Assessment of Consistency and Bayesian Approaches for Demonstrating Efficacy in Pediatric Populations

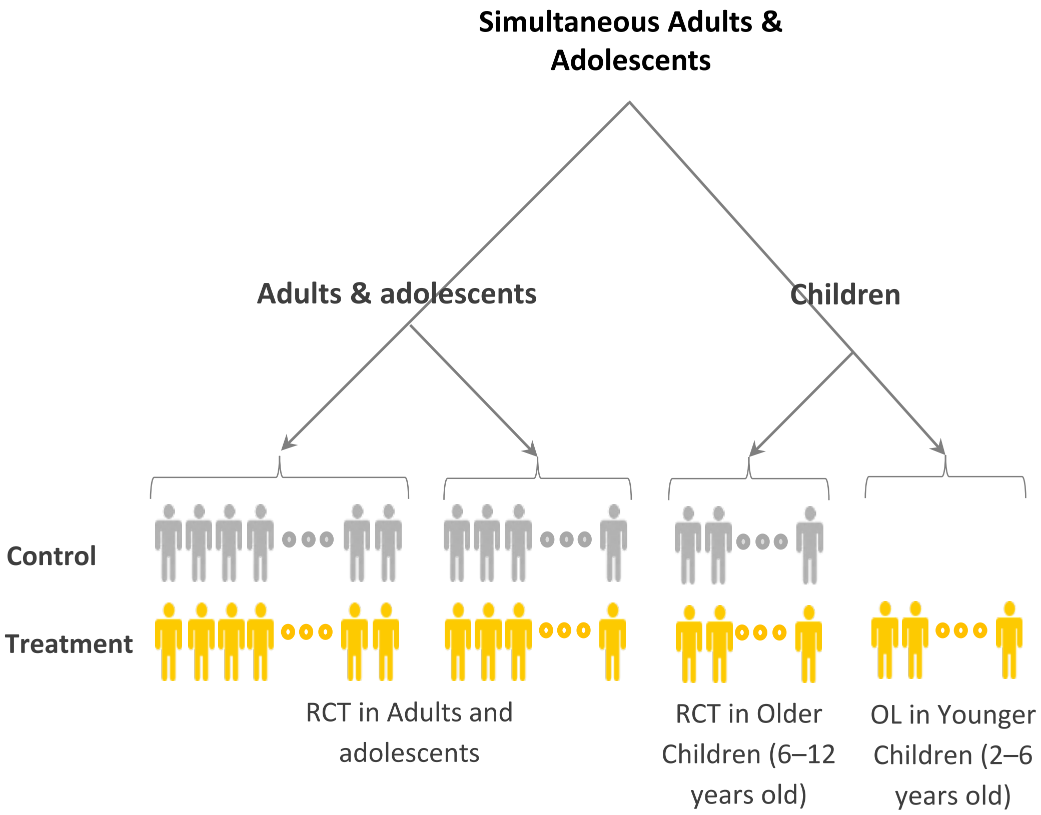
Margaret Gamalo, PhD

Senior Director – Biostatistics, Medical Dermatology

Global Biometrics and Data Science, Pfizer

Advancing the Development of Pediatric Therapeutics (ADEPT 7)
Complex Innovative Design, Public Workshop
September 1 - September 2, 2021

Two Hypothetical Development Plans



Simultaneous Adults & Adolescents

Adults & adolescents

Children

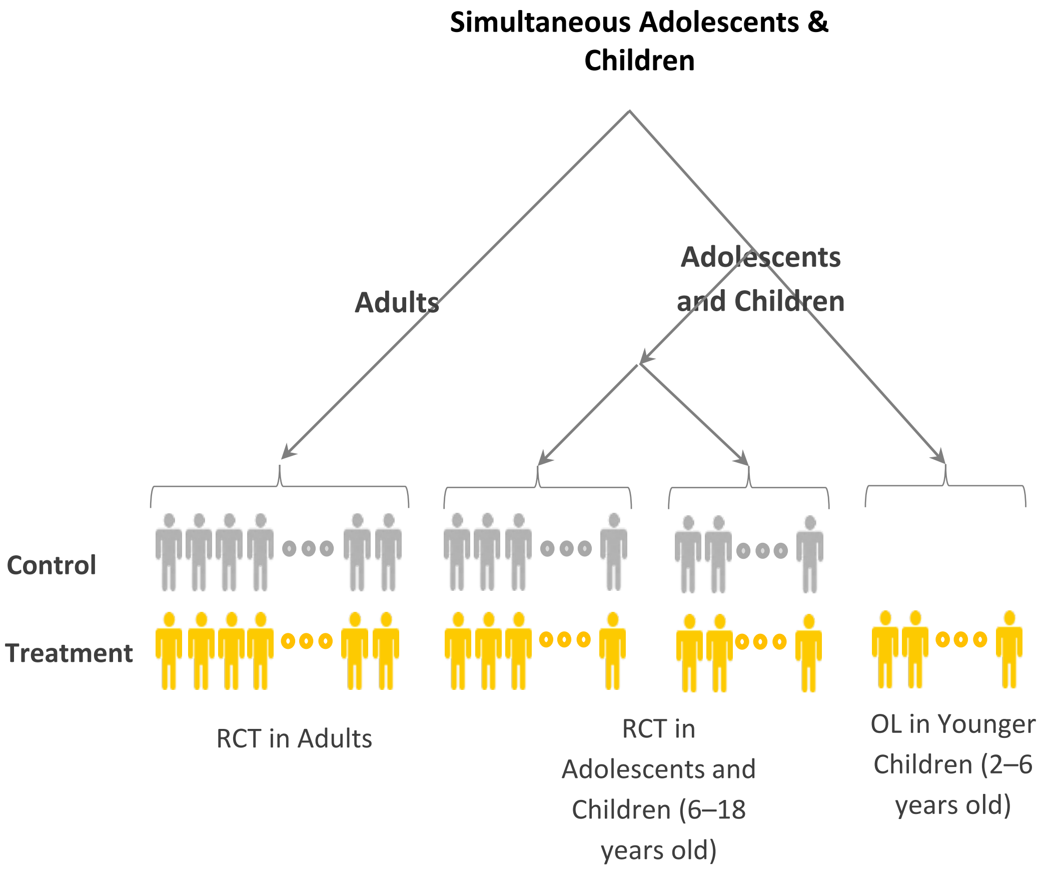
Control

Treatment

RCT in Adults and adolescents

RCT in Older Children (6–12 years old)

OL in Younger Children (2–6 years old)



Simultaneous Adolescents & Children

Adults

Adolescents and Children

Control

Treatment

RCT in Adults

RCT in Adolescents and Children (6–18 years old)

OL in Younger Children (2–6 years old)

What does efficacy in each of the cohorts look like?
How can that be used to design trials?

Extent of Hypothetical Development Plan

Assumption: Overall Effect size is 15%, Randomization is 1:1 (iTx: Placebo)

	Development A (Simultaneous adult and adolescents)	Development B (staggered, combined adolescents and children)	Development C (fully staggered)
Pivotal Trials in adults:	Two 400 patient pivotal RCT including 15% adolescent (60)	Two 400 patient pivotal RCT in adults only	Two 400 patient pivotal RCT in adults only
Trials in adolescents	Included in adults: 120 with 60 exposed to iTx	Stand alone adequately powered RCT including children	**Stand alone adequately powered RCT?**
Trials in children 6-12 yr olds	**Should this be adequately powered stand-alone trial?**	Combined with adolescents	**Stand alone adequately powered RCT?**
Trials in children 2-6 yr old	Open label 50 pts	Open label 50 pts	Open label 50 pts

Should extent of development in children be similar under different clinical development strategies? Or should a strategy that makes access to drugs for children be given incentive?

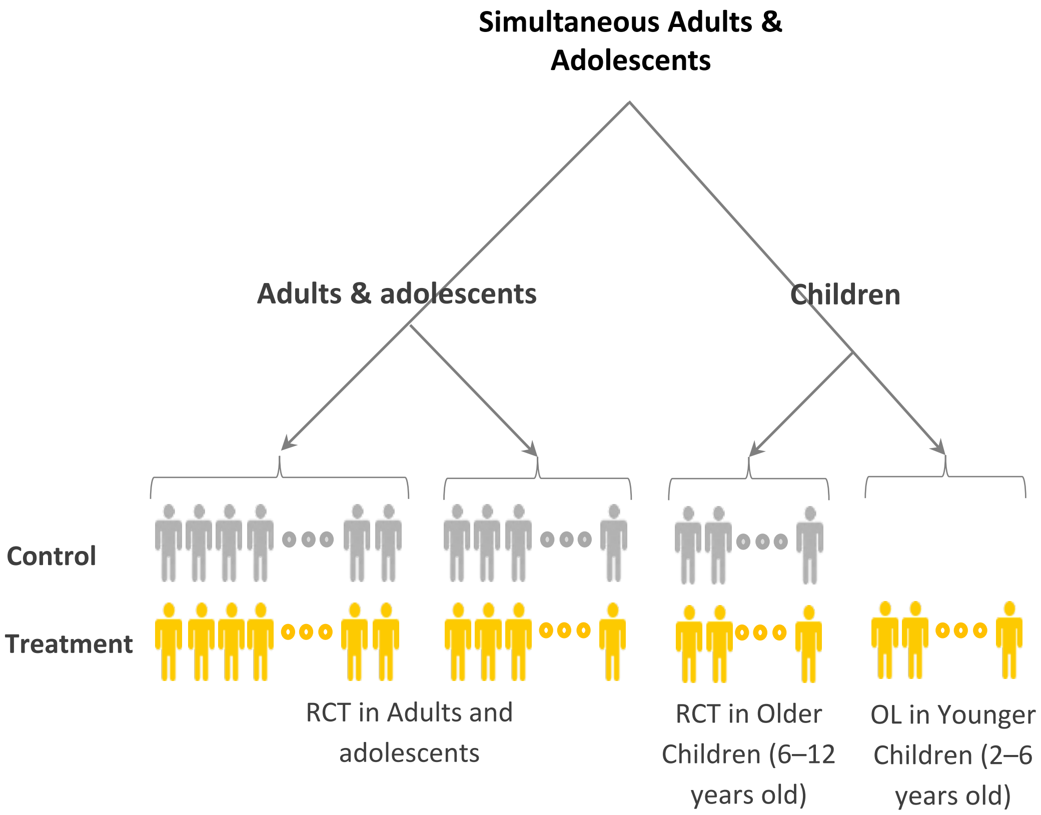
Extent of Development Plan

Assumption: Overall Effect size is 15%, Randomization is 1:1 (iTx: Placebo)

	Development A (Simultaneous adult and adolescents)	**Development B (staggered, combined adolescents and children)**	**Development C (fully staggered)**
Pivotal Trials in adults:	Two 400 patient pivotal RCT including 15% adolescent (60)	Two 400 patient pivotal RCT in adults only	Two 400 patient pivotal RCT in adults only
Trials in adolescents	Included in adults: 120 with 60 exposed to iTx	Stand alone adequately powered RCT including children	**Stand alone adequately powered RCT - 194**
Trials in children 6-12 yr olds	**194 (no borrowing)– not logical!**	Combined with adolescents	**Stand alone adequately powered RCT - 194**
Trials in children 2-6 yr old	Open label 50 pts	Open label 50 pts	Open label 50 pts
Total pediatric patients in development	**364**	**244**	**438**

Is Development B truly the optimal? On paper, yes. But probably not in reality! How can Development A be optimized?

Hypothetical Development Plan A


Simultaneous Adults & Adolescents
Adults & adolescents
Children
Control
Treatment
RCT in Adults and adolescents
RCT in Older Children (6–12 years old)
OL in Younger Children (2–6 years old)

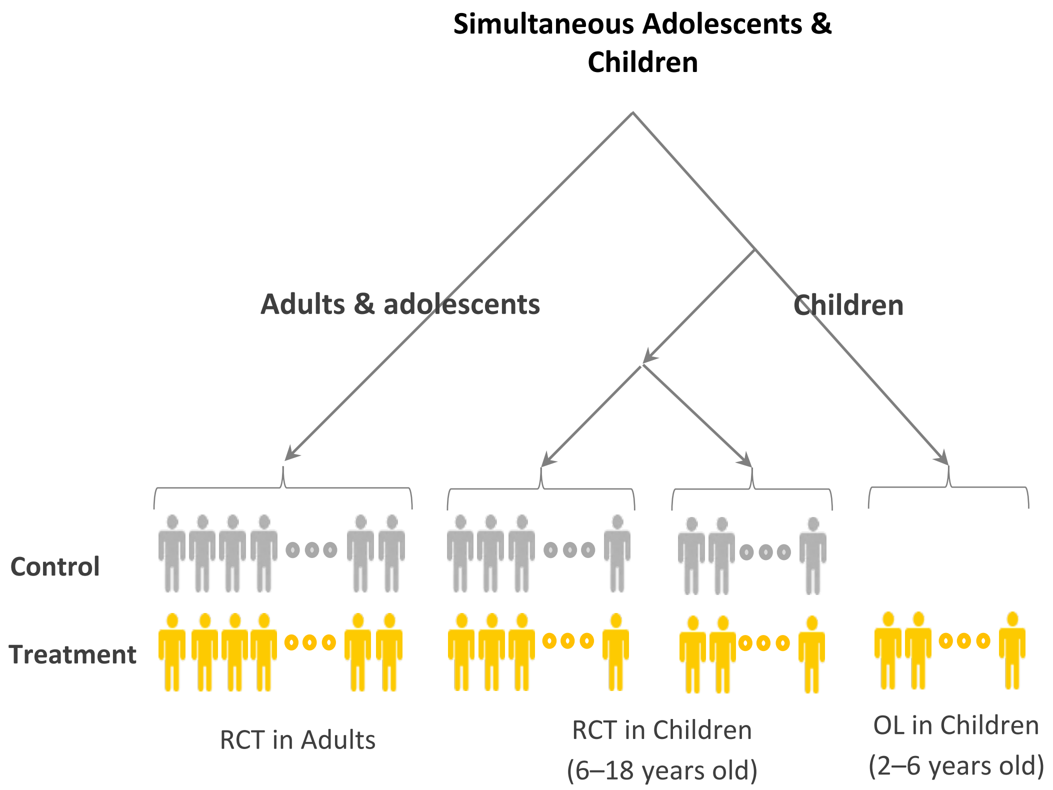
Assumption for adult and adolescent RCT: Treatment benefit in adults and adolescents are similar

- Trial wants to determine whether there is heterogeneity
- In the frequentist sense, show presence of interaction. However, interaction not significant is not proof of no heterogeneity.
- Show **consistency** of **all** cohorts.

Should we have the same assumption for 6-12 yr olds? i.e., treatment effect is consistent across all cohorts?

- If not, show **efficacy** in 6-12 yr old?
- Logical restriction – less patients exposed to research risk in younger cohorts than adjacent older cohort, i.e.,
    - If enrollment in adolescents is only 120, shouldn't the 6-12 yr old trial be less than or equal to 120?
- Partial extrapolation seems imperative, i.e., borrow data from older cohorts to show difference.
    - Is there preference to borrowing adjacent cohort than non-adjacent cohort?

Hypothetical Development Plan B


Simultaneous Adolescents & Children
Adults & adolescents
Children
Control
Treatment
RCT in Adults
RCT in Children (6–18 years old)
OL in Children (2–6 years old)

Assumption: Assumes that the treatment benefit in adults and children may or may not be the same or there are safety concerns.

Objective: Show efficacy

- o in the overall pediatric population, or
- o for each cohort, i.e., show efficacy in adolescents and show efficacy in children?

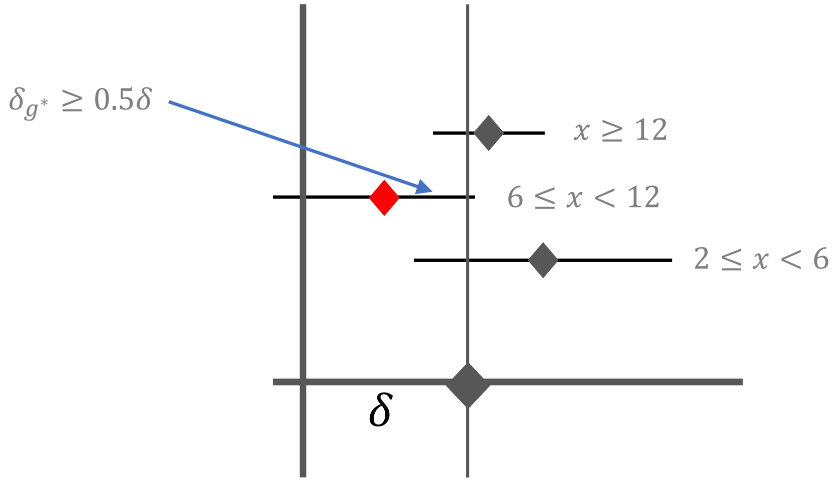
- If development is staggered, i.e., adolescents first before children, show efficacy for each?
- If simultaneous development, show consistency?

It appears that the criterion for efficacy is dependent on whether the development is simultaneous or staggered, i.e., **if simultaneous the criterion is consistency whereas for staggered it is efficacy.**

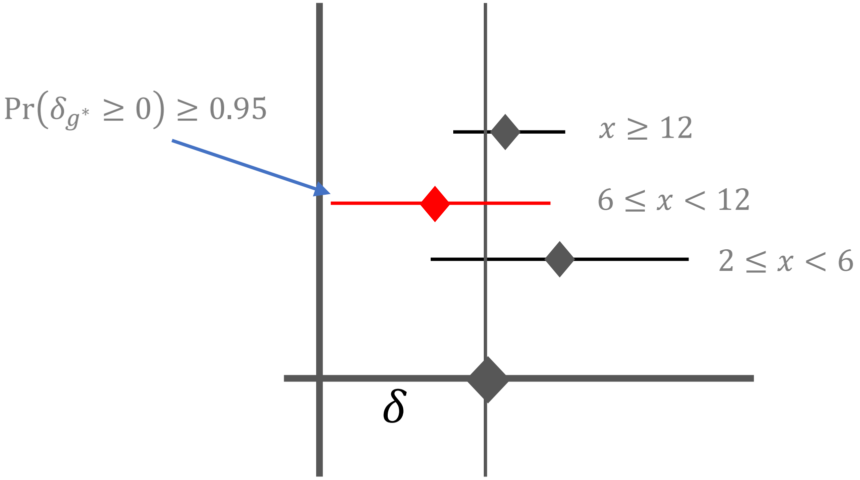
- Decision to pool cohorts is a decision of similarity of diseases!
- Creates a hurdle for diseases that are dissimilar (perceived or known) or with concerns of potential safety risks.

Efficacy through Assessment of Consistency

- **Assumption**: All cohorts are similar. Typically, want to confirm that the response is robust and consistent across cohorts; not to show difference in cohorts.
- **Hypothesis**: $H_0: \delta_{g^*}$ not consistent vs $H_1: \delta_{g^*}$ consistent (reverse!)
- **Criterion for Efficacy**: Consistency assessment based on retaining a portion of the overall effect, i.e., $\Pr\left(\delta_{g^*} \geq c\delta\right) \geq p^*$



Consistency based on the Point estimate



Consistency based on point estimate and level of uncertainty

Optimizing Developments A-C

- If it is shown that adults and adolescents have consistent treatment effects (Development A), **“line of reasoning” for extrapolation is established**!
  - There is sufficient rationale that 6-12 yr olds should have similar effect as well.
- In Development C, “line of reasoning” has not been established for adolescents.

			Consistency	
Development	Cohort	No borrowing	PMDA method 1*	PMDA method 2**
A	Adolescents	120 - included in adult RCT	~110	~100
	6-12 yr olds	194	~90 260	~80 250
C	Adolescents	194	~90	~80
	6-12 yr olds	194	~70 210	~80 210

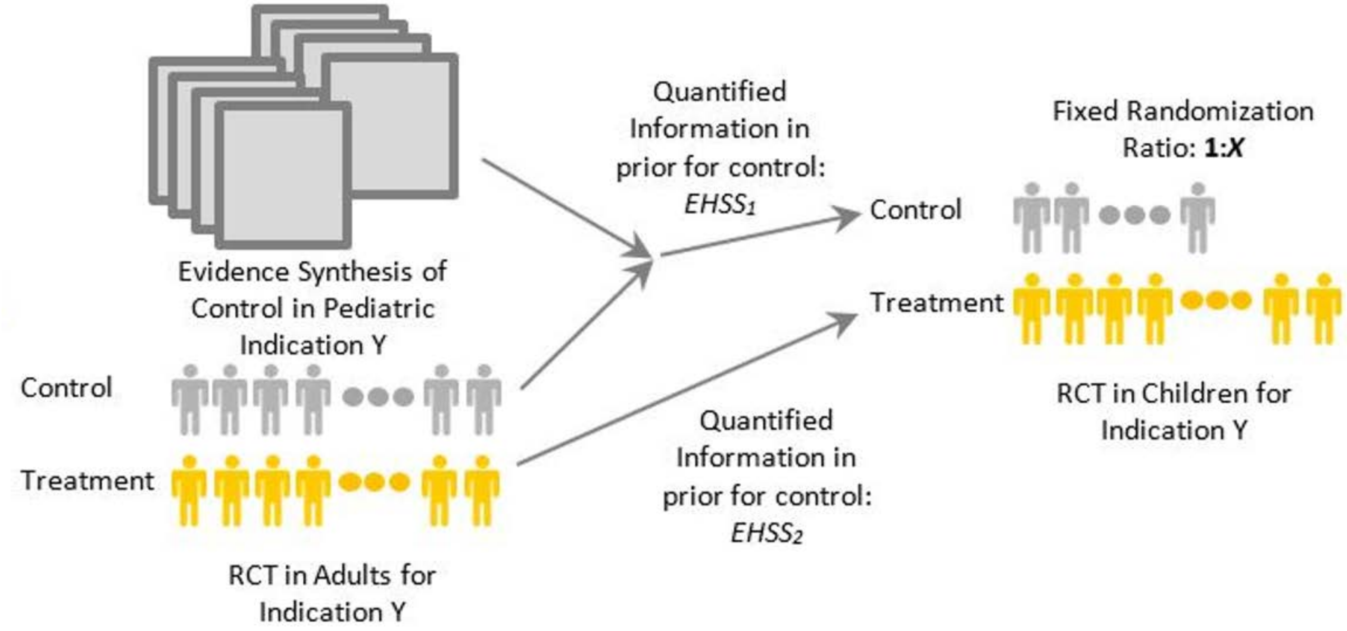
*PMDA method 1: $\Pr(\delta_{g^*} \geq 0) \geq 0.95$;

**PMDA method 2: $E[I(\delta_{g^*} \geq t)] \approx \Pr(\delta_{g^*} \geq t)$; MCMB = 5%, $t = \max(0.05, 0.5\delta)$

Efficacy through showing non-zero effect

Assumption: Cohorts may or may not be the similar. The prior can be chosen to incorporate this uncertainty.

Hypothesis: $\boldsymbol{H_0}$: treatment effect $= 0$ vs $\boldsymbol{H_1}$: treatment effect > 0



Evidence Synthesis of Control in Pediatric Indication Y

Control

Treatment

RCT in Adults for Indication Y

Quantified Information in prior for control: *EHSS₁*

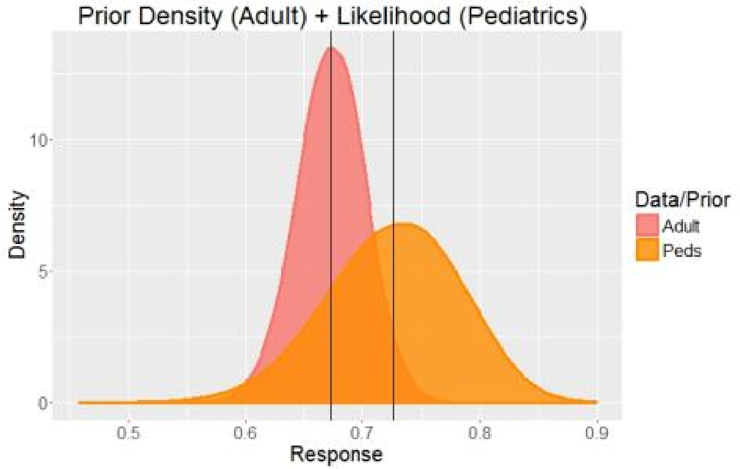
Quantified Information in prior for control: *EHSS₂*

Control

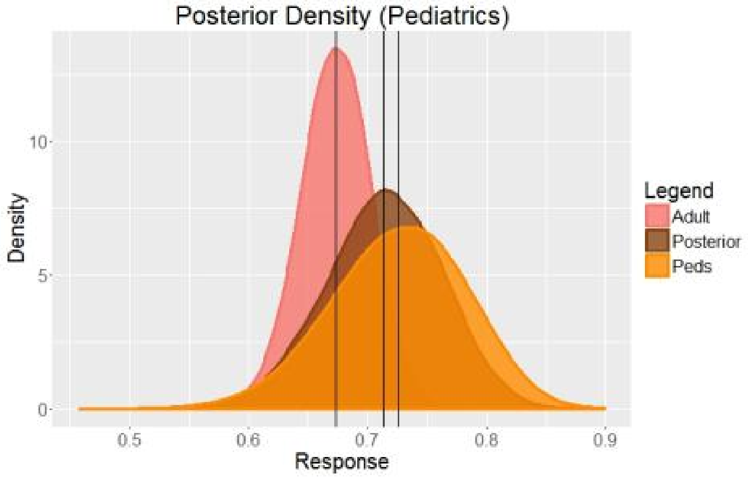
Treatment

Fixed Randomization Ratio: **1:*X***

RCT in Children for Indication Y



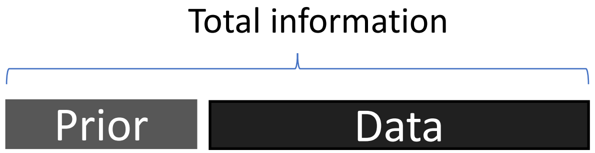
Prior Density (Adult) + Likelihood (Pediatrics)



Posterior Density (Pediatrics)

Efficacy through showing non-zero effect

Two schools of thought on borrowing:

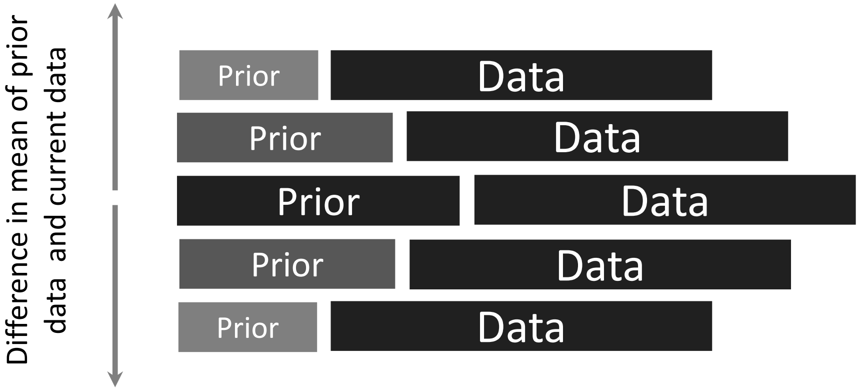


Total information

Prior

Data

Pre-specify amount of borrowing depending on similarity of disease; regardless of outcome



Difference in mean of prior data and current data

Prior

Data

Prior

Data

Prior

Data

Prior

Data

Prior

Data

Amount of data borrowed is dependent on similarity of outcomes; amount of borrowing is dependent on specific outcome

Extrapolation assumption: Is the assumption of *exchangeability* justified?

- exchangeability is with respect to similarity of populations; not just outcomes!)
- what flexibility can be accommodated for trials where monotherapy cannot be administered in children?

Variability and robustness: To what extent does the prior need to be robust?

- Posterior mean of treatment is not shifted by borrowing too much!
- Is there a way to measure the effective sample size or its upper bound accurately?

Outcome Choice and Extrapolation

Continuous endpoints over time provide better picture of progression of disease and consistency of treatment over time

Endpoints derived from dichomization are inefficient and are highly influenced by sample size

Leveraging of information through Bayesian methods that adjust level of borrowing based on differences in response, are extrapolating based on similarity of disease and variability

Proposal: Use continuous efficacy measures for primary endpoint and support by dichotomized endpoint that leverages available information or showing consistency

Characteristics of an <u>ideal</u> extrapolation index (score)

- Similarity of disease and treatment response implies a fixed proportion of borrowing, i.e., $w \in [0,1]$.
- Priors should have the ability to pivot if data are not the same, or data is worse.
- Proportion of borrowing generally within a range, $\mathrm{a}_0 \leq \pi(w) \leq b_0$
  - b_0 is pre-specified (elicited) maximum borrowing; and
  - a_0 is ***validation*** or how low borrowing can be in case of prior-data conflict

Optimizing Developments A-C

- If it is shown that adults and adolescents have consistent treatment effects (Development A), **“line of reasoning” for extrapolation is established**!
  - There is sufficient rationale that 6-12 yr olds should have similar effect as well.
- In Development C, “line of reasoning” has not been established for adolescents.

			Efficacy-Bayesian Approach		
Development	Cohort	No borrowing	Robust Prior $(\nu = 0.5)^{\S}$	Power Prior $(w = 0.5)$	Commensurate Prior $(\nu = 0.5)^{\S}$
A	adolescents	120 - included in adult RCT	NA	NA	NA
	6-12 yrs old	194	80 250	<50 220	<50 220
C[§§]	adolescents	194	80	<50	<50
	6-12 yrs old	194	60 190	<50 150	<50 150

*PMDA method 1: $\Pr(\delta_{g^*} \geq 0) \geq 0.95$;

**PMDA method 2: $E[I(\delta_{g^*} \geq t)] \approx \Pr(\delta_{g^*} \geq t)$; MCMB = 5%, $t = \max(0.05, 0.5\delta)$

[§] Tuning parameter rather than proportional to amount of information borrowed

[§§] In Development C, the Adult Trial is composed of 2 RCTs

Simulated False-positive decision (concluding efficacy for an ineffective drug in target cohort)

- Typically, decision criteria for concluding efficacy under the Bayesian approach mimics Frequentist rules, e.g., false positive conclusion to be less than 0.05.
- Borrowing information through informative priors will always increase Type-I error
- Strict control of Type-I error limits the ability to extrapolate, particularly, in diseases that are similar in reference and target cohorts
- **Proposal: The more similar the diseases, the tolerance for type I error needs to be higher!**

Label Language in Section 14

- Language on efficacy drawn from innovative analytics is challenging to effectively communicate.
- Estimation of mean (confidence interval) without borrowing can give false impression that treatment is ineffective (if the interval is crossing 0 or 1).
- Use of priors is related to methodology which creates additional challenge in communication.
- If Bayesian credible intervals are not used, use of side-by-side data comparing target and reference populations can be helpful to provide contextual efficacy.

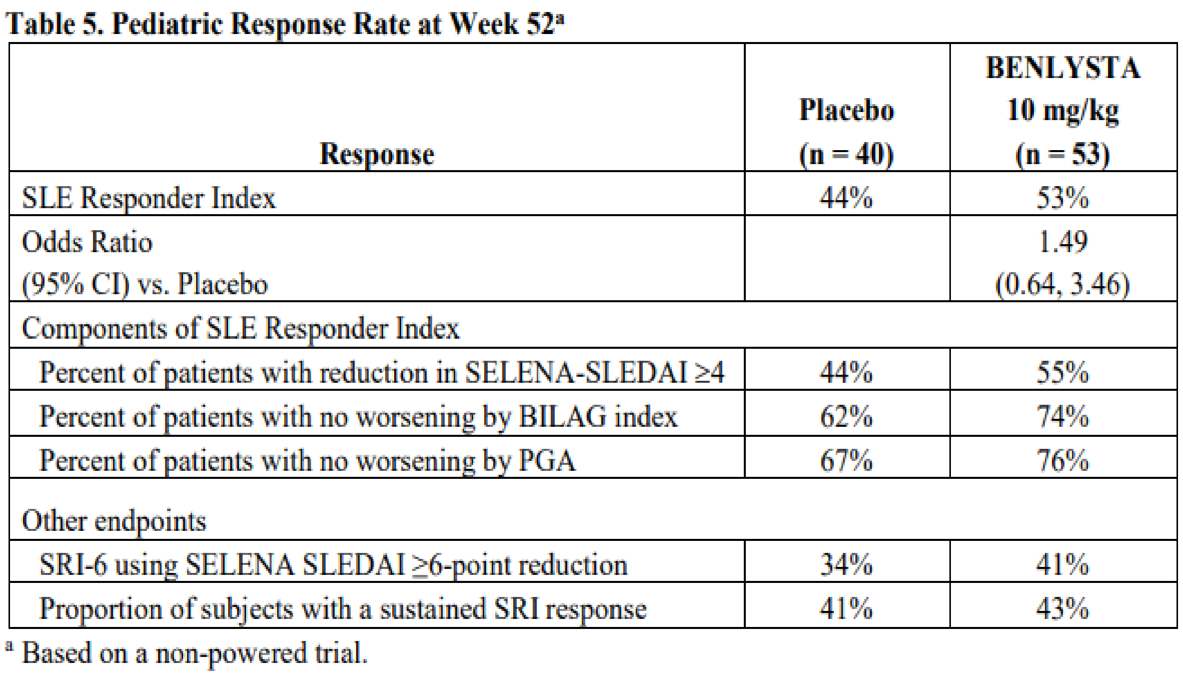


Table 5. Pediatric Response Rate at Week 52[a]

Response	Placebo (n = 40)	BENLYSTA 10 mg/kg (n = 53)
SLE Responder Index	44%	53%
Odds Ratio (95% CI) vs. Placebo		1.49 (0.64, 3.46)
Components of SLE Responder Index		
Percent of patients with reduction in SELENA-SLEDAI ≥4	44%	55%
Percent of patients with no worsening by BILAG index	62%	74%
Percent of patients with no worsening by PGA	67%	76%
Other endpoints		
SRI-6 using SELENA SLEDAI ≥6-point reduction	34%	41%
Proportion of subjects with a sustained SRI response	41%	43%

[a] Based on a non-powered trial.

Table 5 in Belimumab USPI describing efficacy in multiple endpoints

Safety and extrapolation

- As trials become leaner through efficient analytical ways, there will be less and less data for the assessment of safety.
- Large safety exposures to ascertain signal and precision of treatment emergent adverse events and long-term effects of the drug can make use of innovative designs moot.
    - If a drug has no on-target effects on safety domains of interest, what would be the number of patients needed and the length of the follow-up to have an adequate safety database.
    - Can there be room for extrapolation of safety?

Conclusion

- Current framework for establishing efficacy in pediatrics creates a possibility of illogical extent of development since requiring adequate and well controlled trials in cohorts will require younger cohorts having bigger sample size than older cohorts, exposing more children to research risks.
- Efficacy criteria based on consistency can be conservative if not chosen appropriately. If consistency is the measure for efficacy in cohorts, it needs to be based it on ensuring that the least benefitting subgroup retains positive benefit risk.
- If the criteria is demonstrating efficacy, leveraging of information through Bayesian approach that has a prior that is both not dependent on one outcome but also flexible is ideal.
- The type-I error is always increased under extrapolation. This has to be countered by an increase in tolerable uncertainty guided by similarity of disease.
- Adoption of innovative analytics create challenge in communication of efficacy in label language that needs to be addressed.
- There needs to be an avenue for extrapolation of safety. How it can be implemented requires further discussion.