EXTRAPOLATION CHALLENGES IN PEDIATRIC PAH

Possible solutions for a feasible, global study

M. Bacchi, A. Morganti, P. Cornelisse, J. Bolognese, C. Lesage, A. Kümmel, P. Nilsson

PEDIATRIC PAH – THE SETTING

- Rare and progressive disease
  - Annual incidence 2-3/million. Global pediatric development program needed for feasibility
  - Progression to right heart failure and death. Progression is highly predictive of increased risk of death
  - Disease characteristics and response to treatment similar to adults
- No medicinal product globally approved
  - Medicines approved for adults regularly used in children
  - Limited scope for placebo-controlled studies

PEDIATRIC PAH – THE SETTING (2)

- No PD/intermediate endpoint that can be defined across pediatric subsets
  - Effect on pulmonary vascular resistance requires invasive approach, unacceptable in children (nowadays)
  - Exercise capacity can only be assessed in developmentally able children
- Demonstration of efficacy in children would have to be based on outcome events (morbidity/mortality)

DRUG TO BE INVESTIGATED – MACITENTAN

- Well-proven mode of action in adult PAH
  - Dual endothelin receptor antagonist (ERA)
- Well-characterized pharmacokinetics
  - PK profiles and inter-subject variability in adult PAH patients from large Phase 3 study well-described by one compartmental population model with linear elimination and first order absorption
- First PAH medicine to demonstrate benefit on outcome events (morbidity/mortality), in monotherapy and as add-on
  - Data suggest best-in-class characteristics

PEDIATRIC PROGRAM OBJECTIVE AND CHALLENGES

- Aim for one single, global pivotal study
  - Macitentan alone or as add-on is better than standard-of-care (SoC) for morbidity/mortality (M/M) outcome in pediatric PAH
    - This will be the first study to provide outcome data in children with PAH
    - Data from this study would further substantiate extrapolation from adult to pediatric PAH treatment effects of PAH medicines
- How to make this feasible in a global program?

STANDARD SUPERIORITY DESIGN, EVENT-DRIVEN

Sample size for a standard, phase III study randomized (1:1), active (SoC) controlled, event-driven, superiority design

Type I error (1-sided)	
0.05	0.025
for HR= 0.65, 138 events for a power of 80% ~250 subjects	for HR= 0.65, 180 events for a power of 80% ~300 subjects
number of subjects assumes a predefined enrolment rate and study duration (~6 years)	

HR= hazard ratio

Feasibility concerns

SOME POSSIBLE STATISTICAL SOLUTIONS

1. Change/relax the type I error to 0.05 1-sided

2. Make use of external information via Bayesian approach:
   - Largest, multiregional, longitudinal pediatric PAH registry provides contemporary, long term and ongoing control information (SoC)
   - A Bayesian approach using an informative prior for log(HR). For closed form Bayesian inference for survival endpoints, consider the Schoenfeld normal approximation of the log(HR)

[Berger 2012] Clinical features of paediatric pulmonary hypertension: a registry study. The Lancet 2012; DOI:10.1016/S0140-6736(11)61621-8
[Schoenfeld 1981] Schoenfeld D. The Asymptotic Properties of Nonparametric Tests for Comparing Survival Distributions. Biometrika. 1981; 68(1):316-319

SUITABILITY OF PEDIATRIC PAH REGISTRY

TOPP Registry (Berger)

Age at diagnosis: 3 months to 18 yrs
WHO group PH: 1*, 3, 4, 5
*Including CHD with residual PH following surgery, with no residual left side obstruction (PCWP mean <= 11mmHg)

Variables and inclusion criteria of the registry sufficiently cover the target population

Hard endpoints

Fit with Pocock criteria for use of external control

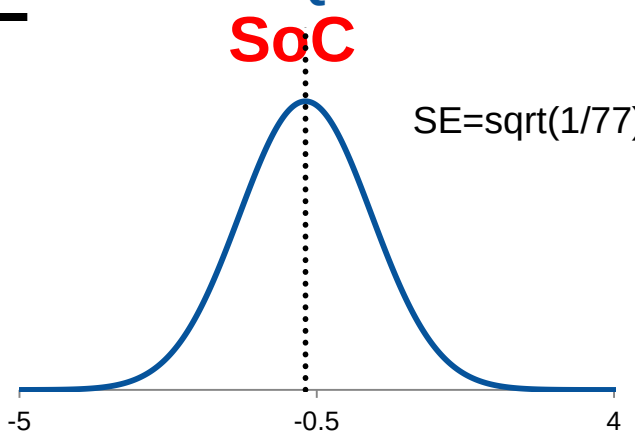
- Same SoC treatments
- Contemporary (2008→)
- Long-term composite endpoint (CHMP)
- WHO group 1, age: 2 years to 18yrs
- Similar geographic landscape
- Expected similarity of results between the randomized and external controls

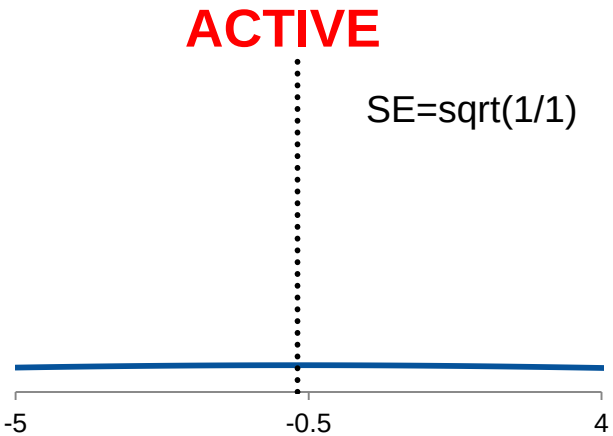
[Pocock 1975] Pocock S. The Combination of Randomized and Historical Controls in Clinical Trials. J.Chron Dis 1976; Vol 29:175-178

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)

SAMPLE SIZE (BAYESIAN INFORMATIVE DESIGN)

Sample size estimation was performed using simulations to yield:

– 80% of cases meeting the efficacy criteria for true HR=0.65

Success criterion for posterior probability	
>95%	>97.5%
if true HR= 0.65, 96 events for 80% probability of success ~180 subjects	if true HR= 0.65, 136 events for 80% probability of success ~224 subjects
assumes a predefined enrolment rate and study duration (~5 years)	

HR= hazard ratio

Initial proposal to HAs

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)

HETERONOGENEITY: TYPE I/II ERRORS INFLATION

Type I Error Inflation (HAs concern):

- The type I error can be inflated by incorporating a prior distribution with higher SoC hazard rate than observed in the study
  - Based on this, a strict control of type I error of 0.025 1-sided (HA requirement) was needed

Type II Error Inflation (Sponsor concern) **:**

- The type II error can be inflated by incorporating a prior distribution with lower SoC hazard rate than observed in the study

POTENTIAL SOLUTION: SUBJECT MATCHING

To limit the concerns of heterogeneity between SOC rates:

Subject matching was proposed by Dr Wang (FDA) to mininize heterogeneity of SoC hazard rates between prior (registry) and observed in the study (King 2015)

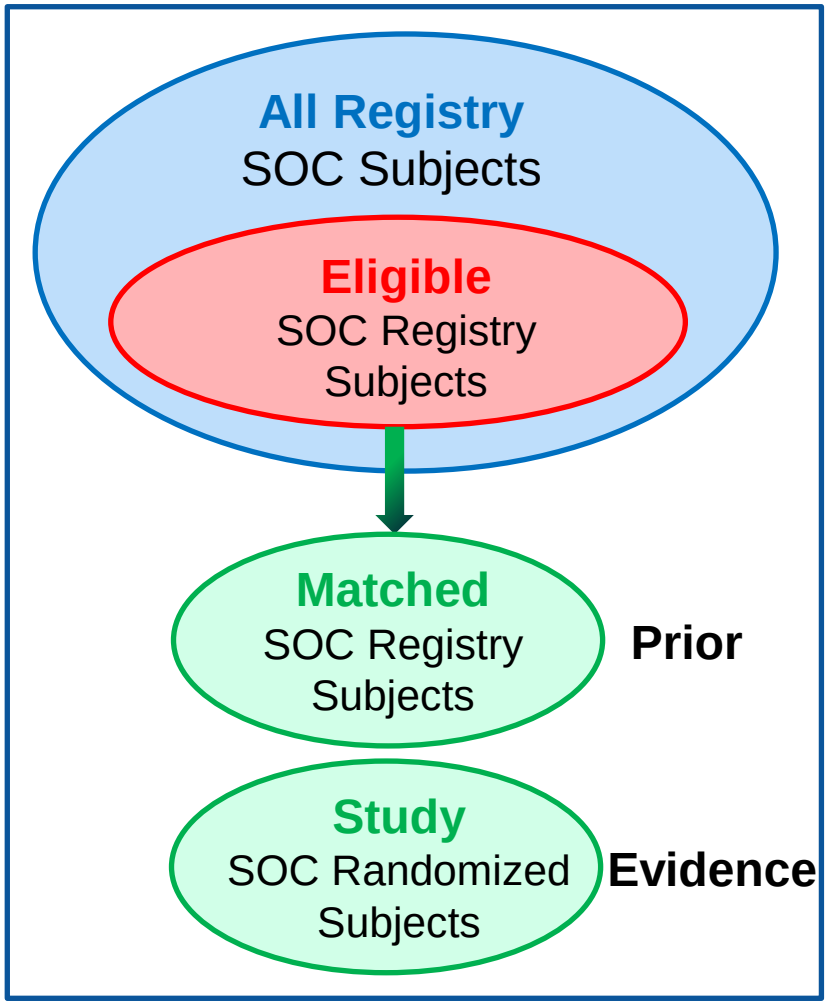
"amounts to a search for a data set that might have resulted from a randomized experiment hidden within an observational data set."

[King 2015] King G and Nielson G. Why propensity scores should not be used for matching. Harvard University and Massachusetts Institute of Technology, 2015.

SELECTION OF MATCHED COHORT FROM PEDIATRIC REGISTRY

Steps for matching (based on individual subjects data):

1. Restrict registry subjects to mimic the eligibility criteria of the study
2. Define/obtain statistically and clinically relevant prognostic variables for the M/M endpoint (Cox)
3. Consider variables correlated with important unmeasured confounders
4. Order variables in terms of prognostic importance
5. Predefine rules for statistical matching



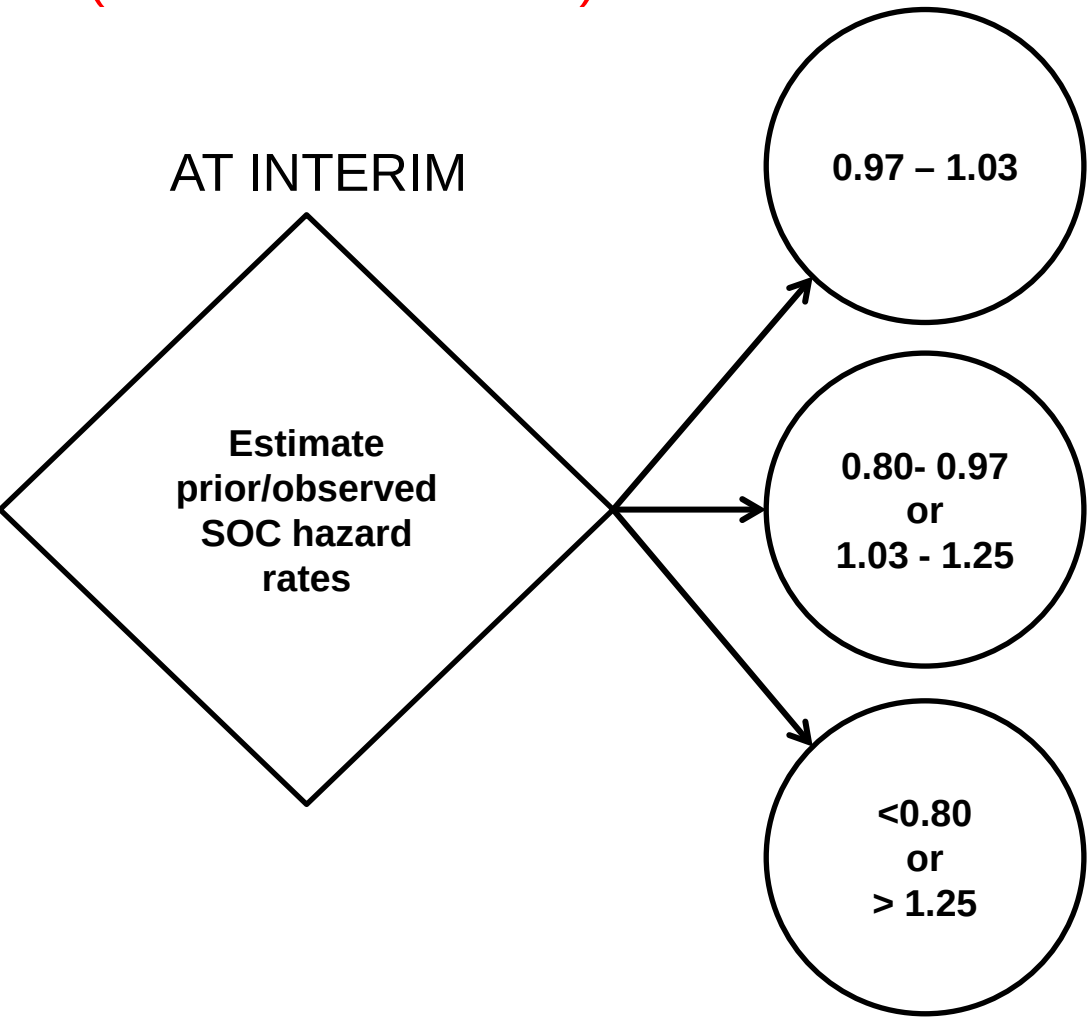
SOLUTION: INFORMATIVE BAYESIAN ADAPTIVE

To incorporate both the matching and the control of errors, Actelion proposed a refined, informative Bayesian adaptive design:

- At interim (performed by independent statistician):
  - Identify the matched cohort from the registry
  - Estimate hazard rates and heterogeneity between matched cohort and study SoC data
  - Depending on magnitude of heterogeneity determine the final sample size of the study using a pre-defined, adaptive algorithm

NEW BAYESIAN ADAPTIVE DESIGN OPTION: EXAMPLE

Depending on magnitude of heterogeneity, the following could happen (based on simulation):



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

- Stop study at 137-179 events
- Primary analysis: Informative Bayesian

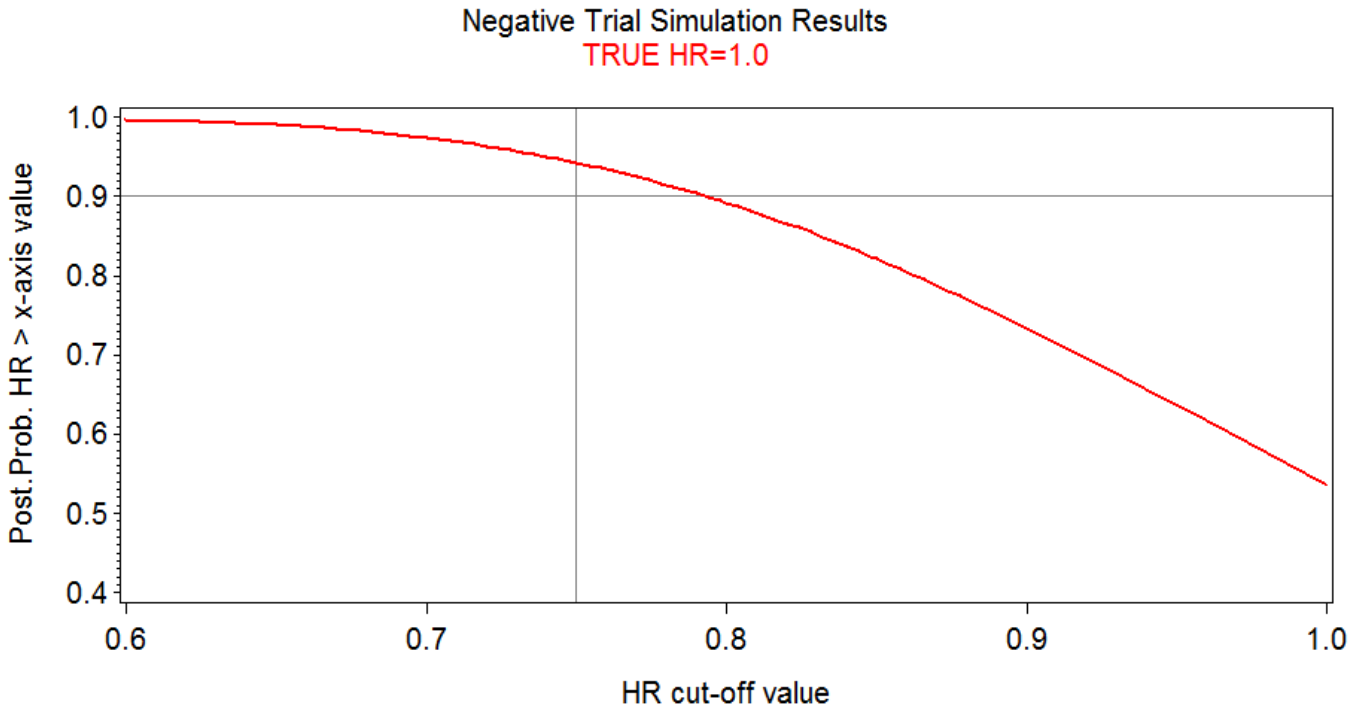
- Stop study at 180 events
- Primary analysis: Un-informative Bayesian

CONCLUSIVENESS OF THE BAYESIAN ADAPTIVE APPROACH

- Sample size is already computed with adequate power for expected treatment effect
- However, in case of non-significant treatment difference, it is important to obtain 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., “Macitentan highly likely to have negligible effect in children”

SUMMARY OF PROPOSED DESIGNS (0.025, 1-SIDED)

Standard Frequentist superiority	Bayesian Informative	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

HR = 0.65

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 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)
3. Complexity of implementation, and no precedent in pediatric PAH

Actelion decided for a standard, frequentist, group sequential design

WHAT DID WE LEARN?

1. The community (HAs and sponsors) is 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 and reports are required for the HAs interactions and in the protocol to document code, methods and performance characteristics (especially type 1 error control and conclusiveness)

4. Case specific limitation: only one source of prior data for control was available.

WHAT DO WE RECOMMEND FOR THE FUTURE?

- Bayesian and adaptive designs should continue to be considered:
  - If historical data for prior are expected to be similar (for example if Pocock criteria are met)
  - Accounting for interim heterogeneity checks (with/without subject matching)
  - Plan ahead for extensive simulations based on PK/PD and clinical information

- Need for clear and global guidance (and metrics) from HAs on what is recommended for a successful application of Bayesian designs beyond devices

BACK-UP

DOSE SELECTION STRATEGY DURING CLINICAL STUDY

COMPARABLE EXPOSURE ASSUMED TO BE ACHIEVED BY DOSE SCALING BASED ON BODY-WEIGHT AND AGE (I.E., ENZYME MATURATION)

- Exposure prediction using population PK model scaled to children
  - Clearance scaled by allometry and ontogeny of CYP3A4
  - Volume scaled by allometry
- Doses for children to achieve adult therapeutic exposures

	Weight-based dose groups for 2-17 year old children			
WT cut off (kg)	<15	15 to <25	25 to <50	≥50
Dose (mg)	3.5	5	7.5	10

- Dose adaptation during study conduct
  - Enrolment of children from ≥2 to <18 years
  - Conduct of an interim PK analysis
  - Update of PK model, definition of doses for children from ≥1 month to <2 years
  - Enrolment of children from ≥1 month to <2 years