Cytel

An Adaptive Sequential Bayesian Design for a COVID Vaccine Trial

Ingrid Lönnstedt, Ph D, Principal Consultant Statistics
and Site Manager Uppsala

FMS Autumn Meeting 24 Nov 2025

Ingrid Lönnstedt

PhD Mathematical Statistics

Work experience

Pharmaceutical industry

- Consulting in drug development and life science as biostatistician ~15 years
  (last > 5 years with SDS Life Science/Cytel: clinical trial designs and exploratory analyses, regulatory)
- Management and biostatistics in global big pharma company 3 years (CSL)

Academia

- Bioinformatics research position 4 years

STATISTICS IN BIOPHARMACEUTICAL RESEARCH
2022, VOL. 14, NO. 1, 42–51
https://doi.org/10.1080/19466315.2021.1979641

Taylor & Francis
Taylor & Francis Group

A Sequential Predictive Power Design for a COVID Vaccine Trial

Rajat Mukherjee[a], Pranav Yajnik[b], Natalia Muhlemann[b], and Caroline Morgan-Bouniol[b]

[a]Cytel, Strategic Consulting, Barcelona, Spain; [b]Cytel, Inc., PA

ABSTRACT

Medical investigations for therapeutics and vaccines for combating a pandemic such as COVID-19, call for flexible and adaptive trial designs that are capable of producing robust results amidst uncertainties. Here, we present a Bayesian sequential design to study the efficacy of Bacillus Calmette–Guérin (BCG) in providing protection against COVID-19 infections via its known "trained-immunity" mechanism. The main design consideration is to provide a framework to rapidly establish a proof-of-concept on the vaccine efficacy of BCG under a constantly evolving incidence rate and in the absence of prior efficacy data. The trial design is based on taking several interim looks and calculating the predictive power with the current cohort at each interim look. Decisions to stop the trial for futility or stopping enrollment for efficacy are made based on the current cohort predictive power computation. At any interim, if any of the above decisions cannot be taken then the study continues to enroll till the next interim look. Via extensive numerical studies, we show that the proposed design can achieve the desired frequentist operating characteristics, currently required by regulatory bodies while offering greater flexibility in terms of sample size and the ability to make robust interim decisions.

ARTICLE HISTORY
Received November 2020
Accepted September 2021

KEYWORDS
Bayesian adaptive designs; COVID-19; Frequentist operating characteristics; Predictive power; Vaccine trials

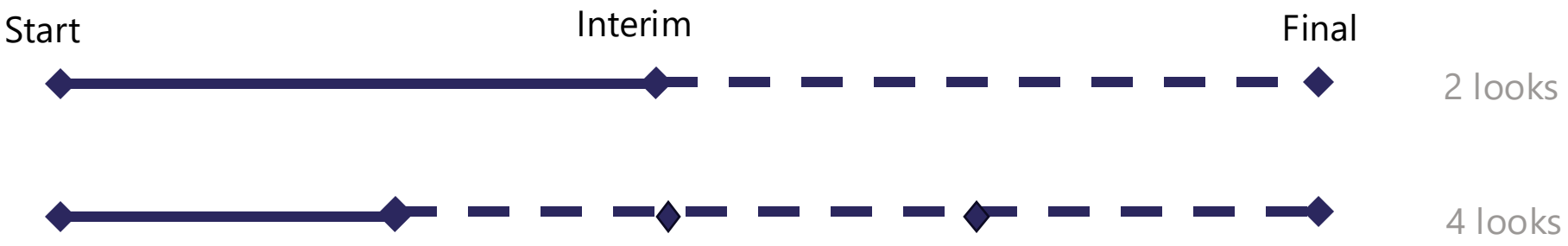
An Adaptive Sequential Bayeisan design

Cytel

Background

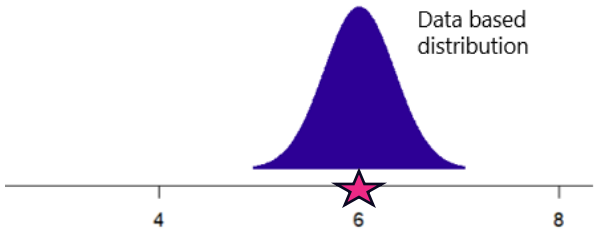
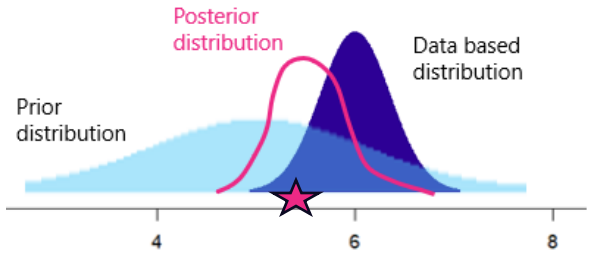
Confirmatory trials - Hypothesis testing framework

Evaluate the study sequentially at interim analyses



Interim data will be used to make predictive judgement on the final study outcome.

Two approaches to predict and evaluate study outcomes

	Frequentist (traditional)	Bayesian
Interim analysis e.g. for **S**ample-**S**ize **R**e-estimation	**Conditional power** Probability of positive study Given interim effect estimates 	**Predictive power** Probability of postitive study Given posterior interim effect distribution 
Intermediate: Assurance		
Final analysis	**P-value < 0.05 of treatment effect** Absence of any treatment effect is very unlikely, based on the study data	**Posterior probability of treatment effect > 0.95** Posterior probability of treatment effect based on the study data and prior distribution

Prior/posterior = Learn/confirm = Adaptive
Flat prior or borrow information

Cytel

Covid vaccine trial

Optimized for speed

Pfizer's RAM vaccine and I_SPY Covid 19 platform trials were both Bayesian designs.

Primary objective

Short term efficacy of BCG (Bacillus Calmette-Guérin), Proof of Concept
(BCG was developped against tuberculosis, there are other BCG studies on Covid, too.)
If successful, a long-term efficacy trial was planned.

Primary endpoint

Confirmed symptomatic Covid-19 (x = yes/no)
3 months follow-up after vaccination

π_B incidence proportion with BCG
π_C incidence proportion with control (placebo)

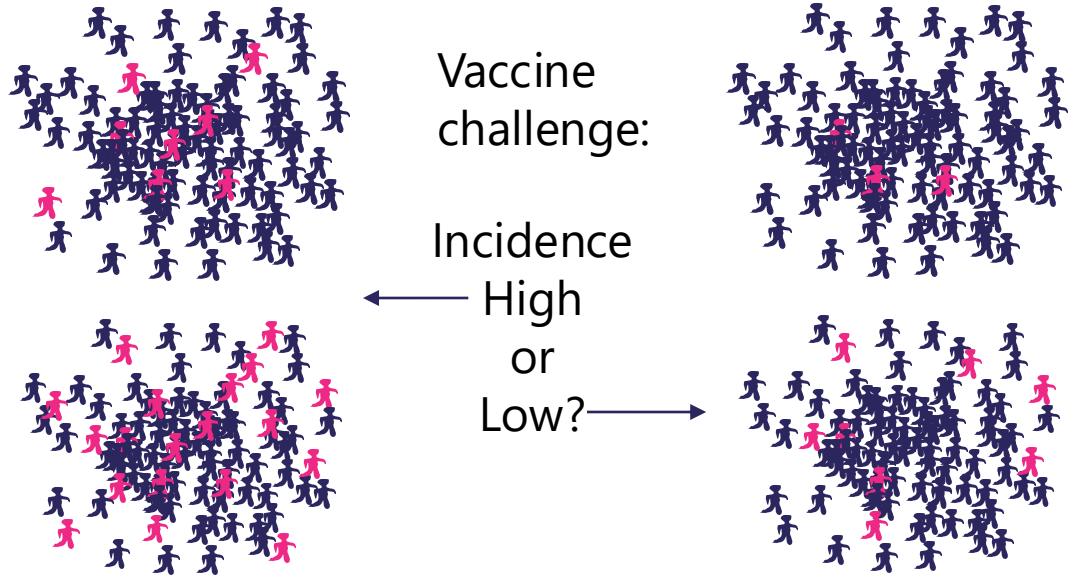
H_0: $\pi_B - \pi_C \geq 0$
H_A: $\pi_B - \pi_C < 0$

FDA prerequisites

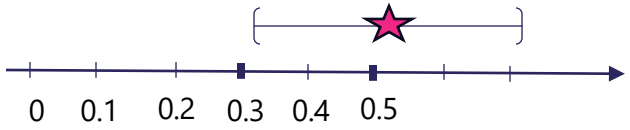
π_B incidence proportion with BCG
π_C incidence proportion with control (placebo)

H_0: $\pi_B - \pi_C \geq 0$
H_A: $\pi_B - \pi_C < 0$

VE = vaccine efficacy = $1 - \frac{\pi_B}{\pi_C}$ (will be > 0 if treatment effect)



FDA requirements:
$\widehat{VE} \geq 0.5$ reducing incidence by at least 50%
& $CI > 0.3$
(i.e. superiority margin of 0.3 -> large study)



One-sided α = 2.5%
90% power

Testing for difference in proportions with pooled variance, assuming π_C = 20%, with **0 superiority margin**
->
645 patients needed, 8% dropout -> **700 patients** randomized 1:1 to detect 50% reduction.

Lower exposure -> lower π_C, underpowered!

Assuming π_C = 6 – 8 % and VE = 70%
->
1200 patients needed

Head for a study design with flexible sample size!

Longer survival studies use hazard ratios. Here we have a fixed follow-up time.

An Adaptive Sequential Bayesian Design

Interim analyses throughout the study

Adaptive Sequential Bayesian

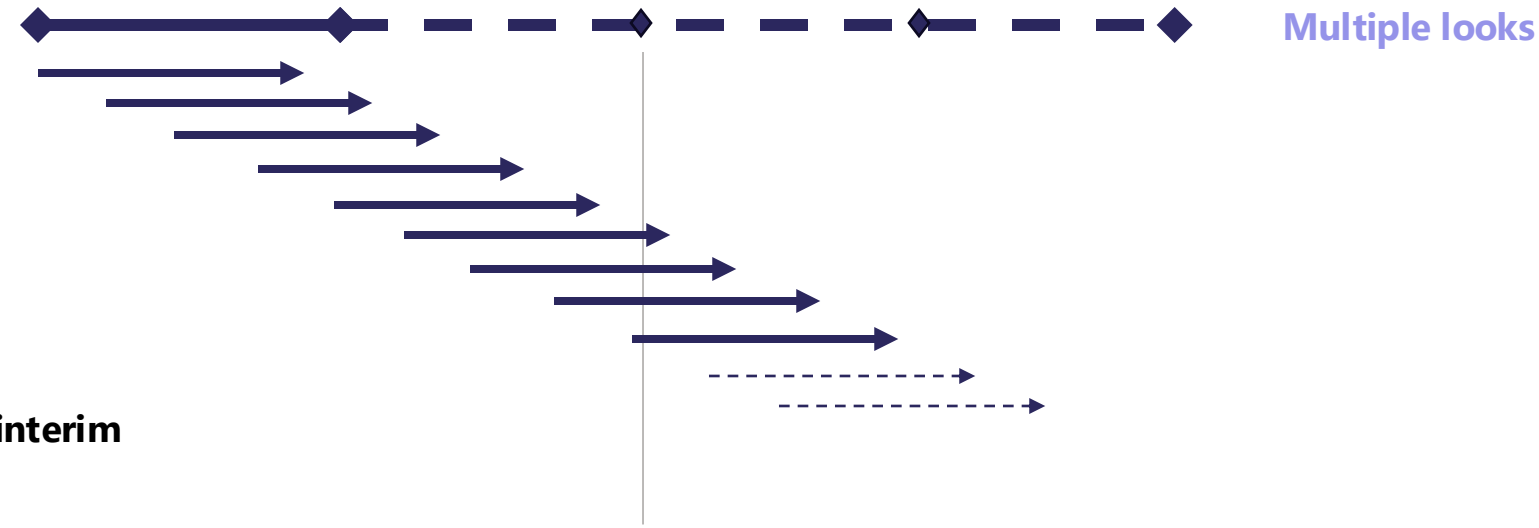
Adjust to learnings from data

Use predictive power and posterior distribution for decisions

Adaptive Sequential Bayesian – at it's extreme

What?

Interim analyses after say 30%, 40%, 50%, … have been enrolled.



Allow sequential possibility to adapt

- Stop accrual early for futility
- Stop accrual early for predicted efficacy
- Re-estimate sample size and continue

How?

Predict treatment effect of enrolled patients at interim

Early stopping boundaries on predictive power

- S success threshold 80-99%
- F futility threshold 2.5-20%

Note!

- Useful when small to moderate delay between enrollment and observing primary outcome
- Extensive simulations needed

Cytel

Study procedure

N_plan = 700 patients
N_max = 1200 patients

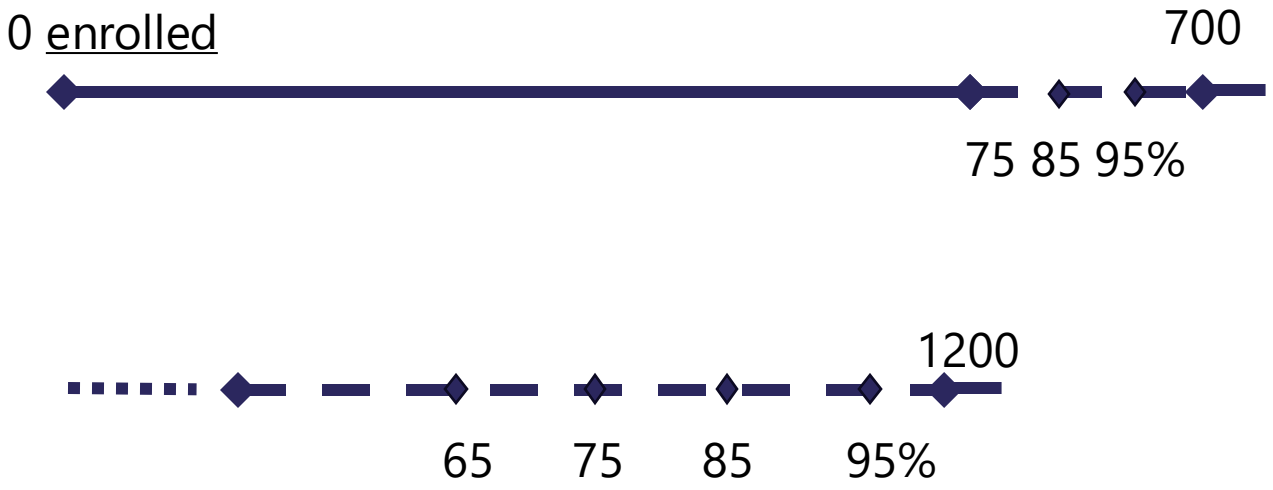
Plan the 1st interim look when 75% of N_plan patients have been enrolled (not finished), and subsequent interim looks at 10% increments.

At k:th interim look, derive Predictive Power (PP):
- If PP ≥ S = **90%** stop enrolling for efficacy
- If PP < F = **20%** stop enrolling for futility

At last interim look before N_plan recruited:
- If F < PP ≤ 0.5 carry out study as planned with N_plan = 700 patients.
- If 0.5 < PP < S increase sample size to N_max = 1200 patients and continue enrollment to next interim look at 65% of 1200.

➔ max 7 interim looks, at 75%, 85%, 95% of 700, then 65%, 75%, 85% and 95% of 1200.

Study is positive if postierior probability of $H_A > \gamma$.



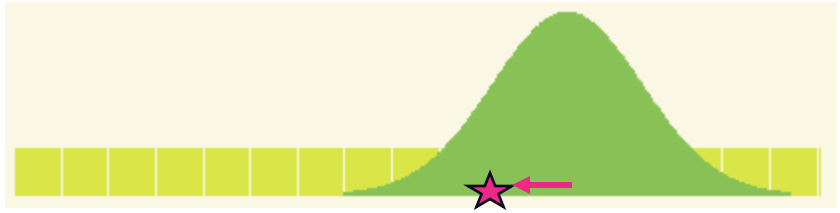
Bayesian analysis

π_B incidence proportion with BCG
π_C incidence proportion with control (placebo)

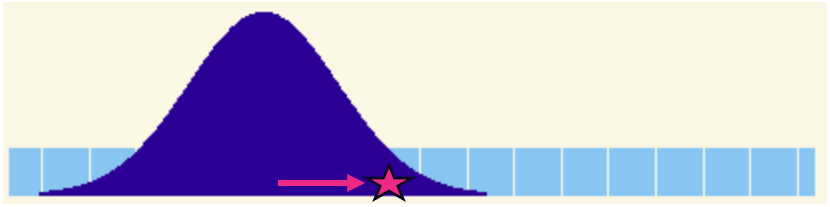
H_0: $\pi_B - \pi_C \geq 0$
H_A: $\pi_B - \pi_C < 0$

Reject if the postierior probability $Pr(H_A|Data) > \gamma$, where γ is set by simulation.

Beta Binomial conjugate prior-posterior derivation:
Each $\pi \sim \beta(\alpha = 1, \beta = 1)$ (uniform prior)
Each x$\sim Bin(n, \pi)$
->
Posterior distribution $(\pi|x) \sim \beta(\alpha + x, \beta + n - x)$



π_C estimate



π_B estimate

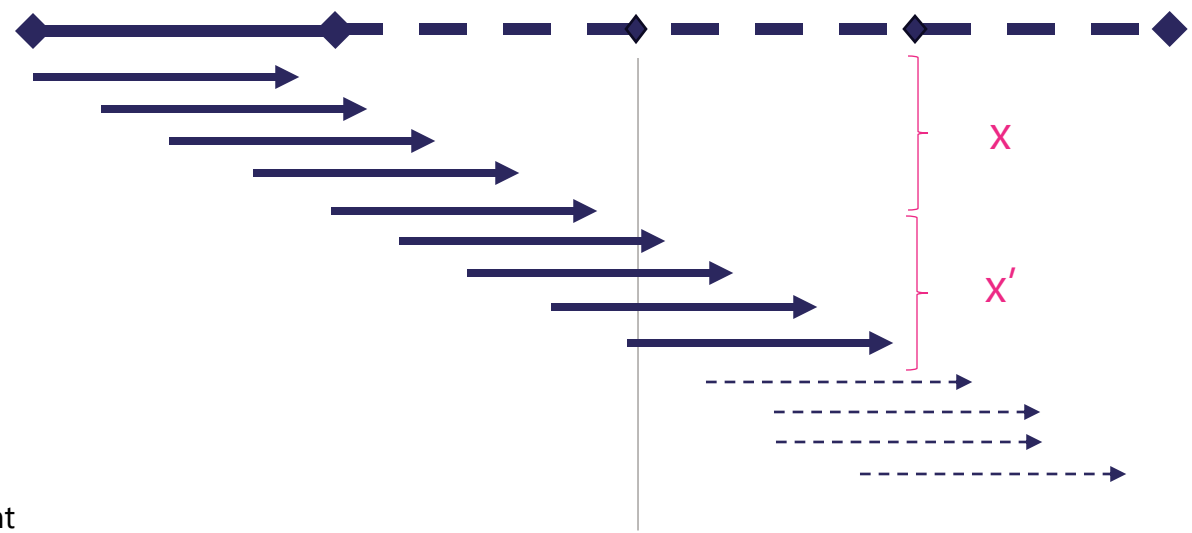
More advanced versions: include observations taken so far on started patients (dichotomous or survival analysis)

Posterior disributions are pulled closer to eachother - conservative!

Multiple testing/type I error control

No study evaluation at interim:

- we look at PP of those enrolled, not at the posterior distribution of all patients.
- Outcome of x′ is predicted, not known.



Derive Predictive Power PP:

$$PP = E_{p(x'|x)}\{Pr(H_A|\underline{x}, \underline{x}') > \gamma\}$$

Posterior predictive distribution

Data so far
x finished patient
x′ started but unfinished patient

Upon enrollment stop, the last enrolled patients will be finished before final study evaluation.

➔ Multiple testing issue less pronounced, we earn some alpha ← Type I error risk, false positive risk

Use simulations to set γ that preserves Type I error risk < 2.5% one-sided.

No "special" rules by authorities for Bayesian studies, they need to adhere to the same risk limits as traditional studies.

Evaluations

Simulation results (1/3)

Table 1. Simulation results under the null scenario with 0% vaccine efficacy (BCG vs. Control), varying recruitment rate (Rec.) and success criteria γ. Up to 7 interim looks, $\eta = 0.9$, $\eta_1 = 0.5$, and $\eta_f = 0.2$.

	Success criteria (γ)	No. of interim looks mean (SD)	Sample size min, mean(SD), med, max	Prob. early efficacy stop enroll.	Prob. early futility stop	Prob. sample size inc. ($N = 1200$)	Type I error rate	
Rec.: 50								
	0.975	1.22(0.76)	525, 705.79(68.34), 700, 1200	0.0236	0.9471	0.0217	0.02777	Not OK
	0.977	1.21(0.74)	525, 705.50(66.37), 700, 1200	0.0220	0.9501	0.0204	0.02568	
	0.978	1.20(0.74)	525, 705.40(65.49), 700, 1200	0.0211	0.9518	0.0200	0.02438	
	0.980	1.19(0.72)	525, 705.16(63.63), 700, 1200	0.0195	0.9550	0.0188	0.02221	
Rec.: 100								
	0.975	1.47(1.12)	525, 716.68(98.87), 700, 1200	0.0239	0.9019	0.0451	0.02688	Not OK
	0.977	1.45(1.11)	525, 716.14(97.00), 700, 1200	0.0224	0.9060	0.0433	0.02494	
	0.978	1.45(1.10)	525, 715.87(96.03), 700, 1200	0.0216	0.9081	0.0424	0.02393	
	0.980	1.43(1.07)	525, 715.05(93.50), 700, 1200	0.0203	0.9126	0.0401	0.02197	
Rec.: 150								
	0.975	1.86(1.41)	525, 730.13(129.96), 700, 1200	0.0350	0.8271	0.0760	0.02679	Not OK
	0.977	1.84(1.39)	525, 729.22(128.01), 700, 1200	0.0335	0.8321	0.0734	0.02468	
	0.978	1.83(1.38)	525, 728.77(127.02), 700, 1200	0.0327	0.8348	0.0721	0.02365	
	0.980	1.81(1.36)	525, 727.90(125.01), 700, 1200	0.0311	0.8406	0.0697	0.02178	

NOTE: Results are summarized over 200,000 simulated trials.

Type I error rate must be < 2.5%

Cytel

Simulation results (2/3)

Table 2. Simulation results under the null scenario with 0% vaccine efficacy (BCG vs. Control), varying maximum number of interim looks (Max IAs) and recruitment rate (Rec.) per month.

	Rec.	No. of interim looks mean (SD)	Sample size min, mean(SD), med, max	Prob. early efficacy stop enroll.	Prob. early futility stop	Prob. sample size inc. ($N = 1200$)	Type I error rate
Max IAs: 2							
	50	1.04(0.19)	525, 716.52(95.82), 700, 1200	0.0124	0.9139	0.0391	0.02266
	100	1.07(0.26)	525, 733.34(129.46), 700, 1200	0.0129	0.8262	0.0733	0.02235
	150	1.15(0.36)	525, 770.26(179.58), 700, 1200	0.0204	0.6882	0.1500	0.02270
Max IAs: 4							
	50	1.13(0.45)	525, 707.84(72.15), 700, 1200	0.0190	0.9476	0.0227	0.02331
	100	1.27(0.63)	525, 718.68(101.55), 700, 1200	0.0194	0.8989	0.0452	0.02292
	150	1.51(0.77)	525, 733.00(132.74), 700, 1200	0.0285	0.8174	0.0771	0.02271

NOTE: Success criteria $\gamma = 0.978$, $\eta = 0.9$, $\eta_1 = 0.5$, and $\eta_f = 0.2$. Results are summarized over 200,000 simulated trials.

OK

Type I error rate must be < 2.5%

Simulation results (3/3)

Table 3. Simulation results under design alternatives with 50% vaccine efficacy (VE, BCG vs. Control), varying maximum number of interim looks (Max IAs) and recruitment rate (Rec.) per month. Success criteria $\gamma = 0.978$, $\eta = 0.9$, $\eta_1 = 0.5$, and $\eta_f = 0.2$.

	Rec.	No. of interim looks mean (SD)	Sample size min, mean(SD), med, max	Prob. early efficacy stop enroll.	Prob. early futility stop	Prob. sample size inc. ($N = 1200$)	Power
Max IAs: 7							
	50	1.64(1.17)	525, 596.48(114.03), 525, 1200	0.7695	0.1149	0.0889	0.91360
	100	2.00(1.48)	525, 622.49(143.46), 525, 1200	0.7274	0.0723	0.1613	0.92110
	150	2.25(1.62)	525, 643.32(160.60), 595, 1200	0.6999	0.0522	0.2009	0.92040
Max IAs: 4							
	50	1.42(0.73)	525, 614.88(139.53), 525, 1200	0.7593	0.1068	0.1015	0.91760
	100	1.64(0.88)	525, 648.17(172.23), 525, 1200	0.7161	0.0640	0.1746	0.92580
	150	1.80(0.95)	525, 675.99(189.87), 665, 1200	0.6805	0.0463	0.2215	0.92630
Max IAs: 2							
	50	1.21(0.40)	525, 660.06(196.84), 525, 1200	0.6063	0.0928	0.2065	0.92220
	100	1.31(0.46)	525, 707.40(225.74), 525, 1200	0.5364	0.0518	0.3078	0.93040
	150	1.39(0.49)	525, 747.86(237.79), 700, 1200	0.4626	0.0355	0.3871	0.93160

NOTE: Results are summarized over 10,000 simulated trials.

More flexible

Less flexible

Lower power

Higher power

OK

Power must be > 90%

Compare to Group Sequential design
(Lan-DeMets alpha spending with O'Brien-Flemming boundaries)

	GSD-OB	ASB
VE = 50% π_C = 20% 100 patients/month	97% power SS average 1005 P(early stopping) = 98% More effective	92% power SS average 622 P(early stopping) = 72%
VE = 70% π_C = 6% 100 patients/month	85% power SS average 1199	90% power SS average 666 More effective

Up to 8 looks including the final
Max 1200 patients
Overall alpha 2.5% one-sided.
1st interim look at 30% of 1200 completed/enrolled
Subsequent looks every 10%

$\gamma = 0.978, \eta = 0.9, \eta_1 = 0.5$, and $\eta_f = 0.2$.
strictly non-binding

Adaptive Sequential Bayesian stops earlier when motivated!

Cytel

Wrap up

An Adaptive Sequential Bayesian Design for a COVID Vaccine Trial

A Goldilocks approach to sample size selection – not too big, not too small

1. Sample size will be taylored to vaccine incidence, to be adequate even with low spreading of the disease.
2. A "temporary high" in efficacy has a chance to be corrected in the final analysis including also the last enrolled patients.
3. Only recruitment rates of 50 – 150 patients per month. Check from case to case!
4. Infinite variations possible: add new arms, historical data in informative priors, …
5. Wild and crazy design powered to prove Proof of Concept, i.e. not pivotal.
    1. Does not show VE>0.3, and 3 months is too little safety.
    2. If $\widehat{VE} > 0.5$ a confirmatory trial will follow!
6. This trial could inform a confirmatory trial as informative priors.
7. Multiple looks -> heavy operational burden. Keep interim results blinded to investigators, investors, sponsors -> preserve the intregrity of trial results.
8. Bayesian analysis – beware, in clinical trias we strive to be objective!

Some adaptive sequential designs successfully implemented with FDA

Julian, T. B., Blumencranz, P., Deck, K., Whitworth, P., Berry, D. A., Berry, S. M., Rosenberg, A., et al. (2008). Novel intraoperative molecular test for sentinel lymph node metastases in patients with early-stage breast cancer. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 26(20):3338–3345.

White, W. B., Grady, D., Giudice, L. C., Berry, S. M., Zborowski, J., Snabes, M. C. (2012). A cardiovascular safety study of LibiGel (testosterone gel) in postmenopausal women with elevated cardiovascular risk and hypoactive sexual desire disorder. *American Heart Journal* 163(1):27–32.

Wilber, D. J., Pappone, C., Neuzil, P., De Paola, A., Marchlinski, F., Natale, A., Macle, L., Daoud, E. G., Calkins, H., Hall, B., Reddy, V., Augello, G., Reynolds, M. R., Vinekar, C., Liu, C. Y., Berry, S. M., Berry, D. A., for the ThermoCool AF Trial Investigators. (2010). Comparison of antiarrhythmic drug therapy and radiofrequency catheter ablation in patients with paroxysmal atrial fibrillation: A randomized controlled trial. *Journal of the American Medical Association* 303(4):333–340.

Cytel
Sweden

Contact us and we will solve it!

Biostatistics & Adaptive Trial Design
Regulatory Strategy
Clinical Pharmacology
RWE & RWD
Clinical Project Management & Sponsor Oversight
Clinical Development Strategy & Planning
Medical Writing
Pharmacutical Development & CMC