SSL/EFSPI Webinar Dose Finding in Oncology

Dose-optimization strategies in late development

February 10th, 2026

Bristol Myers Squibb®

Background

- **Optimal dose may not be known by the time Phase 3 study starts**
- **Competing landscape and business needs may require the Phase 3 study to start as soon as possible**
- **Key aspects of dose-optimization in Phase 3**
- Multiplicity (Family-Wise Error Rate must be strongly controlled)
- Sample size (which may include Phase 2 sample size)
- Study duration and enrollment
- Probability of success (undermined by the probability of choosing the wrong dose?)

Motivating example

Primary efficacy endpoint is a survival endpoint

Median in control arm: 15 months

Targeted HR: 0.70

Enrollment rate

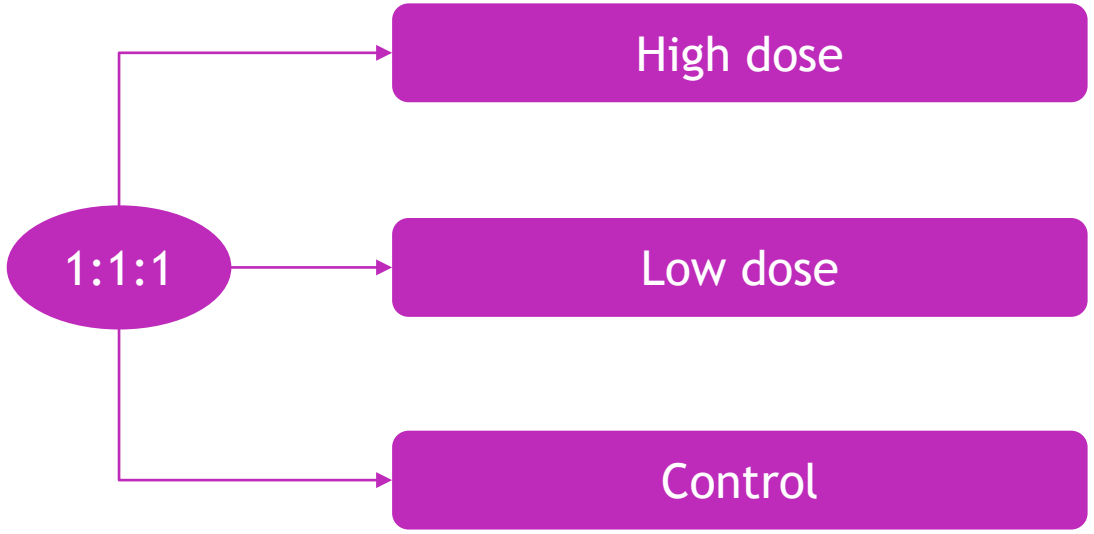
- **at most 40 patients/month with a 12-month ramp-up period (linear)**
- **For accelerate enrollment, let's consider a 2-month ramp-up period (see in seamless design sections)**

For sack of simplicity,

- **Let's consider only a Final Analysis (with an interim analysis only for dose-selection if any)**
- **Same primary efficacy endpoint for dose-selection and final analysis**
- **Let's target a ratio Randomized patients/Events ~ 60%**

Parallel arms

Study design

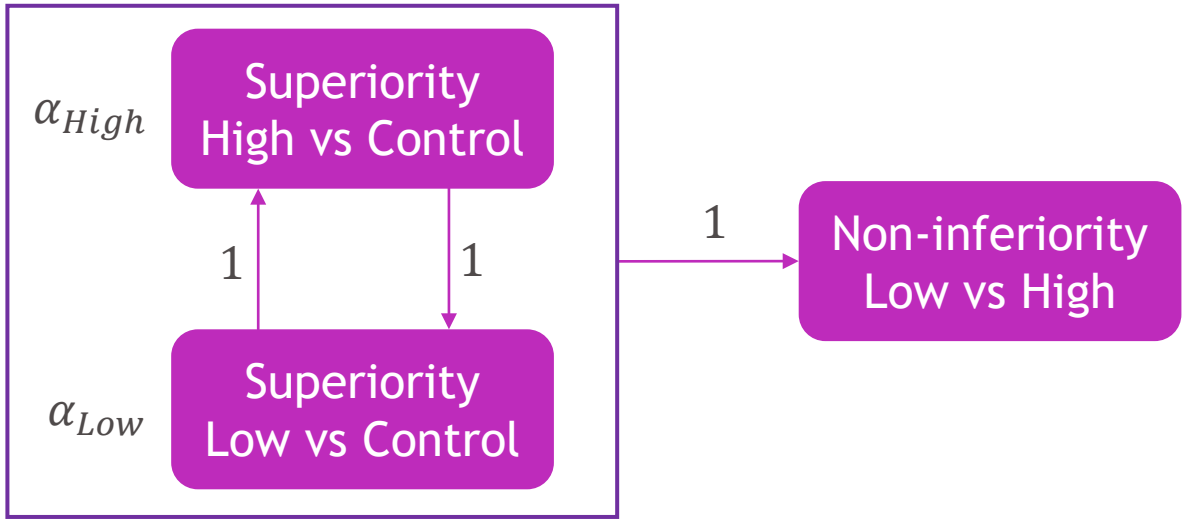


- Multiplicity
  - High dose vs Control
  - Low dose vs Control (no need to compare doses if low dose is not better than control)
  - Non-inferiority of low vs high dose? Superiority regarding a Benefit-risk score?
- Targeted Power and Sample size

Testing strategies

- **Hierarchical procedure**

- **Gatekeeping procedure**



Sample sizes

Design	Alpha (1-sided)	Power	Sample size	Events at FA	cHR at FA	Accrual	Study duration
Hierarchical	2.5%	90%	828*	480*	0.806	26mo	42mo
Gatekeeping	$\alpha_{High} = 1.25\%$ $\alpha_{Low} = 1.25\%$	84%-90%	828*	480*	0.781	26mo	42mo
Gatekeeping		90%	975**	567**	0.797	30mo	44mo

* For each drug comparison, 551 patients and 331 events
**For each drug comparison, 650 patients and 391 events

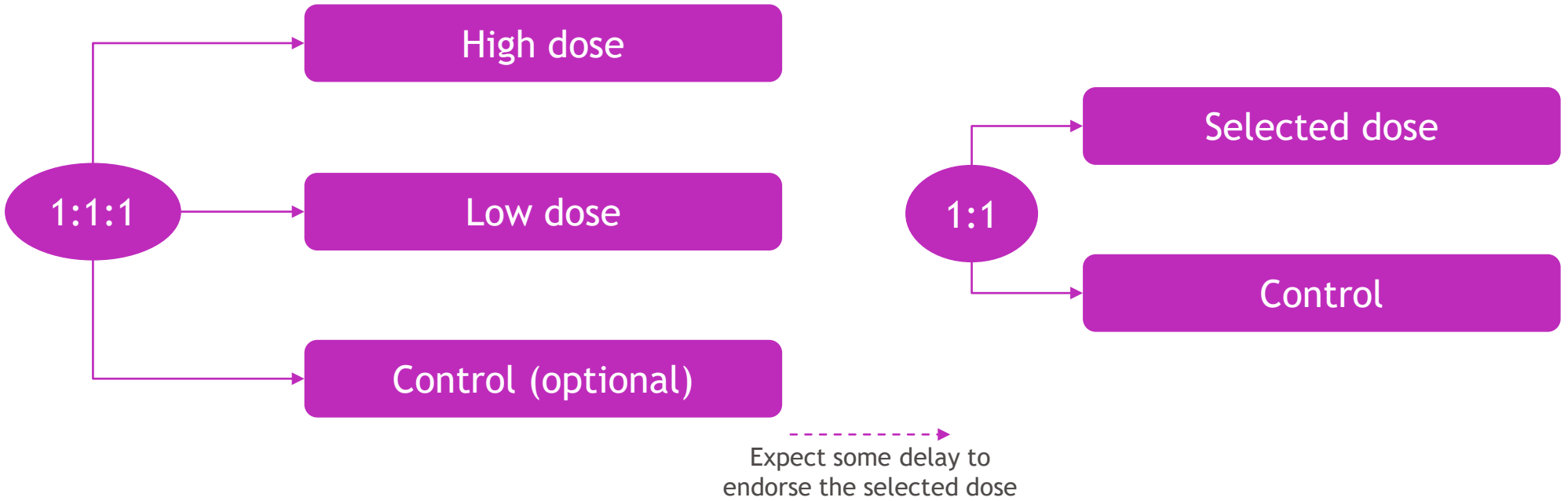
- Hierarchical procedure is convenient if there is clear dose-response relationship
- Gatekeeping procedure provides more flexibility, but at a cost (loss of power or larger sample size). It could be convenient if there is a safety concern (which may undermine the drug efficacy) regarding the high dose.
- Overall sample size is much larger than a usual clinical development Phase 2 + Phase 3 (Phase 2 ~100-150 patients)
- 3-arm study is slightly longer (4-5 months longer) than a 2-arm study (assuming there is no need for dose-optimization)
- A non-inferiority analysis of low vs high dose is not properly powered (~49%)
- A sponsor might be tempted to limit the size of low dose arm, but this strategy is at risk if the low dose provides similar efficacy

Summary

- **Easy to implement**
- **Dose-optimization is postponed to final analysis (or even to HAs' review), relying on mature data**
  - Primary endpoints are fully evaluable, but also any kind of benefit-risk score (since Overall Survival requests the longest follow-up)
- **Leading to larger studies**
- **Some sponsors may prefer enrolling half the sample size in the low dose experimental arm**
  - No formal testing of low dose vs high dose
  - Saving time and resources
  - Clearly assuming that low dose is less effective
  - Not back-up plan if the low dose is showing similar efficacy

Operationally seamless design

Study design



Single protocol

Sample sizes

Design	Alpha (1-sided)	Power	Sample size	Events at FA	cHR at FA	Accrual	Study duration
Phase 2 part	10%	84%	225 (75/arm)	112 (across 3 arms)	0.742	12mo	26mo
Phase 3 part	2.5%	90%	551	331	0.806	19mo*	38mo*
Overall		75.6%**	776				>64mo

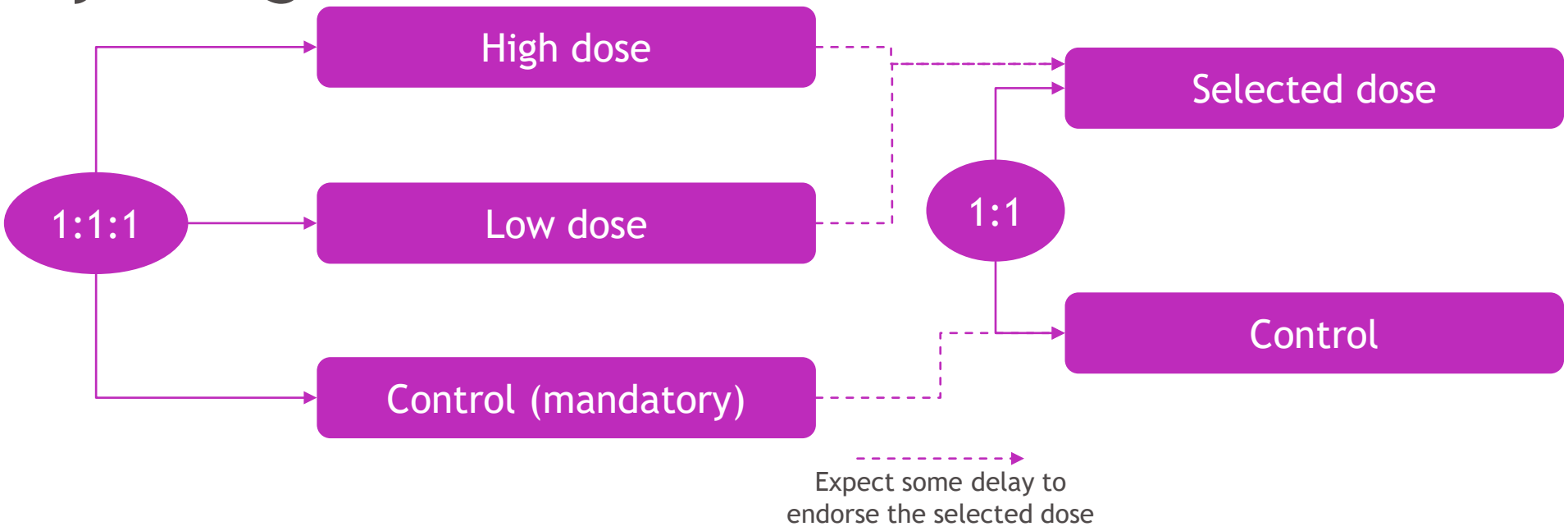
* Phase 3 study part is shorter than a 3-arm study thanks to the faster accrual
**Go/no-Go decision after Phase 2 study part would act as an aggressive futility analysis

- Phase 3 study part is designed as a standalone Phase 3 study
- Overall sample size is significantly lower than a 3-arm study
- Waiting for Phase 2 outcome significantly delays the clinical development
- **Phase 3 study part can be accelerated if most of investigator sites are activated in Phase 2. Assuming a 2-month ramp-up enrollment:**

Design	Alpha (1-sided)	Power	Sample size	Events at FA	cHR at FA	Accrual	Study duration
Phase 3 part	2.5%	90%	551	331	0.806	14mo*	33mo*
Overall		75.6%**	701				>59mo

Inferentially seamless design

Study design



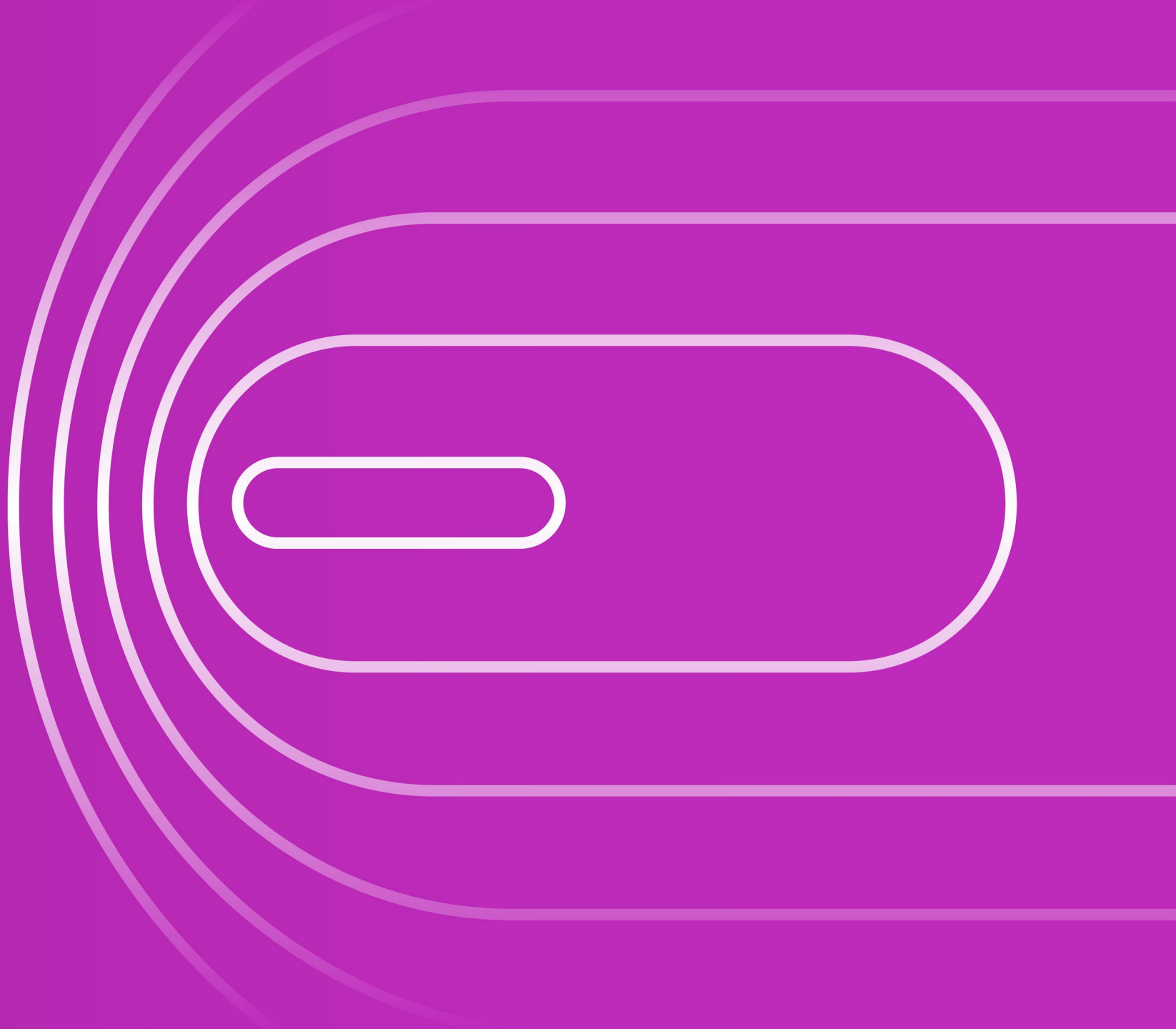
- Experimental treatments are compared to the control treatment using Dunnett's many-to-one testing procedure
- In this example, only the best experimental arm is selected at dose-selection analysis (regardless of effect size). Other selection rules are possible (select more than 1 experimental arm, select all with meaningful effect size…)
- Data from the interim and final analyses for the final outcome measure are combined together using either the inverse normal combination test

Sample sizes

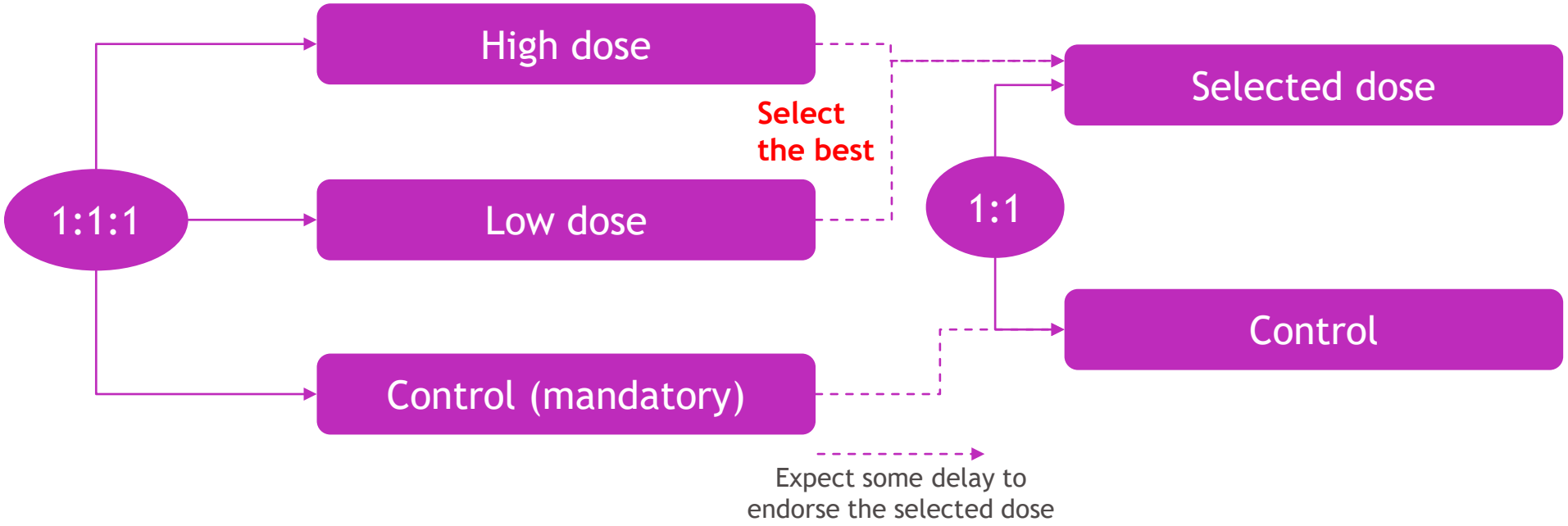
Design	Alpha (1-sided)	Power	Sample size	Events at FA	Accrual	Study duration
Phase 2 part			225 (75/arm)	123 (across 3 arms)	11mo	26mo
Phase 3 part	2.5%	90%	401 (selected arm + control)	312(selected arm + control)	7mo*	>= 41mo
Overall			626			

- Combination test based on patients randomized before vs after the dose selection
- If allowing multiple experimental arms to continue to Phase 3 study part, then the sample size 1) may significantly increase and 2) becomes a random variable (which may be raised as an issue for clinical operations and study planning).

Drop-the-losers design



Study design



- Binding selection rule based on primary endpoint
- Efficacy is not tested (vs control arm) at dose-selection analysis
- Final analysis is based on all data (from Phase 2 and Phase 3 study parts) for control and selected dose arms

Drop-the-losers framework

Contrasts

Selected dose is better than dropped dose at Interim Analysis	$Z^{(IA)}_{Selected} > Z^{(IA)}_{Dropped}$
Selected dose is significantly better than control at Final Analysis	$Z^{(FA)}_{Selected} > c^{(FA)}$

By considering the correlation between test statistics across time and experimental arms (sharing the same control arm), the distribution of $Z^{(FA)}_{Selected}$ can be derived. The critical threshold $c^{(FA)}$ is chosen to control Family-Wise Error Rate (FWER).

$$FWER = 2 \times \mathbb{P}\left(Z_1^{(IA)} > Z_2^{(IA)} \text{ AND } Z_1^{(FA)} > c^{(FA)} \mid \text{Null hypotheses}\right)$$

Proof of strong control of FWER is provided in Wason et al. (A multi-stage drop-the-losers design for multi-arm clinical trials. Stat Methods Med Res. 2017)

Sample sizes

Design	Alpha (1-sided)	Power	Sample size	Events	cHR at FA	Accrual	Study duration
Dose selection	NA	NA	225 (75/arm)	112 (across 3 arms)	NA	11mo	26mo
Phase 3 part	2.5% 1.68% (adjusted)	82%	401 (selected arm + control)	331** (selected arm + control)	0.792	7mo*	>= 41mo
Phase 3 part	2.5% 1.63% (adjusted)	90%	767 (selected arm + control)	550** (selected arm + control)	0.833	20mo*	>= 46mo

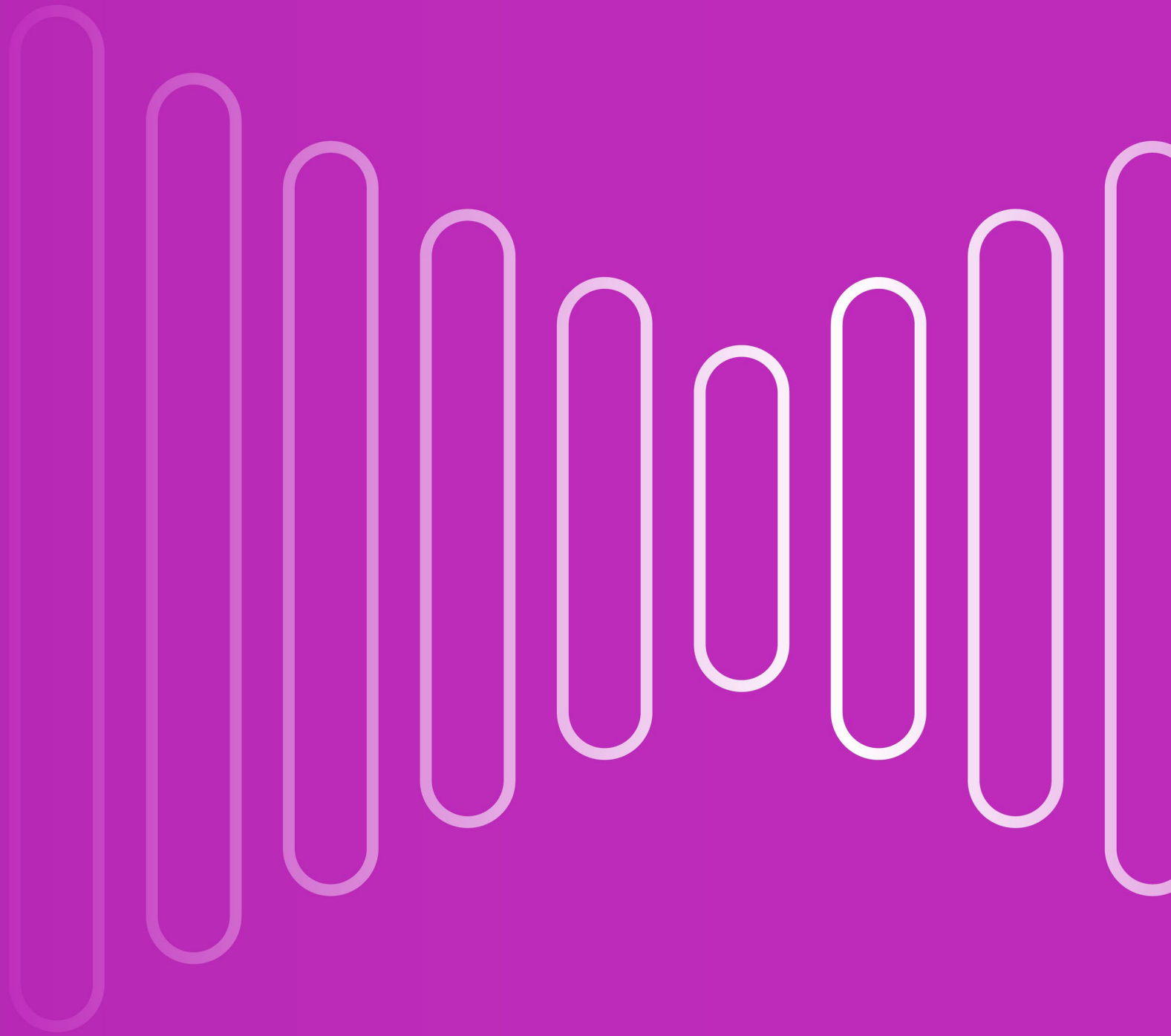
* 2-month ramp-up enrollment

** Including Phase 2 events

- Early binding dose-selection can lead to significant increase of Phase 3 sample size
- Type II error at dose-selection can only be balanced by high conditional power in Phase 3 study part
- **Larger Phase 2 study part may be preferred**

Design	Alpha (1-sided)	Power	Sample size	Events	cHR at FA	Accrual	Study duration
Dose selection	NA	NA	360 (120/arm)	180 (across 3 arms)	NA	14-15mo	28mo
Phase 3 part	2.5% 1.63% (adjusted)	90%	398 (selected arm + control)	383** (selected arm + control)	0.805	11mo*	>= 42-43mo

Conclusion



Summary

Approach	Pros	Cons
Parallel arms	• Most robust approach (if properly powered) • Unbiased estimators • Shorter late development	• Most expensive studies
Operationally seamless design	• Saving time by 1. anticipating dose-selection rule and Go for Ph3 decision, 2. activating investigator sites during Phase 2 study part • Unbiased estimators (in Phase 3 study part)	• Same sample size as Phase 2 + Phase 3 studies • (Lack of) of correlation between early and late efficacy endpoints
Inferentially seamless design	• Flexible dose-selection rules • Not discarding Phase 2 data	• Complex methods with strong adjustment • Can be heavily undermined by non-proportional hazard scenarios • Biased estimators • (Lack of) of correlation between early and late efficacy endpoints
Drop-the-losers	• Limited adjustment for multiplicity • Strong control of FWER • Not discarding Phase 2 data	• Binding dose-selection rule based on primary endpoint • Can be heavily undermined by non-proportional hazard scenarios • Biased estimators • (Lack of) of correlation between early and late efficacy endpoints

Key messages

- **Dose-optimization is usually easier in Phase 2 (fewer requirements)**
- **Dose-optimization in Phase 3 requires careful considerations of**
  - Sample size
  - Study duration
  - Scenarios of interest:
    1. Low dose providing some efficacy
    2. Non-proportional hazard (undermines the dose-optimization)
    3. Correlation between primary efficacy endpoint(s) and endpoint(s) supporting the dose-optimization
- **Pay attention to Phase 2 sample size, regardless of statistical design**
  - More balanced sample size allocation between Phase 2 & Phase 3 is statistically more efficient
- **Health Authorities usually recommend to consider flexible designs, instead of binding selection rules**
- **Assess the final estimators' characteristics (specifically bias due to dose selection)**

References

Inferentially seamless designs

Thall PF, Simon, R, Ans Ellenberg SS. Two-stage selection and testing designs for comparative clinical trials. Biometrika 1989;75,303-310.

Bauer P, Kieser M. Combining different phases in the development of medical treatments within a single trial. Statistics in Medicine 1999;18:1833-1848.

Stallard N, Todd S. Sequential designs for phase II and phase III clinical trials incorporating treatment selection. Statistics in Medicine 2003;22:689-703.

Posch M, Koenig F, Branson M, Brannath W, Dunger-Baldauf C, Bauer P. Testing and estimation in flexible group sequential designs with adaptive treatment selection. Statistics in Medicine 2005;24:3697-3714.

Bretz F, Schmidli H, Koenig F, Racine A, Maurer W. Confirmatory seamless phase II/III clinical trials with hypotheses selection at interim: General concepts. Biometrical Journal 2006;48:623-634.

Koenig F, Brannath W, Bretz F, Posch M. Adaptive Dunnett tests for treatment selection. Statistics in Medicine 2008;27:1612-1625.

Stallard N, Friede T. A group-sequential design for clinical trials with treatment selection. Statistics in Medicine 2008;27:6209-6227.

Drop-the-losers design

Wason J, Stallard N, Bowden J, Jennison C. A multi-stage drop-the-losers design for multi-arm clinical trials. Stat Methods Med Res. 2017 Feb;26(1):508-524. doi: 10.1177/0962280214550759. Epub 2016 Sep 30. PMID: 25228636; PMCID: PMC5302074.

Software R

Parallel design: R-package rpact

Operationnally seamless design: R-package rpact

Inferentially seamless design: R-package asd

Drop-the-losers design: In-house programming (sample size) and R-package rpact (study duration)

Thank you


Bristol Myers Squibb®