# Case study using Bayesian methods to leverage existing clinical efficacy data in paediatric trials

Nicky Best and Dawn Edwards, GSK

PSI Extrapolation Meeting

22 Nov 2017

# Acknowledgements


- Wen Wu
- Lynn Dunsire
- Shuying Yang
- Clara Dominguez

# Outline

- Background
- Case study: Partial extrapolation from adults to adolescents
  - Summary of extrapolation strategy
  - Dynamic Bayesian methods for historical borrowing/extrapolation
  - Design and operating characteristics
- Concluding remarks

# Paediatric Drug Development

- Use of unlicensed / off-label medicines in children is widespread
- Majority of drugs prescribed for children have not been properly tested in paediatric populations
- Now a regulatory requirement for sponsors to agree a paediatric development plan for all new drugs under development (unless a waiver is granted)
- Paediatric populations present several challenges and opportunities for clinical trial design and analysis
  - Practical and ethical constraints on sample size ⇒ fully powered efficacy studies are often not feasible or justified
  - In contrast to other situations, available data have been sufficient for licensing a new drug in adults, and PK-PD and mechanism of action are usually well understood ⇒ strong rationale to extrapolate efficacy and safety from adults / other sources to children

# Paediatric Drug Development

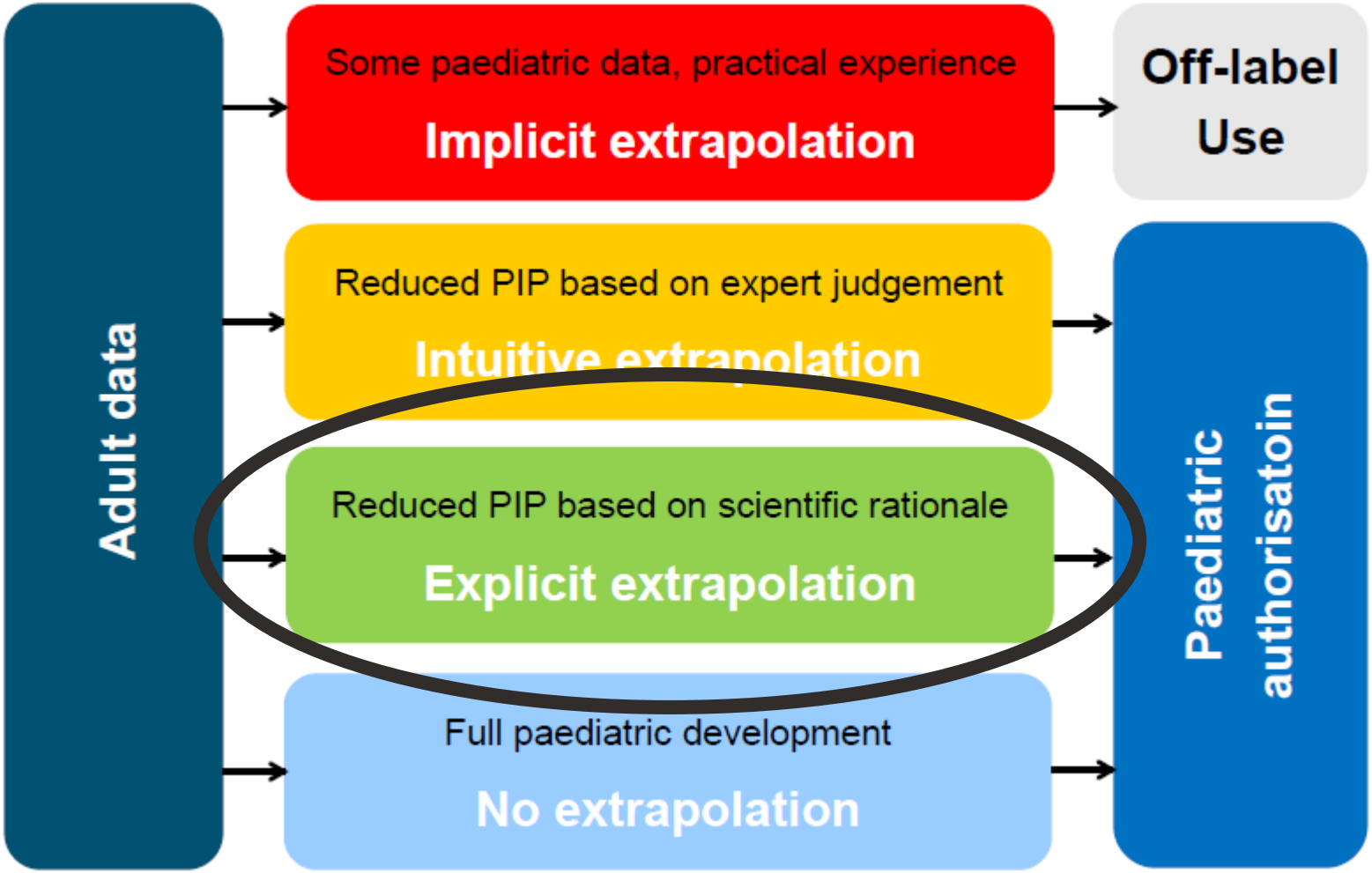



Figure from EMA Extrapolation Workshop May 2016 presentation by Christoph Male (PDCO)

# Reducing data requirements clinical trials in children

EMA 2016 Reflection paper:

- What design features can be altered:
  - Using higher type 1 error threshold (? requires quantitative justification)
  - Widening usually accepted NI margin (clinical interpretation may be different)
  - Using Bayesian methods to explicitly borrow information (from adult trials, control groups, other paediatric trials)

# Case Study:
# Partial extrapolation from adults to adolescents in a phase III respiratory trial

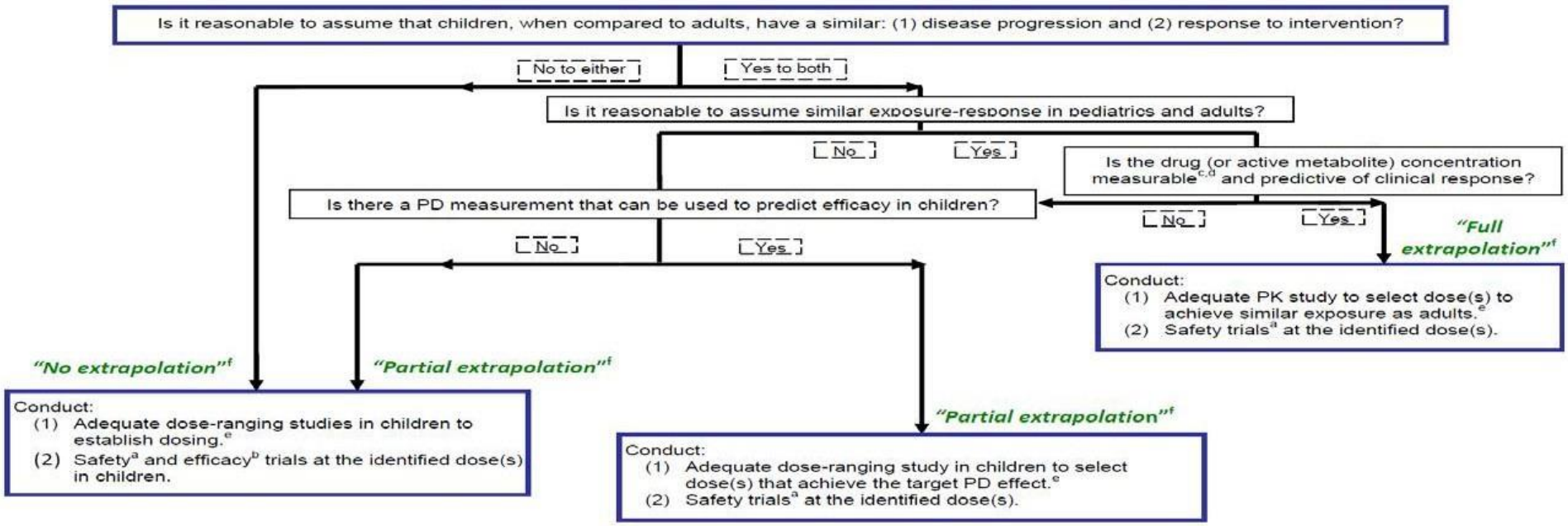
# Case study: Background

- Study background:
  - Phase III adult respiratory study (18+ years) – Ongoing
  - Phase III adolescent respiratory study (12-17years) - Planned
- Study Design
  - Parallel group – both studies
- Primary endpoint
  - Lung function improvement as measured by change from baseline in trough FEV1 for both studies
- Study population
  - Similar population for both studies (uncontrolled asthma)
- Treatment arms
  - Treatment A+B (experimental) vs Treatment A (control) in both studies

# Case study: Rationale for Partial Extrapolation Strategy

- Partial extrapolation of adult efficacy to adolescents is proposed, using Bayesian techniques
- Rationale for partial extrapolation
  - Adolescents & adults previously studied together in similar clinical settings
  - Prevalence lower in children than adults $\Rightarrow$ recruitment challenges
  - Based on standard care in this disease it's acceptable for a study in adolescents to have similar objectives and endpoints to the adult study, recruiting patients with the same degree of disease severity
  - It is reasonable to assume that adolescents, when compared to adults, have a similar disease progression and response to intervention
  - Key differences between adult and adolescents may have led to negative outcomes in the past and they have been carefully considered in the design of the adolescent study

# FDA Pediatric Study Planning & Extrapolation Algorithm




Footnotes:

a. For locally active drugs, includes plasma PK at the identified dose(s) as part of safety assessment.
b. For partial extrapolation, one efficacy trial may be sufficient.
c. For drugs that are systemically active, the relevant measure is systemic concentration.
d. For drugs that are locally active (e.g., intra-luminal or mucosal site of action), the relevant measure is systemic concentration only if it can be reasonably assumed that systemic concentrations are a reflection of the concentrations at the relevant biospace (e.g., skin, intestinal mucosa, nasal passages, lung).
e. When appropriate, use of modeling and simulation for dose selection (supplemented by pediatric clinical data when necessary) and/or trial simulation is recommended.
f. For a discussion of no, partial and full extrapolation, see Dunne J, Rodriguez WJ, Murphy MD, et al. "Extrapolation of adult data and other data in pediatric drug-development programs." Pediatrics. 2011 Nov;128(5):e1242-9.

# Collating Historical Adult and Adolescent Data

- Searched for internal and external studies and publications in a similar patient population with similar treatment arms (to the planned experimental and control treatments) and at similar timepoints for adult and adolescent subjects for the primary endpoint trough FEV1
- Built a excel spreadsheet containing this historical adult and adolescent data
  - Collected information such as study design, population, treatment, time, estimated treatment difference, SE etc
- Reviewed the spreadsheet to assess which studies were the most relevant for constructing the prior
- Where studies had combined adult and adolescent data we performed subgroup analyses for the adolescent population (if possible) to pull out treatment differences and confidence intervals
- Generally the sample sizes were quite high for the adult data (100-900 per arm), but much lower for the adolescent data (from a handful to 140 per arm)

# Collating Historical Adult and Adolescent Data

- Discussed the spreadsheet and our findings with clinical team
  - To understand whether they would expect differences in adult and adolescents and if so what is the medical rationale for these differences
  - To explain the Bayesian methodology and ensure they understand and buy into it
  - To ensure we've captured appropriate historical data
- Explored the relationship between adult and adolescent treatment difference and SD for trough FEV1
  - Methods described on later slides
- Explored Bayesian methodology for incorporating historical data (see later slides)

# Ongoing activities

- Refining how to build the priors to be incorporated into the adolescent study
- Simulation study to investigate:
  - Prior weights
  - PoS and precision
  - Type I and Type II error
- Developing similar dynamic Bayesian design to extrapolate from adolescents to paediatric population

- Next slides will explain the development of the priors and methods for incorporating them in the adolescent study in more detail

# Dynamic Bayesian methods for historical borrowing

# How does it work?

*Bayes' Theorem (combine historical belief with new data) is our natural tool*

– Historical data used to generate predictive prior distribution for mean response in new trial

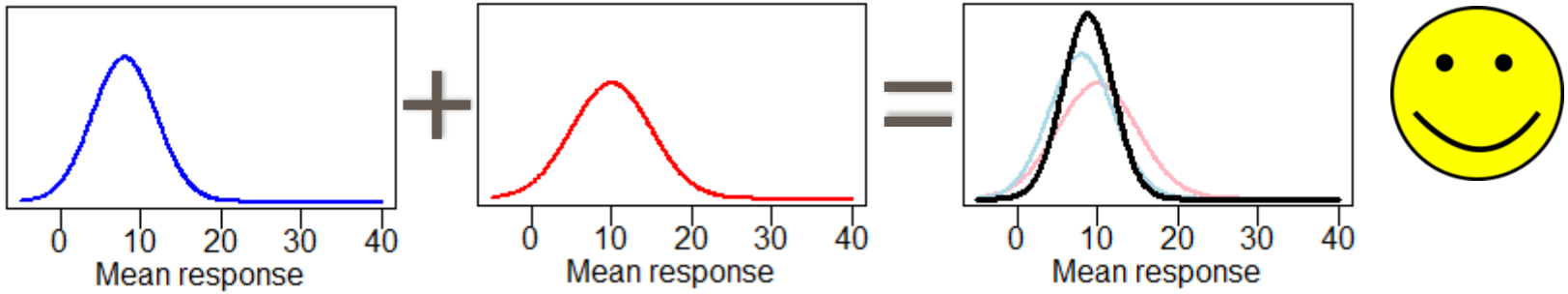
Predictive distribution for what we believe about future responses based on the historical studies ("prior")

What we see in the new study ("sampling distribution")

"Posterior" distribution: weighted average of prior and new trial data

**Scenario 1**

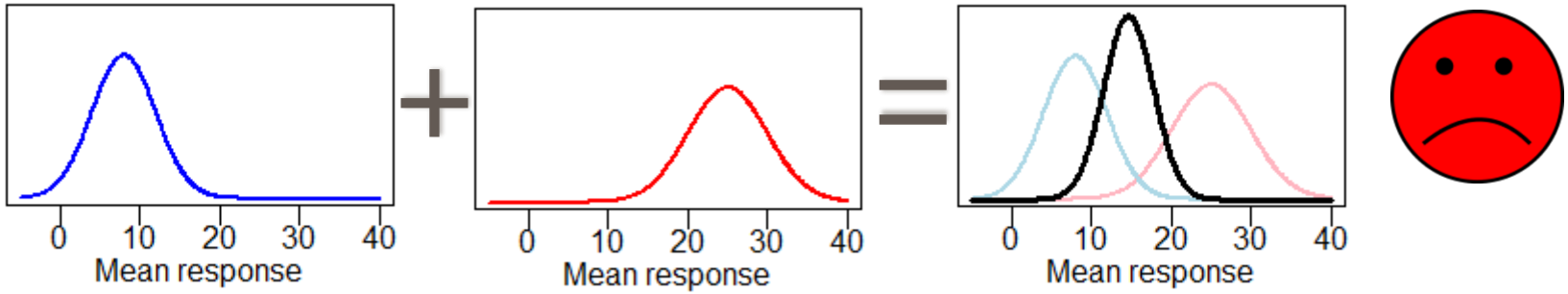
Historical and new data are consistent




– But, can result in potentially unrealistic estimates if historical data conflicts with new data

**Scenario 2**

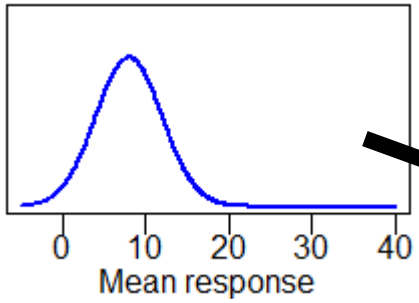
Historical and new data in conflict



# “I beseech you, in the bowels of Christ, think it possible that you may be mistaken” (Cromwell’s rule)

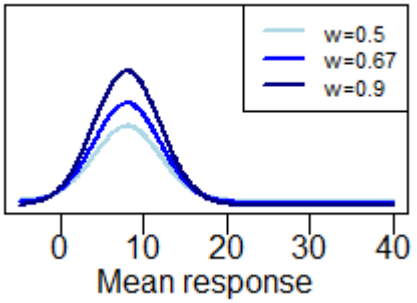
– Robust (dynamic) Bayesian prior can be used to address the problem of conflict between historical and new data

Prior assuming historical data are relevant

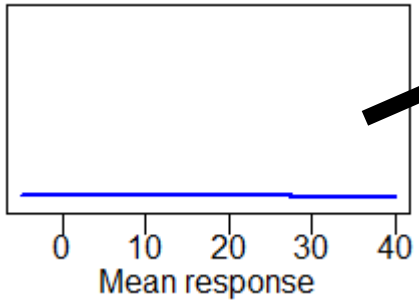



w

Robust prior = weighted mixture of these 2 priors




“In case we are mistaken” prior i.e. assuming historical data are not relevant




1-w

*Schmidli et al. Biometrics (2014)*

# Robust (dynamic) Bayesian models to deal with the historical data / new data conflict

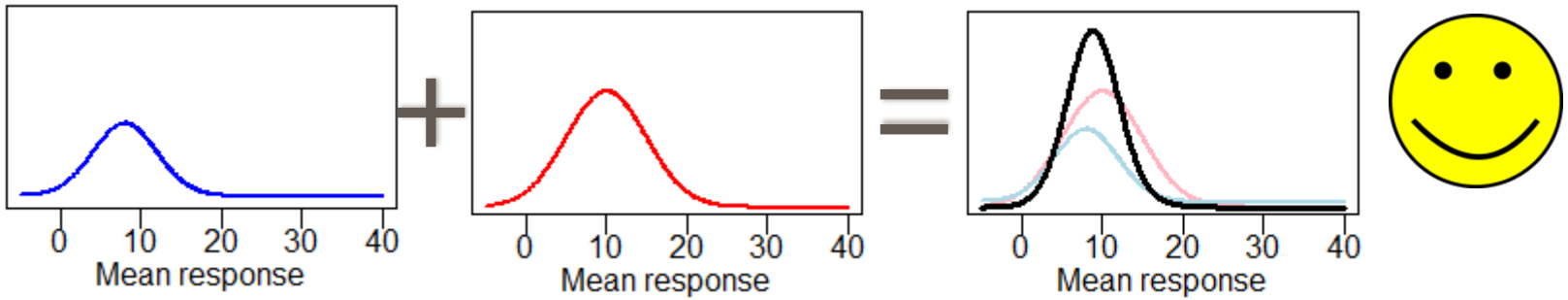
Mixture prior with 50% weight on historical data and 50% weight on flat prior

What we see in the new study ("sampling distribution")

"Posterior" distribution: weighted average of prior and new trial data

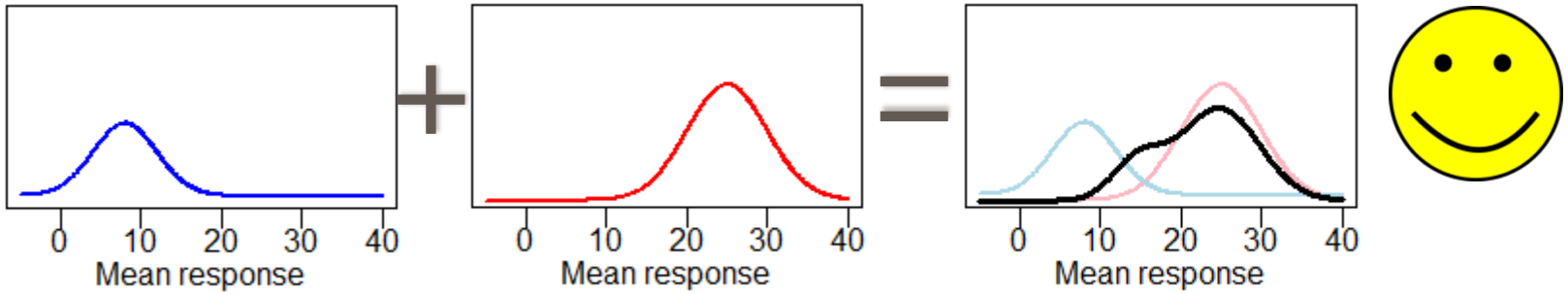
**Scenario 1**

Historical and new data are consistent




**Scenario 2**

Historical and new data in conflict



# Back to case study

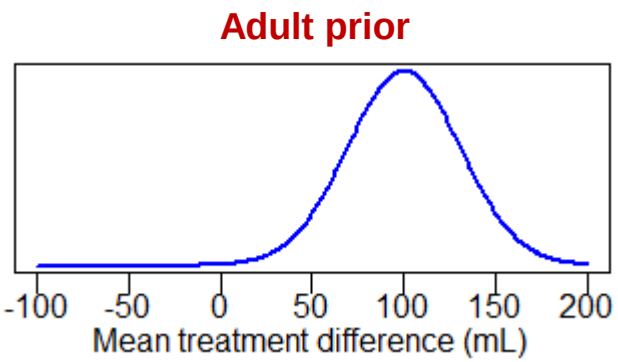
# Building the prior

## 1. Results from Adult Phase III trial

Experimental arm: N=337

Control arm: N=337

Δ = 100 ml
SD = 400 ml



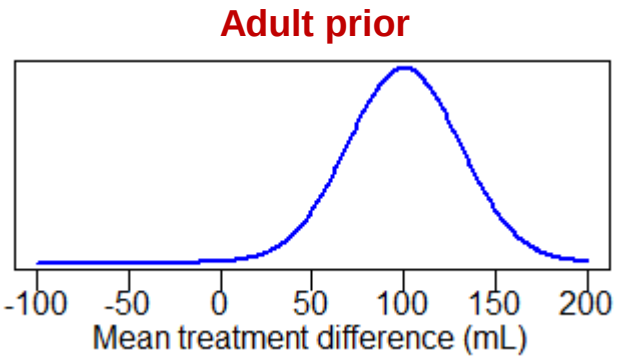
# Building the prior

## 1. Results from Adult Phase III trial

Experimental arm: N=337

Control arm: N=337

Δ = 100 ml
SD = 400 ml






## 2. "Bridging" Data

- Competitor programme comparing active regimen to placebo in same source & target populations
- Used to estimate **bridging factors** to scale Δ and SD to adolescents

| Population | N | Δ | $\frac{\Delta_{adolescent}}{\Delta_{adult}}$ | SD | $\frac{SD_{adolescent}}{SD_{adult}}$ |
|---|---|---|---|---|---|
| Adolescents | 131 | 54 ml | | 664 ml | |
| Adults (study 1) | 229 | 88 ml | 0.61 | 469 ml | 1.42 |
| Adults (study 2) | 224 | 111 ml | 0.49 | 449 ml | 1.48 |

# Building the prior

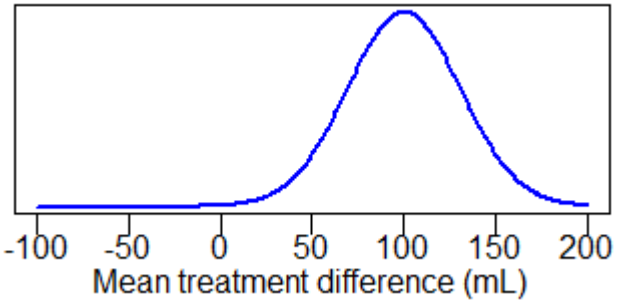
## 1. Assumed Results from (ongoing) Adult Phase III trial

Experimental arm: N=337

Control arm: N=337

Δ = 100 ml
SD = 400 ml

**Adult prior**





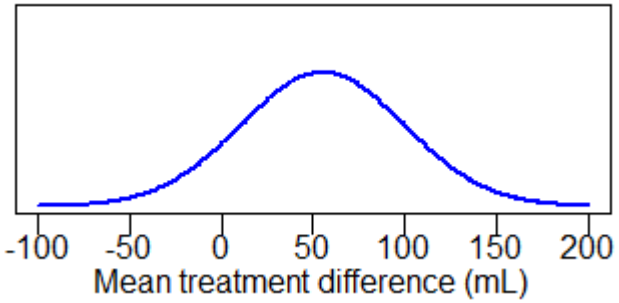

## 2. "Bridging" Data

- Competitor programme comparing active regimen to placebo in same source & target populations
- Used to estimate **bridging factors** to scale Δ and SD to adolescents

| Population | N | Δ | $\frac{\Delta_{adolescent}}{\Delta_{adult}}$ | SD | $\frac{SD_{adolescent}}{SD_{adult}}$ |
|---|---|---|---|---|---|
| Adolescents | 131 | 54 ml | | 664 ml | |
| Adults (study 1) | 229 | 88 ml | 0.61 | 469 ml | 1.42 |
| Adults (study 2) | 224 | 111 ml | 0.49 | 449 ml | 1.48 |

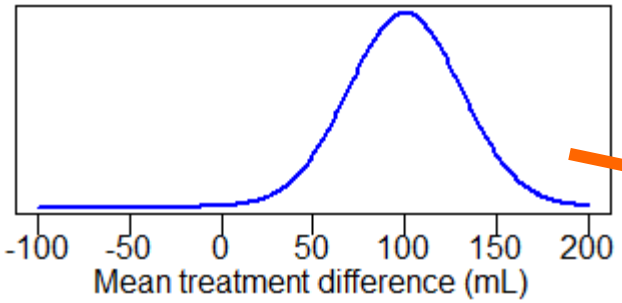
**0.55** **1.45**

**Rescaled adult prior:**
**Δ = 0.55×100 ml; SD = 1.45 ×400 ml**



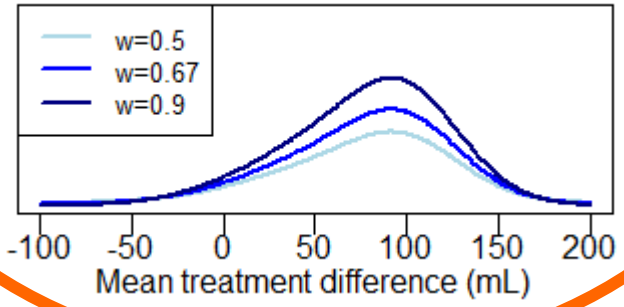
# Building the prior

**Adult prior**




**w/2**

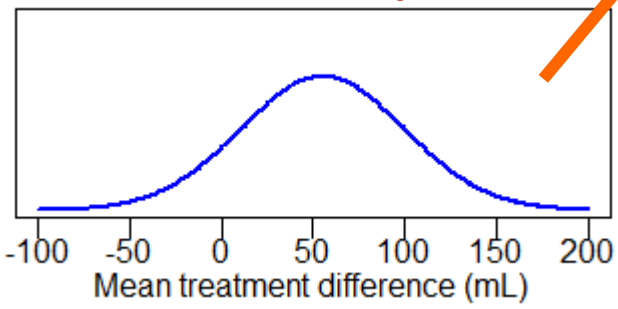
**3 component robust mixture prior for adolescent treatment effect**



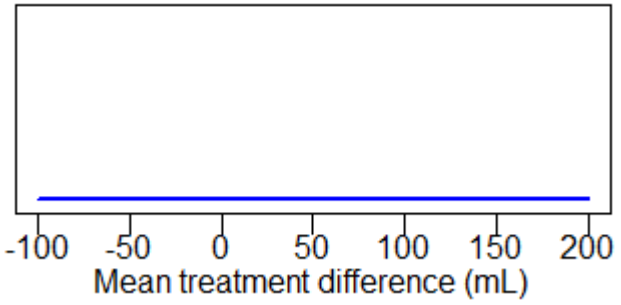

**w/2**

**1-w**

**Rescaled adult prior**



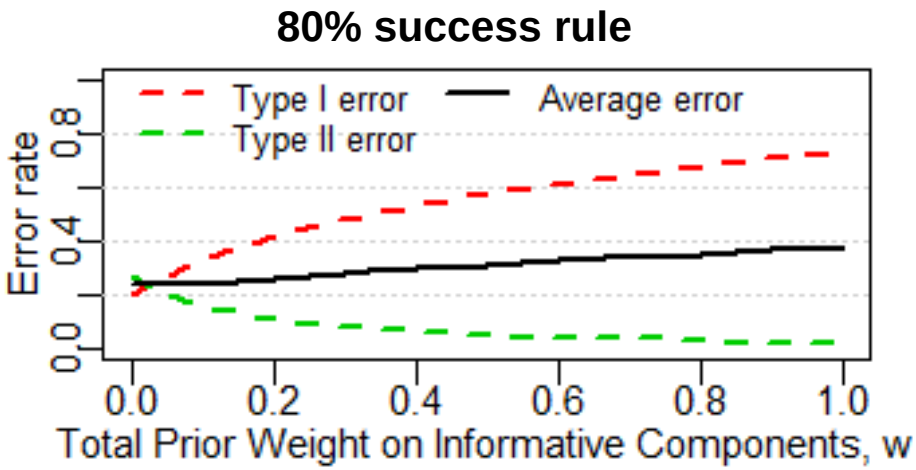

**Vague (unit information) prior**



# Simulation study to investigate frequentist operating characteristics of partial extrapolation design

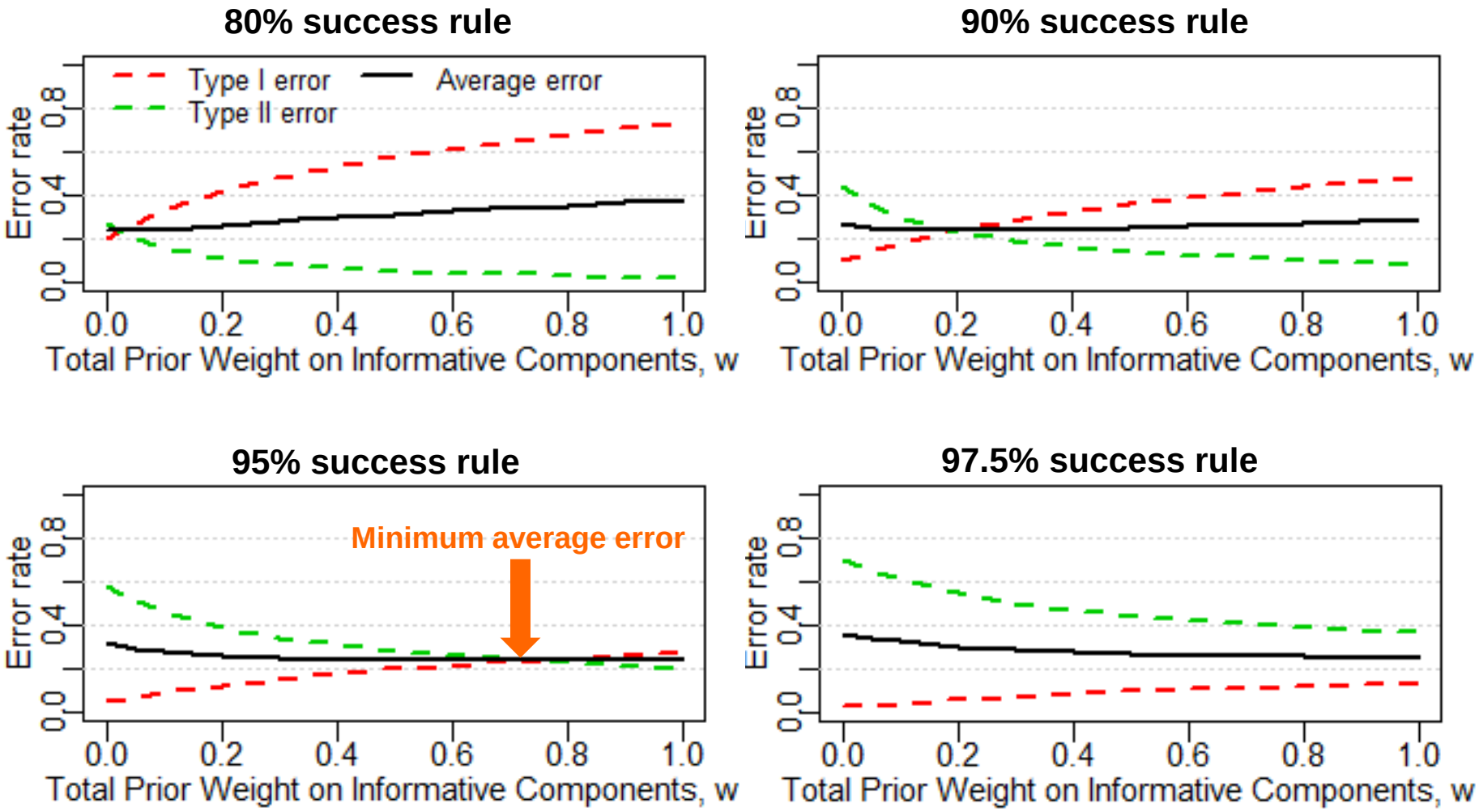
- Prior weights for mixture components should reflect clinical judgement about relevance of adult data to the target (adolescent) population
  - But also need to ensure adult information does not dominate
- Simulation study to explore
  1. Choice of w = “prior extrapolation weight” (total prior weight on informative components, split equally on full & rescaled) for w ∈ (0, 0.1, ..., 0.9,1)
  2. Choice of success rule = Pr(Δ > 0|data) > q% for q ∈ (80%, 90%, 95%, 97.5%)
- Generated 100,000 replicate datasets of size N=130 per arm for adolescent trial under null (Δ =0 ml) and alternative (Δ =90 ml)
- Operating characteristics:
  - Average probability of an error (type I or type II)
  - Prior and posterior precision (halfwidth of 95% CrI for Δ)
  - Prior and posterior weights on each component
  - Effective sample size

# Error Rates




Average error rates: see Grieve (2015). How to test hypotheses if you must. *Pharmaceutical Statistics*

# Error Rates

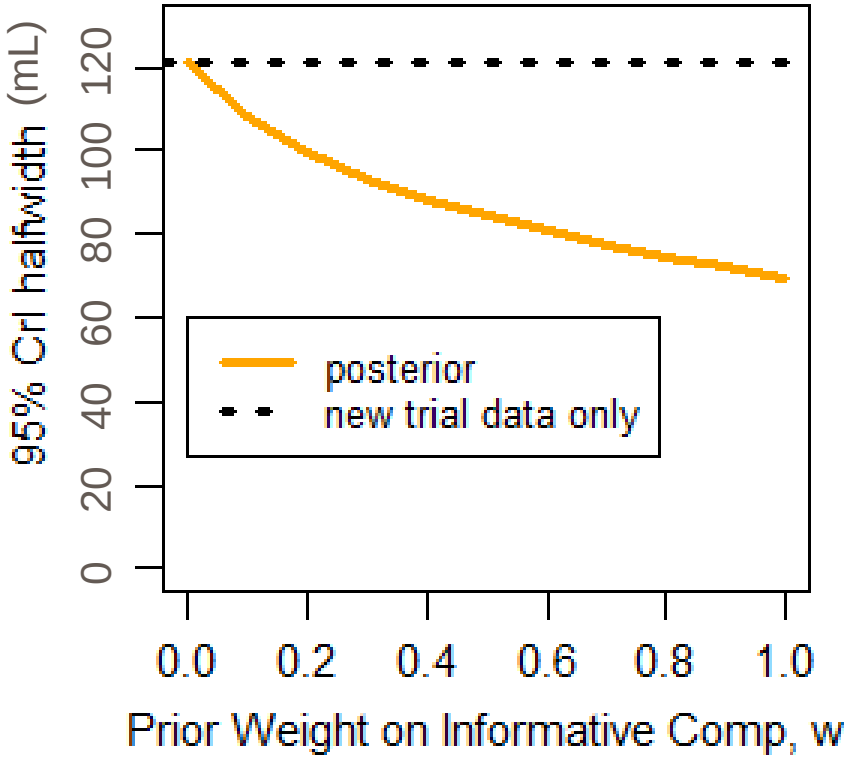



Average error rates: see Grieve (2015). How to test hypotheses if you must. *Pharmaceutical Statistics*

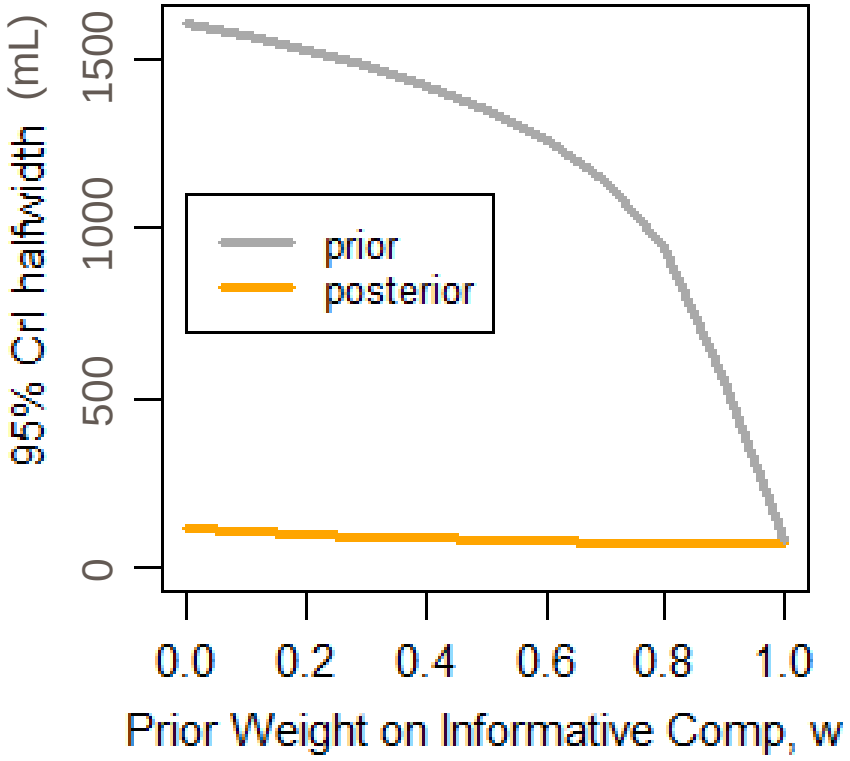
# Precision (95% CrI halfwidth for treatment effect)

## Average precision under alternative (Δ =90 ml)

### Comparing precision of adolescent data alone versus posterior with partial extrapolation




### Comparing prior and posterior precision



# Posterior Effective Sample Size (ESS)

**Proposal:** Define "effective sample size" of posterior as the 'standalone' sample size required to estimate the mean treatment effect with an equivalent margin of error (95% CI halfwidth)

**Example:**

- Suppose posterior 95% CrI HW from extrapolation design = 80 mL
- Assume adolescent sampling SD, $\sigma$ = 500 mL
- 'Standalone' sample size (per arm) required to estimate mean treatment difference $\Delta$ to within ± 80 mL at the 95% confidence level is

$$N = (1.96^2 \times 2\sigma^2)/HW^2 = 300$$

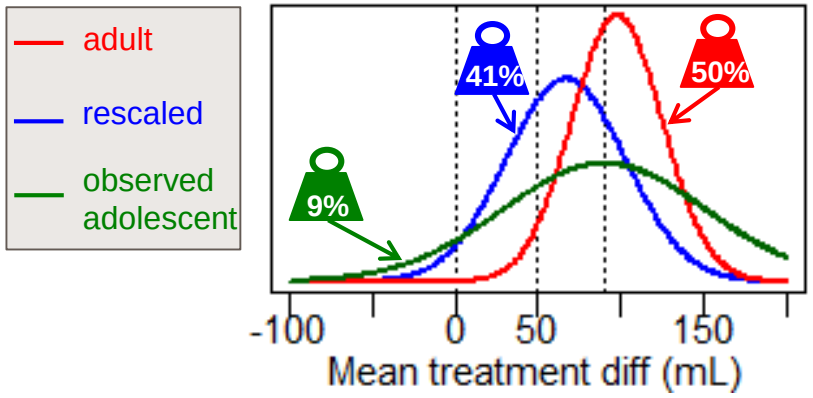
- Actual adolescent sample size is 130 per arm
- Information (gain in precision) borrowed from the prior is equivalent to 300-130 = 170 subjects per arm

# Prior and Posterior weights and ESS

## Scenario 1: Observed adolescent Δ = 90 mL

**Consider a total prior extrapolation weight = 0.5**

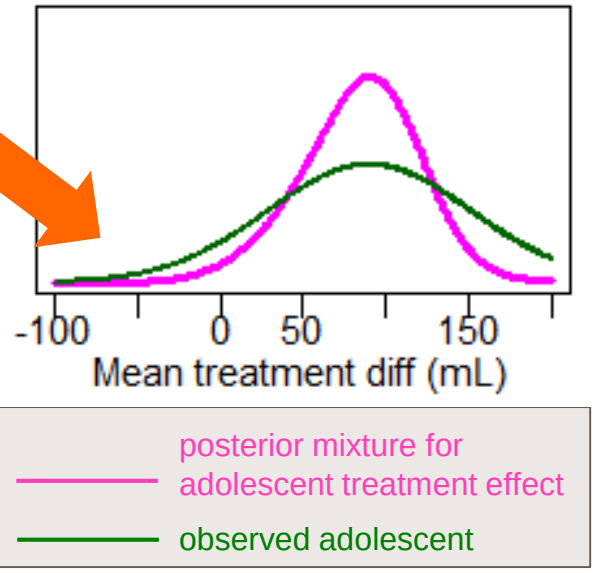
**Updated (posterior) mixture components**




Posterior 95% CrI HW = **75.3 mL**

Posterior ESS = **339**

ESS borrowed from prior = **209**

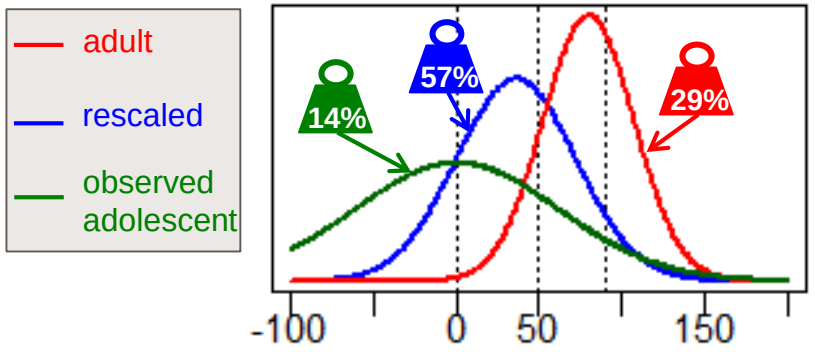


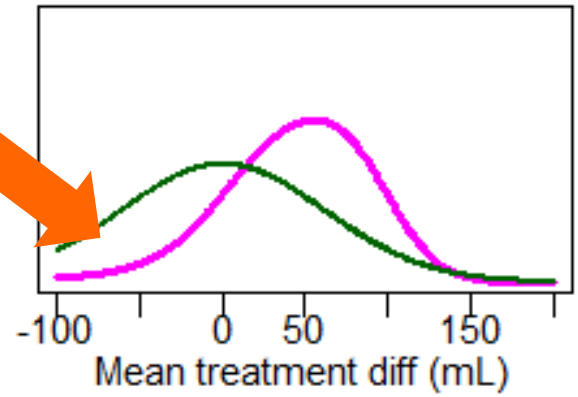
# Prior and Posterior weights and ESS

## Scenario 1: Observed adolescent Δ = 0 mL

**Consider a total prior extrapolation weight = 0.5**

**Updated (posterior) mixture components**




Posterior 95% CrI HW = **91.8 mL**

Posterior ESS = **228**

ESS borrowed from prior = **98**

# Proposed extrapolation design

Proposed extrapolation design:

**95% success rule**

**Total prior weight on informative components = 0.5**

# Proposed extrapolation design: choice of prior weight

Justification for choice of prior weight = 0.5

- 50:50 (coin toss) prior belief that adult efficacy can be extrapolated to adolescents
- Conservative assumption reflecting
  - disease progression and response to intervention expected to be similar
  - adolescents and adults have previously been studied together

  whilst recognising that key differences (e.g. adherence, lung function, prevalence) may exist
- Further conservatism introduced by splitting prior weight equally between directly extrapolating from adults and extrapolation from rescaled adult evidence

# Proposed extrapolation design: how much is borrowed?

| Observed Δ | Posterior weight on: | | Posterior ESS - 130 | Posterior Pr(Δ>0) | Posterior Pr(Δ>0) for reference design (no extrapolation) |
|---|---|---|---|---|---|
| | Adult component | Re-scaled component | | | |
| 90 mL | 0.50 | 0.41 | 209 | **98%** | **93%** |
| 50 mL | 0.42 | 0.49 | 178 | **95%** | **79%** |
| 0 mL | 0.29 | 0.57 | 98 | **84%** | **50%** |
| -90 mL | 0.07 | 0.44 | -27 | **35%** | **7%** |

Prior Pr(Δ>0) = 72%

# Comparison of designs

Compared to reference design (80% success rule + prior weight = 0), proposed extrapolation design (95% success rule + prior weight = 0.5) has:

- Similar power (~72%) and type I error (~20%)
  - In long run, adopting either design strategy for regulatory decision making would result in similar rates of correct / incorrect decisions
- More stringent evidentiary threshold for demonstrating efficacy in adolescents
  - Extrapolation design: total weight of evidence (adult efficacy + prior belief in extrapolation + adolescent data) sufficient to demonstrate efficacy of this drug in adolescents with at least 95% confidence
  - Reference design: total weight of evidence (adolescent data) sufficient demonstrate efficacy of this drug in adolescents with at least 80% confidence
- Increased precision of treatment effect estimate (~30% reduction in 95% CrI HW)

# Concluding remarks

## Benefits of Bayesian partial extrapolation design

- Provides reasonable power to demonstrate a treatment benefit with a pragmatic sample size, whilst maintaining acceptable evidentiary threshold
  - Contrast with low powered standalone adolescent/paediatric study where conclusions are often made on trending data
- Provides substantial gains in precision of treatment effect estimate
- Degree of "similarity" required to demonstrate adolescent efficacy is quantifiable (prior and posterior beliefs in extrapolation assumption)
- Amount of information borrowed (ESS) is a function of the strength of evidence supporting extrapolation

## Challenges

- Requires initial assessment of probability of extrapolation, which is often difficult to make
  - precise amount of belief to place on similarity of adults & adolescents may be unclear
  - who should make this assessment? (Sponsors, Regulators, 3rd party….?)
- If source (adult) and target (adolescent) data conflict, result can be biased treatment effect estimates and inflation of type 1 error
  - Robust dynamic Bayesian methods reduce risk by down-weighting adult data when conflict present

# Thank you for listening