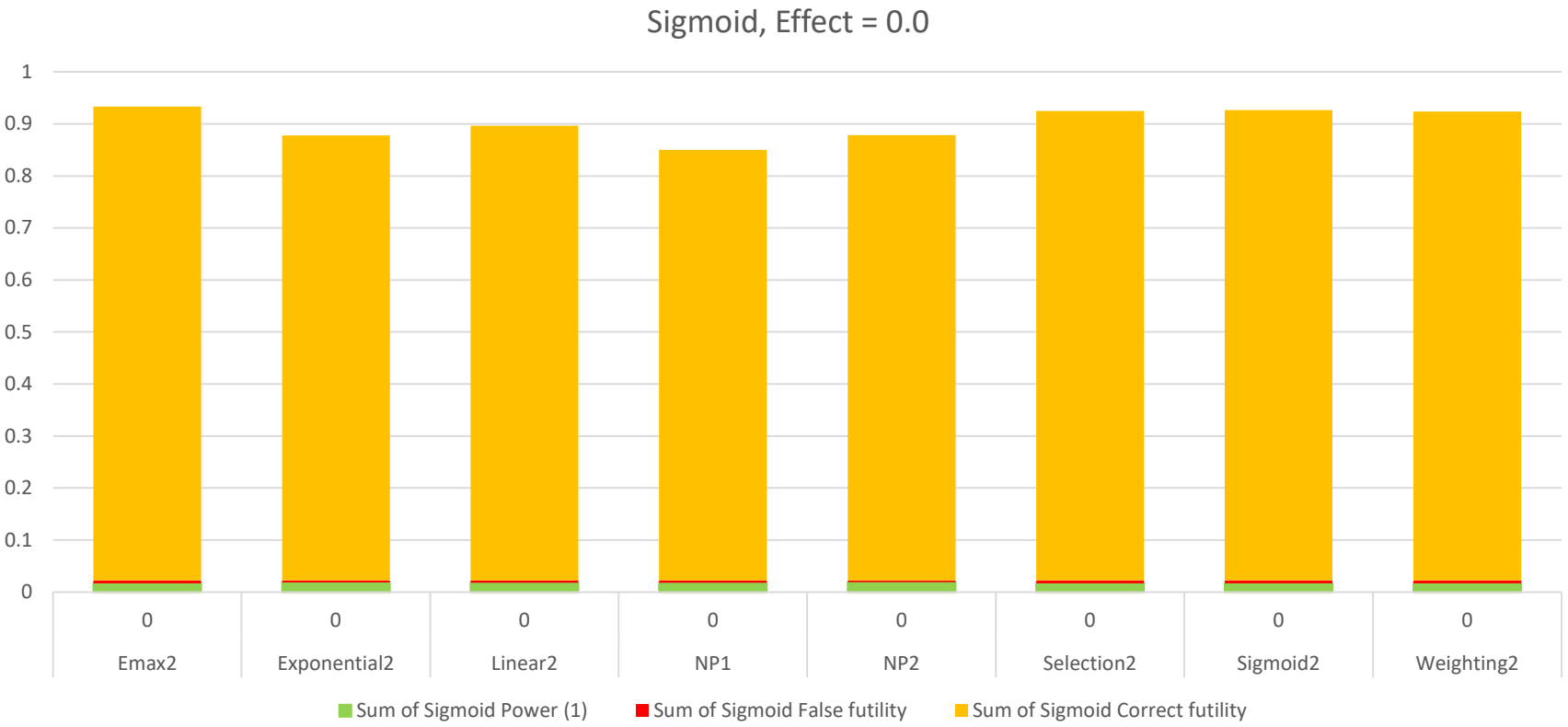
Futility stopping - Exponential scenario (CP)



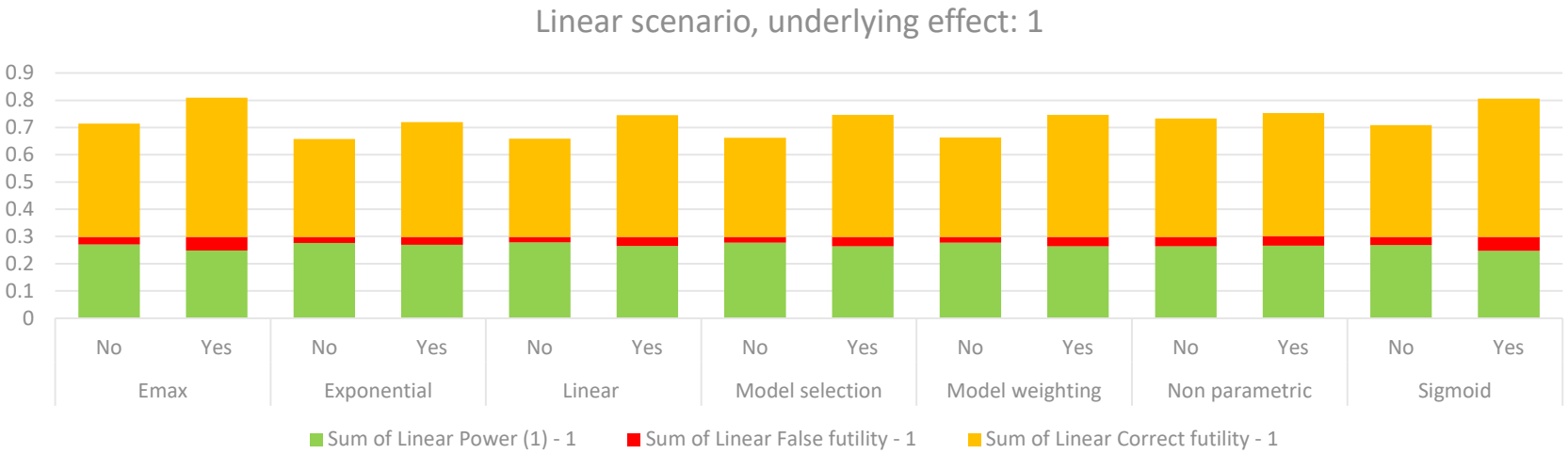
Futility stopping - Sigmoid scenario (CP)



Futility stopping - Sigmoid scenario (CP)



Implications on futility analysis

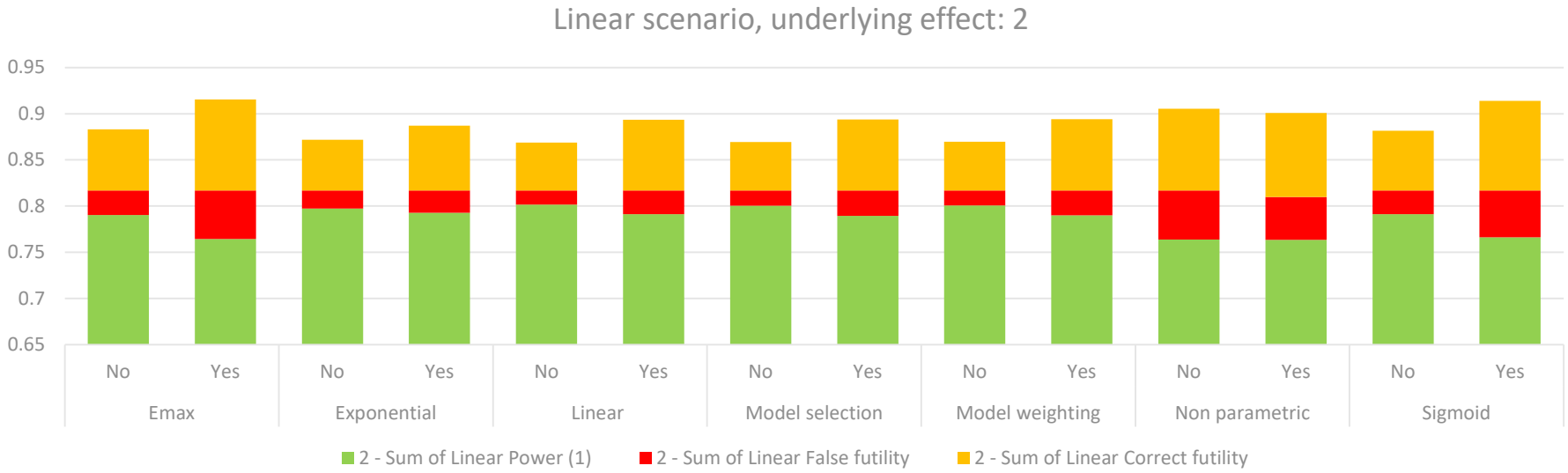


- No = conditional distribution for patients with partial data has not been utilized
- Simulation model: Linear. Cutoff for futility: 20% conditional power.

Result:

- Use of conditional distribution increase probability of futility stops
- Model selection / weighting worked fine in the considered scenarios

Implications on futility analysis



- Specially in effective scenarios, probability of false stops can be decreased using modelling
- … whereas use of conditional vs. unconditional distribution for conditional power calculation has both positive and negative effects, i.e. more correct and false stops.