BAYESIAN STUDY DESIGN TO LEVERAGE INFORMATIVE PRIOR INFORMATION

2 Trial Design Case Studies

Jim Bolognese, Sam Hsiao, Cytel Inc.

Adele Morganti, Peter Cornelisse, Matthieu Villeneuve,

& Marisa Bacchi, Actelion

26-Sept-2017

BAYESIAN STUDY DESIGN TO LEVERAGE INFORMATIVE PRIOR INFORMATION - *OUTLINE*

- Trial 1
  - Bayesian Design with Adaptive Option
    - Rationale
    - Summary of Design
  - Summary of Performance Characteristics via Simulation
- Trial 2
  - Leverage what was learned from Trial 1
  - Overcome limitations found in Trial 1
- Remarks
- Questions / Comments / Discussion

INTRODUCTION AND PURPOSE

- Key Efficacy Objective:
  - Test treatment (Active) superior to standard-of-care (SoC) in rare paediatric disease
    - Limit exposure to new test treatment
- Primary Endpoint:
  - Time to Clinical Worsening (morbidity/mortality)
- Leverage prior Patient Registry data on SoC to lower sample size requirements via Bayesian methods

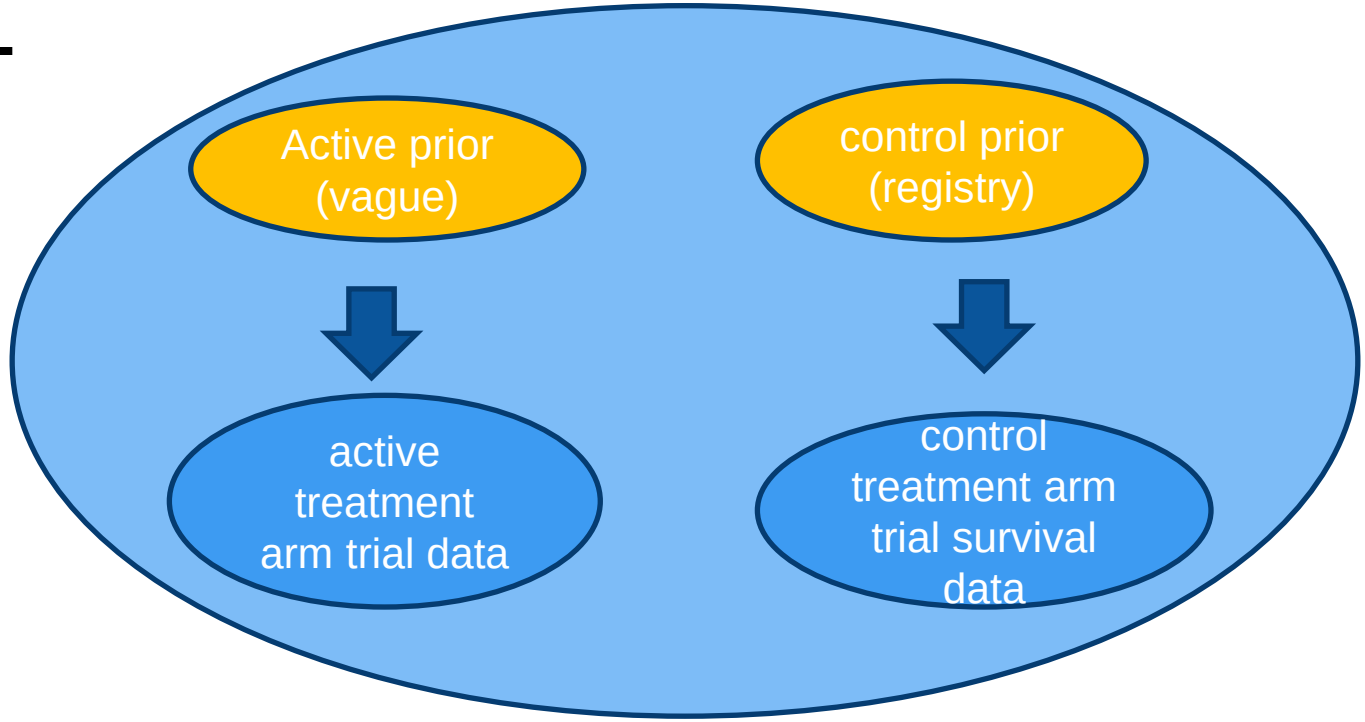
BAYESIAN STATISTICAL APPROACH – STUDY DESIGN

PRIOR
$p(\delta)$

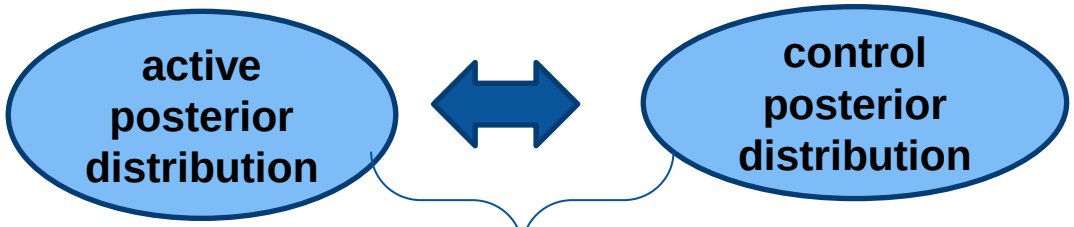
SoC: pediatric registry

LIKELIHOOD
$\ell(t \mid \delta)$

Accumulated trial data



POSTERIOR
$p(\delta \mid t) \propto p(\delta)\, \ell(t \mid \delta)$



Efficacy Criteria: Posterior Probability (TRUE HR < 1) > 0.975

Cytel

NORMAL APPROXIMATION TO LOG(HR)

- Likelihood: log(HR)~normal
  - Mean = log(HR) = log(H1) - log(H2)
    - H1, H2 = hazard for treatment & SoC
  - Variance = (1/E1)+(1/E2)
    - E1, E2 are # events treatment & SoC
- WOLOG prior SoC H2=0; treatment H1=0 or log(HR)
- Variance(H2) = 1/77; Variance(H1)=1/1
  - 77 chosen to be somewhat less than the expected number of events for SoC in the trial
- Posterior mean=log(H1) – log((E2*H2+77*Hsoc)/(E2+77))
  - Variance = (1/(E2+77))+(1/E1)

SUMMARY OF PROPOSED DESIGNS (~80% POWER; 0.025, 1-SIDED; TRUE HR=0.65)

Standard Frequentist superiority	Bayesian Informative Prior	Bayesian adaptive
180 events / ~300 subjects	136 events / ~224 subjects	136 to 180 events/ ~224 to ~300 subjects
study duration: ~6 years	study duration: ~5 years	study duration: ~5 to ~6 years

- Requires assumption that prior and trial SoC hazard rates equal
- If not, need to compute success criteria cut-off value via simulation to control type 1 error

SUCCESS CRITERIA CUT-OFFS TO CONTROL TYPE 1 ERROR AT 2.5%

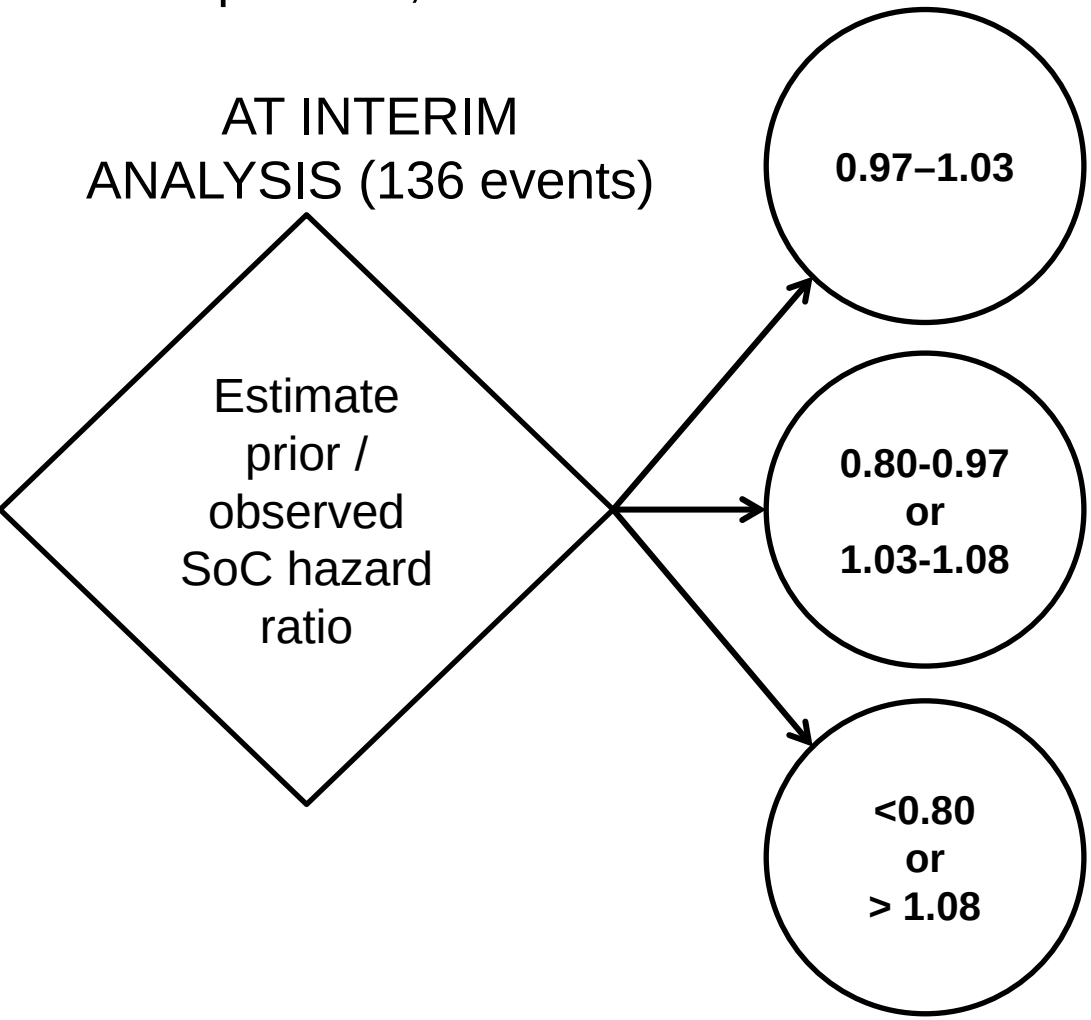
Performance Characteristics based on 10,000 simulations per scenario for the Bayesian informative prior (with prior HR=1)		
Ratio TRUE prior/RCT SoC event rates	bound required for alpha<0.025 (Bayesian success criteria)	Number of events for ~80% power
1	97.50%	136
1.02	97.50%	136
1.05	98.5%	160
1.1	99%	>180
1.2	99.50%	>180
1.3	99.90%	>180
1.4	99.90%	>180
1.5	99.99%	>180
2	>99.99%	>180

Cytel

BAYESIAN ADAPTIVE DESIGN OPTION: EXAMPLE

Based on the level of departure from homogeneity, the following actions were planned, evaluated based on simulation:

AT INTERIM ANALYSIS (136 events)



- Stop study at 136 events
- Primary analysis: Informative Bayesian

- Stop study at 160 events
- Primary analysis: Informative Bayesian

- Stop study at 180 events
- Primary analysis: UN-informative Bayesian

SIMULATION RESULTS

SoC ratio (trial / prior)	success criteria met (% of 10,000 simulations) TRUE HR=0.65 (power)	TRUE HR=1 (type 1 error)
0.5	78.6	2.1
0.75	60.8	2.2
0.8	61.2	1.7
0.95	75.7	1
1	77.5	1.5
1.05	79.2	2
1.25	80.3	2.7
1.33	80.5	2.4
2	81.2	2.6

TRIAL 2 – TWO SOURCES OF PRIOR SOC DATA REGISTRY & TRIAL 1 SOC GROUP!!

- **Borrowing rule based on 1 homogeneity threshold.**
  - Borrow 38 historical events if ≥ 1 historical SoC group has log(HR) within 0.05 of the RCT SoC group; otherwise borrow no events.
  - Randomization switches from 1:1 to 5:1 (test:SoC) if there is borrowing. Use an informative prior.
- **Borrowing rule based on 2 homogeneity thresholds.**
  - Borrow 38 historical events if ≥ 1 historical SoC group has log(HR) within 0.05 of the RCT control group; otherwise borrow 19 events if ≥ 1 historical SoC group has log(HR) within 0.1 of the RCT group;
  - otherwise borrow no events.
  - Randomization switches from 1:1 to 5:1 if there is borrowing.
  - Use an informative prior distribution.
- **No borrowing.** Randomization remains 1:1.
  - Use a vague prior distribution for the log(HR) of the control group.

TRIAL 2 – TWO SOURCES OF PRIOR SOC DATA REGISTRY & TRIAL 1 SOC GROUP!! (SIMULATION RESULTS – 1)

- Under most scenarios, power decreases as #events for the homogeneity assessment increases.
  - most extreme drop was 3% as the homogeneity assessment shifts from 40 to 96 events, but generally the change is modest at 0-2%.
- Inflation of type 1 error evident in several scenarios, notably when historical controls have much higher observed hazard rates than the TRUE hazard rate of RCT controls.
  - Inflation more pronounced under Rule #2, which has greater borrowing than under Rule #1.
  - Borrowing historical events decreases #patients enrolled, particularly #patients randomized to the (potentially inferior) control group.
  - Fewer patients randomized to control group under Rule #2 compared with Rule #1, not surprising since more events are borrowed under Rule #2.

Cytel

TRIAL 2 – TWO SOURCES OF PRIOR SOC DATA REGISTRY & TRIAL 1 SOC GROUP!! (SIMULATION RESULTS – 2)

- performing the homogeneity assessment earlier leads to fewer patients randomized to the control group.
- 5:1 allocation after homogeneity assessment results in fewer patients allocated to SoC compared with keeping 1:1 allocation throughout study.
- randomization ratio has most impact when homogeneity assessment performed at 40 events
  - impact seems negligible when the homogeneity assessment performed at 96 events.
- study duration tends to be slightly shorter (by up to 3 months) if the homogeneity assessment is performed later.

Cytel

CONCLUSIONS

Extensive simulations and interactions with HAs were needed to refine the performance characteristics of Bayesian designs

In summary, due to:

1. Time needed for Bayesian adaptive design optimization and HAs acceptance

2. The proposed Bayesian adaptive design with 1 prior data source yielded sample sizes potentially similar to group sequential due to:

- strict type I error control
- potential SoC heterogeneity
- only one source of prior data (rare disease and new endpoint)

Actelion chose classical group sequential design

▶ Bayesian adaptive design with 2 prior data sources seems promising – Useful in 2nd Trial

▶ Complexity of implementation, and no precedent in pediatric PAH

WHAT DID WE LEARN?

1. HAs are not yet fully comfortable with Bayesian approaches outside of the device world

2. HAs were in favor of using Bayesian approach as sensitivity analysis for phase III studies in children

3. Extensive simulation work required for HAs interactions to document methods and performance characteristics

4. Case specific limitation, only one source of prior data for control was available (Trial 1 example).
    1. When >1 prior source, Bayesian approach seems workable (Trial 2 example)

WHAT DO WE RECOMMEND FOR THE FUTURE?

- Bayesian and adaptive designs should continue to be considered:
  - Historical data for prior can be used if Pocock(1976) criteria are met
  - Accounting for interim heterogeneity checks (with/without subject matching)
  - Plan ahead for extensive simulations to document performance characteristics
  - ➢ Need for clear and global guidance (and metrics) from HAs on what is recommended for a successful application of Bayesian designs beyond devices

QUESTIONS / COMMENTS / DISCUSSION

Jim Bolognese (bolognese@cytel.com)

[supplemental info follows this slide]

STATISTICAL CHALLENGES FOR GLOBAL PROGRAM

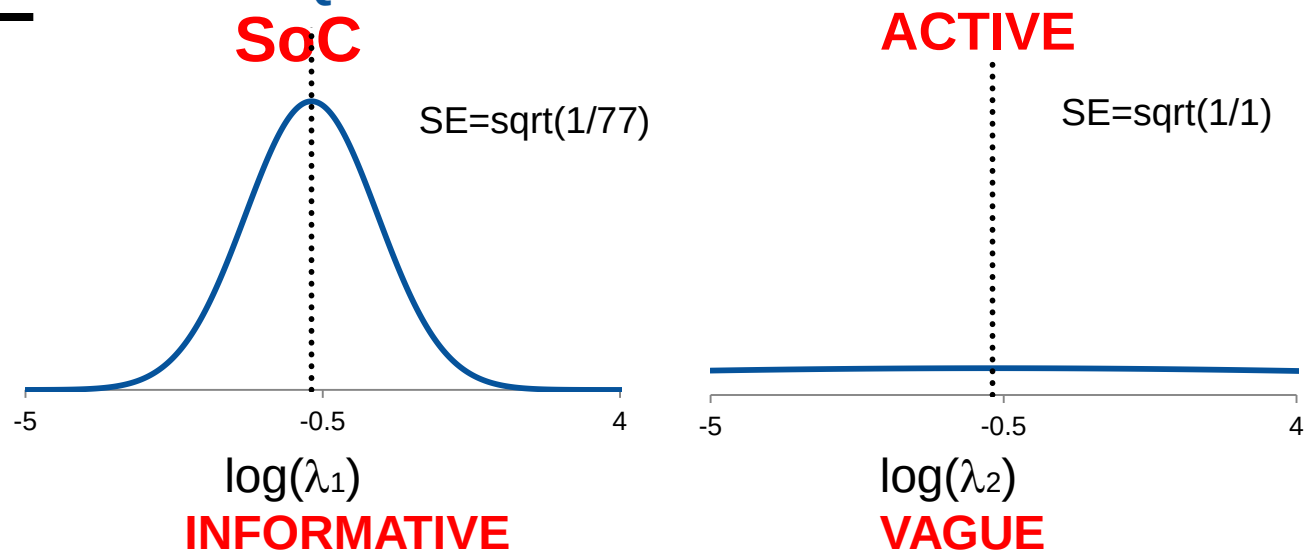
Challenges based on HA discussions:

- Relaxing the “type I” error rate is not acceptable to all HAs
- Bayesian informative approach has potential for type I error inflation in case of heterogeneity between prior and observed SoC hazard rates
- In addition, for a global study:
  - Conclusiveness was required for clear product labeling statements. A negative study needs to rule out a minimum clinically important difference (MCID)

BAYESIAN INFORMATIVE PRIOR

WITH TREATMENT EFFECT EQUAL TO NULL

PRIOR for
log(HR) =
$\log(\lambda_1) - \log(\lambda_2)$



Asymptotic Normal distribution approximation of log (HR) allows use of closed form Bayesian inference

Registry SoC hazard rate is used for the prior

Success Criterion: Posterior Probability (HR < 1) > 0.95

For the prior, treatment effect was assumed to be HR = 1 (null)

BAYESIAN STATISTICAL APPROACH

- For closed form Bayesian inference for survival endpoints (i.e. known probability distribution for $p(\delta \mid t) \propto \ell(t \mid \delta)\, p(\delta)$ the Schoenfeld asymptotic (1981) approach is introduced such :
  - For the prior $p(\delta)$,

    δ = log(hazard ratio) follows a normal distribution
  - Accumulated data are expressed as hazard rates ($\lambda_1(t_{1j})$ and $\lambda_2(t_{2j})$) for the active and control groups, respectively, with survival times t_{ij} observed for each patient of the i_{th} group with:
    - $\ell(t \mid \delta) = \ell(\log\{\lambda_1(t_{1j}) / \lambda_2(t_{2j})\} \mid \delta) = \ell(\log\{\lambda_1(t_{1j})\} - \log\{\lambda_2(t_{2j})\} \mid \delta)$ follows a normal distribution asymptotically

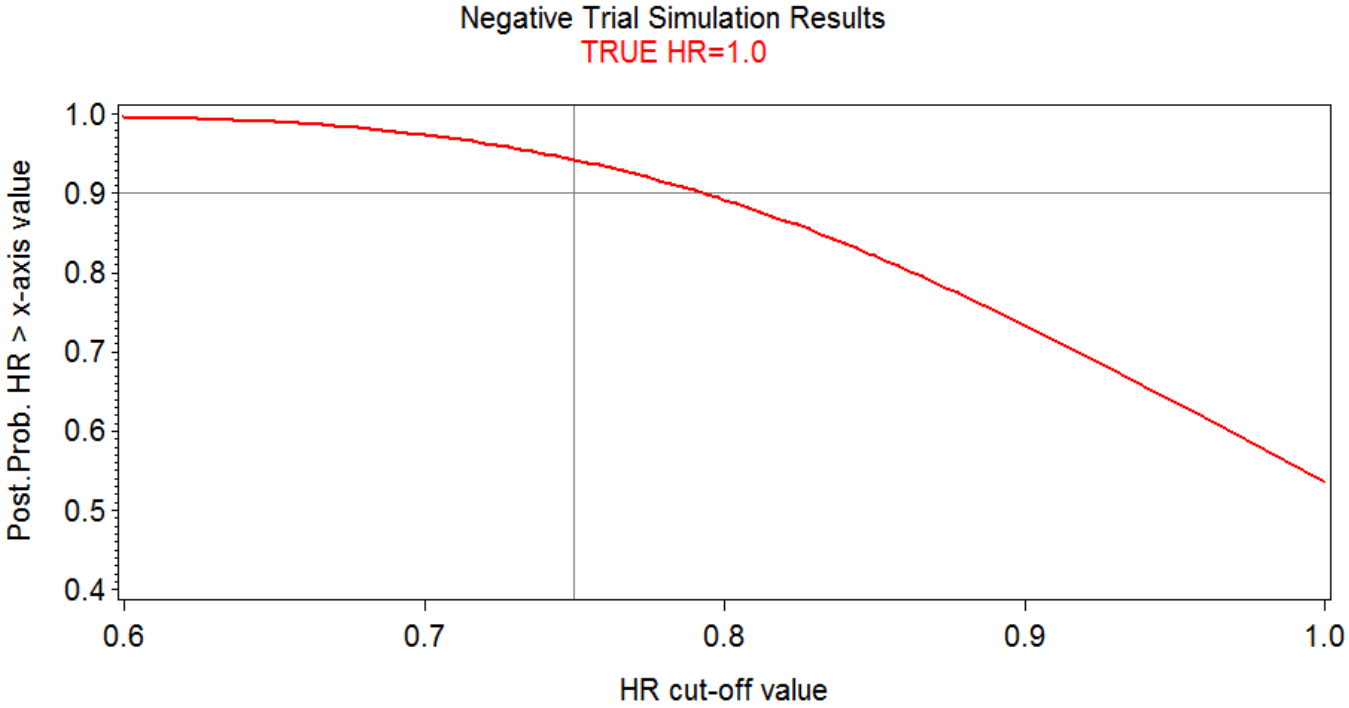
Following this asymptotic approximation for the parameter of interest, the data and the prior follow normal conjugate distributions and as such lead to a posterior asymptotic normal distribution.

CONCLUSIVENESS OF THE BAYESIAN ADAPTIVE APPROACH

- Sample size is already computed with adequate power for expected treatment effect
- However, in case non-significant treatment difference, it is important to provide a reasonable probability to rule out a smaller treatment effect in case the underlying HR=1 to fulfill the requirement of a written request (WR) from FDA
- Simulations were carried out to look at conclusiveness, which we refer to as probability to rule out the MCID

NEGATIVE, BUT STILL CONCLUSIVE RESULTS

Simulations were conducted to investigate conclusiveness for the adaptive Bayesian design. *Simulated results which yielded a negative outcome are summarized*:



If true HR=1: posterior probability (HR>0.75 MCID) ~94%

posterior probability (HR>0.80) ~ 89%

- E.g., “Test treatment highly likely to have negligible effect”

TRIAL 2 – TWO SOURCES OF PRIOR SOC DATA REGISTRY & TRIAL 1 SOC GROUP!! (CONTINUED)

- To assess how the timing of the homogeneity assessment affects performance, designs with homogeneity assessments occurring at 40, 67, or 96 total events were considered.

- After the homogeneity assessment, the randomization ratio in the borrowing designs is switched from 1:1 to 5:1. To assess the impact of this choice of randomization ratio (versus, say, a less extreme ratio such as 2:1 or 3:1), simulations were performed in which the ratio remains at 1:1 throughout the trial. Thus, performance under intermediate ratios such as 2:1 or 3:1 can be roughly approximated by interpolation between the 5:1 and 1:1 scenarios.