Use of frequentist and Bayesian approaches for extrapolating from adult efficacy data to design and interpret confirmatory trials in children

Lisa Hampson, Franz Koenig and Martin Posch

Department of Mathematics & Statistics, Lancaster University

PSI Scientific Meeting: Extrapolation
22nd November 2017

Acknowledgements

This research was funded through a MRC Career Development Award in Biostatistics and a MRC-NIHR Methodology Research Programme grant (LVH). The work of FK and MP has received funding from the European Unions Seventh Framework Programme for research, technological development and demonstration under Grant Agreement no 602552.

We are grateful for helpful discussions with Peter Bauer and Gerald Hlavin.

Developing new medicines for children

Trials have been run predominantly in adults leading to off-label prescribing in children.

Since 2007, EU Paediatric Regulation mandates medicine development in children:

- Sponsor must submit a Paediatric Investigation Plan (PIP) application to European Medicines Agency's Paediatric Committee before adult PK completed.
- PIP outlines all aspects of the development programme in children.
- Similar regulations exist in US, e.g., Pediatric Research Equity Act.

Developing new medicines for children

Trials have been run predominantly in adults leading to off-label prescribing in children.

Since 2007, EU Paediatric Regulation mandates medicine development in children:

- Sponsor must submit a Paediatric Investigation Plan (PIP) application to European Medicines Agency's Paediatric Committee before adult PK completed.
- PIP outlines all aspects of the development programme in children.
- Similar regulations exist in US, e.g., Pediatric Research Equity Act.

Issues of ethics and feasibility mean we wish to limit experimentation in children.

Extrapolation
European Medicines Agency (2012)

European Medicines Agency defines extrapolation as:

'Extending information and conclusions available from studies in one or more subgroups of the patient population (source population) . . . to make inferences for another subgroup of the population (target population) . . . '

Extrapolation

European Medicines Agency (2012)

European Medicines Agency defines extrapolation as:

'Extending information and conclusions available from studies in one or more subgroups of the patient population (source population) . . . to make inferences for another subgroup of the population (target population) . . . '

Can we extrapolate from adult efficacy data to reduce the size of the paediatric trial needed to demonstrate benefit in this age group?

Notation and assumptions

We will use the following notation:

- θ_A measures the advantage of a new therapy relative to placebo in adults.
- θ_C denotes the corresponding effect in children.
- $\hat{\theta}_A$ and $\hat{\theta}_C$ denote maximum likelihood estimates of effects.

We will rely on the following assumptions:

- Responses are continuous with known common variance, so effects represent differences in average outcomes.
- Unit difference in expected outcomes is a clinically meaningful effect.
- Common known response variance of 1 in adults and children.
- A single positive adult Phase III trial is sufficient to justify licensing in adults.
- A paediatric Phase III trial is conducted only if the adult trial is significant.

Extrapolating from adult efficacy data

Hlavin, Koenig, Male et al. (2016)

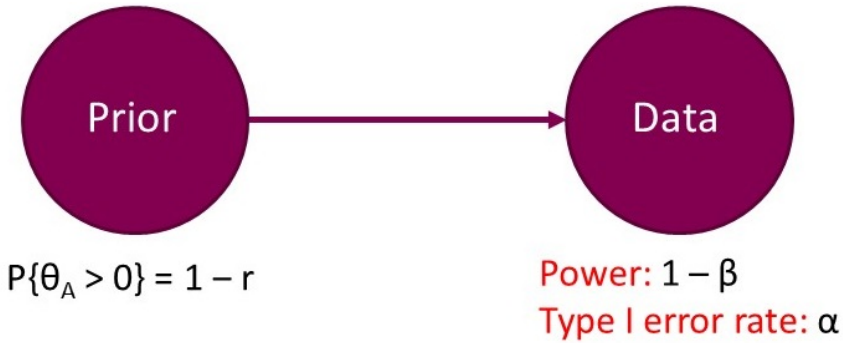
Suppose we first conduct a Phase III trial in adults to test $H_{0A} : \theta_A \leq 0$ versus $\theta_A > 0$.

Prior: 2 point prior based on opinion or historical success rates across similar drugs.

Extrapolating from adult efficacy data

Hlavin, Koenig, Male et al. (2016)

Suppose we first conduct a Phase III trial in adults to test $H_{0A} : \theta_A \leq 0$ versus $\theta_A > 0$.



Prior: 2 point prior based on opinion or historical success rates across similar drugs.

Data: Adult Phase III trial is designed to control frequentist operating characteristics.

Extrapolating from adult efficacy data

Hlavin, Koenig, Male et al. (2016)

Suppose we first conduct a Phase III trial in adults to test $H_{0A} : \theta_A \leq 0$ versus $\theta_A > 0$.

$P\{\theta_A > 0\} = 1 - r$

Sample Size: $n = 4\sigma^2\{\Phi^{-1}(1-\alpha) + \Phi^{-1}(1-\beta)\}^2/\delta^2$

Success Criterion: $\hat{\theta}_A \geq \Phi^{-1}(1-\alpha)\sqrt{4\sigma^2/n}$

Prior: 2 point prior based on opinion or historical success rates across similar drugs.

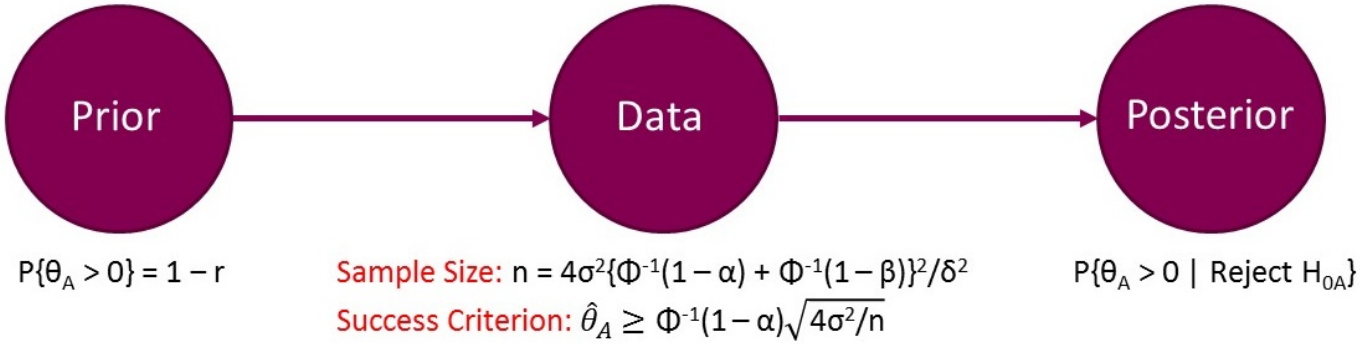
Data: Adult Phase III trial is designed to control frequentist operating characteristics.

Data: Desired power and type I error rate determines sample size and success criterion.

Extrapolating from adult efficacy data

Hlavin, Koenig, Male et al. (2016)

Suppose we first conduct a Phase III trial in adults to test $H_{0A} : \theta_A \leq 0$ versus $\theta_A > 0$.



Prior: 2 point prior based on opinion or historical success rates across similar drugs.

Data: Adult Phase III trial is designed to control frequentist operating characteristics.

Data: Desired power and type I error rate determines sample size and success criterion.

Posterior: If trial is successful, weight of evidence supporting superiority can be summarised by the average positive predictive value of the decision to reject H_{0A}.

Extrapolation model

Hlavin, Koenig, Male et al. (2016)

Can we borrow strength from the adult result to test $H_{0C} : \theta_C \leq 0$ vs $\theta_C > 0$?

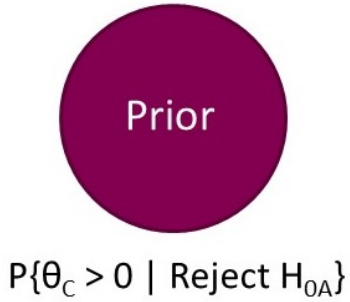
Hlavin et al. countenance two extrapolation scenarios:

Full extrapolation (prob. $1 - s$): $\text{pr}\{\theta_C > 0 \mid \text{Reject } H_{0A}\} = \text{pr}\{\theta_A > 0 \mid \text{Reject } H_{0A}\}$
No extrapolation (prob. s): $\text{pr}\{\theta_C > 0 \mid \text{Reject } H_{0A}\} = 1 - q$.

Extrapolation model

Hlavin, Koenig, Male et al. (2016)

Extrapolation model implies we design the paediatric test of $H_{0C} : \theta_C \leq 0$ as follows:

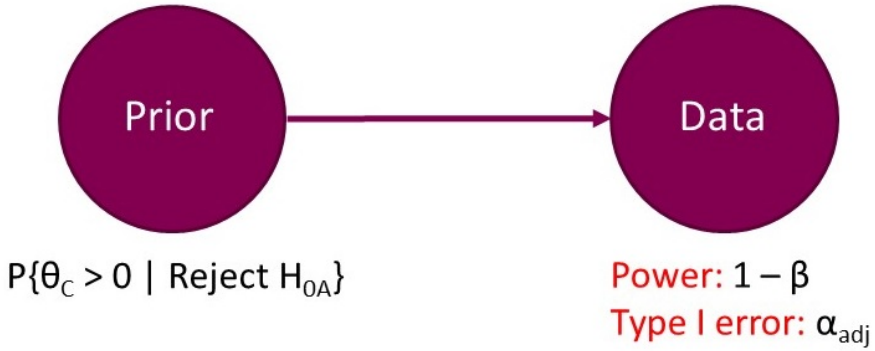


Prior: 2 points: $\text{pr}\{\theta_C > 0 \mid \text{Reject } H_{0A}\} = (1 - s)\,\text{pr}\{\theta_A > 0 \mid \text{Reject } H_{0A}\} + s\,(1 - q)$.

Extrapolation model

Hlavin, Koenig, Male et al. (2016)

Extrapolation model implies we design the paediatric test of $H_{0C} : \theta_C \leq 0$ as follows:



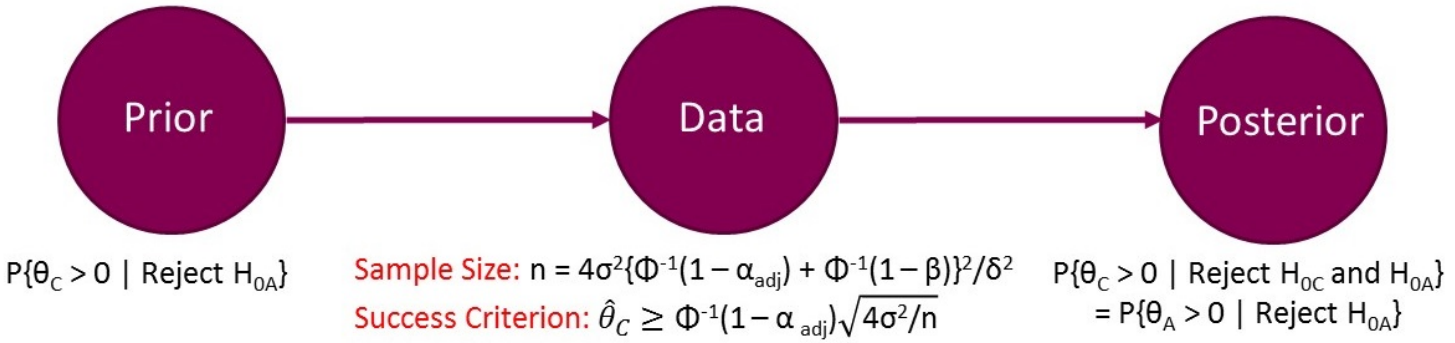
Prior: 2 points: $\text{pr}\{\theta_C > 0 \mid \text{Reject } H_{0A}\} = (1-s)\,\text{pr}\{\theta_A > 0 \mid \text{Reject } H_{0A}\} + s\,(1-q)$.

Data: Paediatric Phase III trial is designed to have power $1-\beta$ at $\theta_C = \delta$
and type I error rate calibrated so that if we reject H_{0C} ...

Extrapolation model

Hlavin, Koenig, Male et al. (2016)

Extrapolation model implies we design the paediatric test of $H_{0C} : \theta_C \leq 0$ as follows:



Prior: 2 points: $\text{pr}\{\theta_C > 0 \mid \text{Reject } H_{0A}\} = (1 - s)\,\text{pr}\{\theta_A > 0 \mid \text{Reject } H_{0A}\} + s\,(1 - q)$.

Data: Paediatric Phase III trial is designed to have power $1 - \beta$ at $\theta_C = \delta$

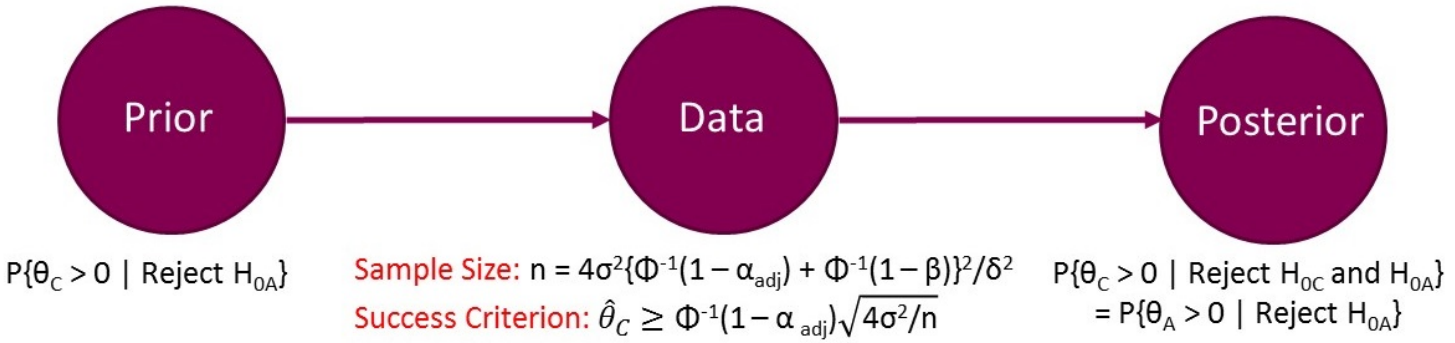
and type I error rate calibrated so that if we reject H_{0C} . . .

Posterior: . . . the total weight of evidence supporting efficacy in children (given significant adult and paediatric tests) equals evidence that supported adult licensing.

Extrapolation model

Hlavin, Koenig, Male et al. (2016)

Extrapolation model implies we design the paediatric test of $H_{0C} : \theta_C \leq 0$ as follows:



Prior: 2 points: $\text{pr}\{\theta_C > 0 \mid \text{Reject } H_{0A}\} = (1 - s)\,\text{pr}\{\theta_A > 0 \mid \text{Reject } H_{0A}\} + s\,(1 - q)$.

Data: Paediatric Phase III trial is designed to have power $1 - \beta$ at $\theta_C = \delta$

and type I error rate calibrated so that if we reject H_{0C} . . .

Posterior: . . . the total weight of evidence supporting efficacy in children (given significant adult and paediatric tests) equals evidence that supported adult licensing.

$\alpha_{adj} > \alpha$ leads to a reduction in the paediatric sample size.

Extrapolating from adult efficacy data

Bauer and Koenig (2016)

Hlavin method can be thought of as a *hybrid* Bayesian-frequentist approach:

- Paediatric trial is analysed using conventional, frequentist, methods . . .
- . . . but α is adjusted with a Bayesian interpretation of the data in mind.

Extrapolating from adult efficacy data

Bauer and Koenig (2016)

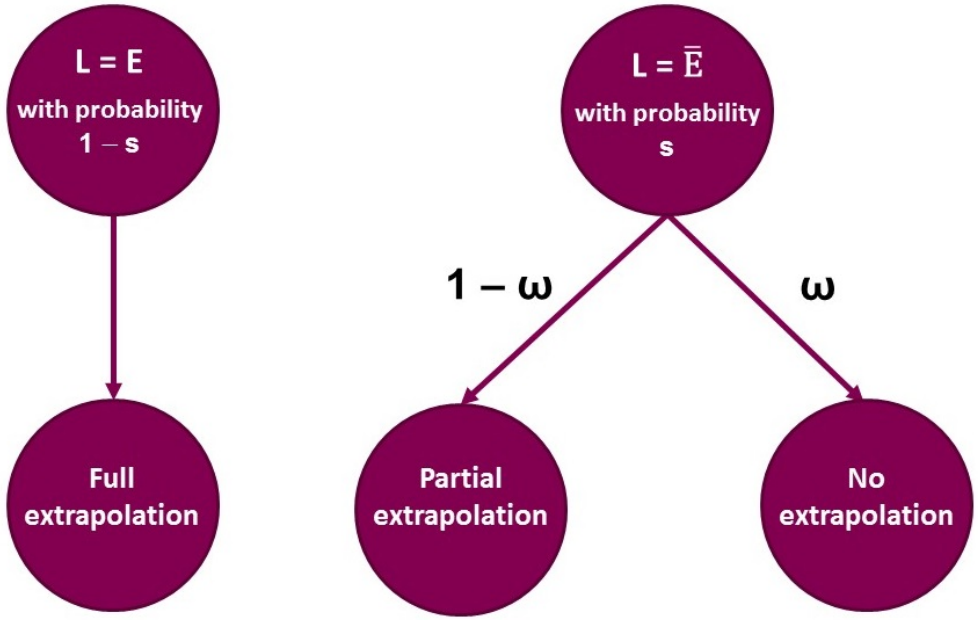
Hlavin method can be thought of as a *hybrid* Bayesian-frequentist approach:

- Paediatric trial is analysed using conventional, frequentist, methods . . .
- . . . but α is adjusted with a Bayesian interpretation of the data in mind.
- Working with two point priors implies α_{adj} is found calibrating lower bounds of positive predictive values of tests.
- Only condition on event $\{\text{Reject } H_{0A}\}$ so paediatric design can be fixed in advance.
- But, $\hat{\theta}_A$ will be known before the paediatric trial begins so makes more sense to condition on this value. This would lead to an adaptive PIP.

How would this method compare with taking a fully Bayesian approach?

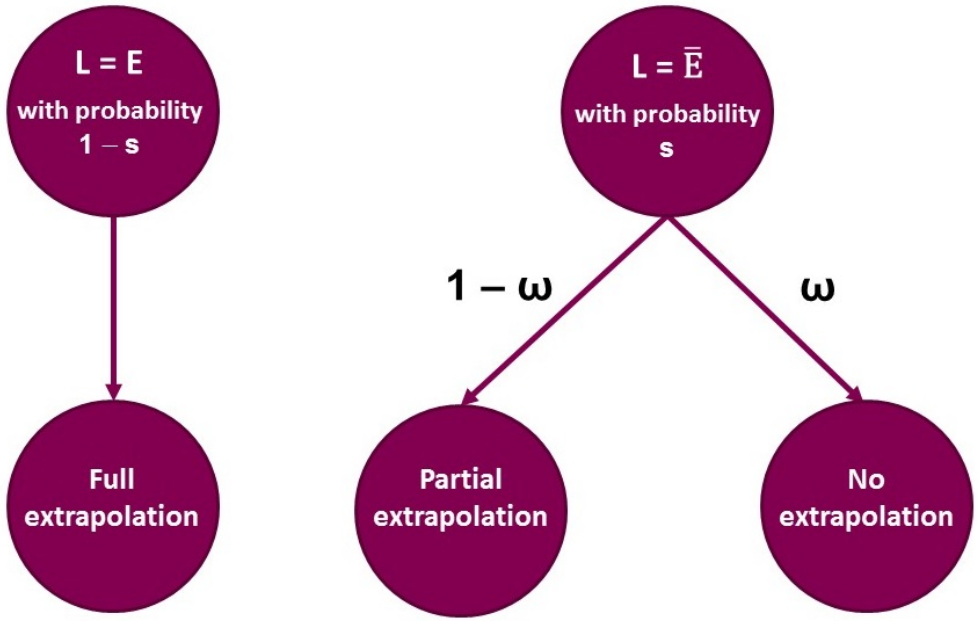
Bayesian extrapolation model

Below is a Bayesian mixture model representing prior opinion on θ_A and θ_C.
In all cases, $\theta_A \sim N(\mu_A, \sigma_A^2)$.



Bayesian extrapolation model

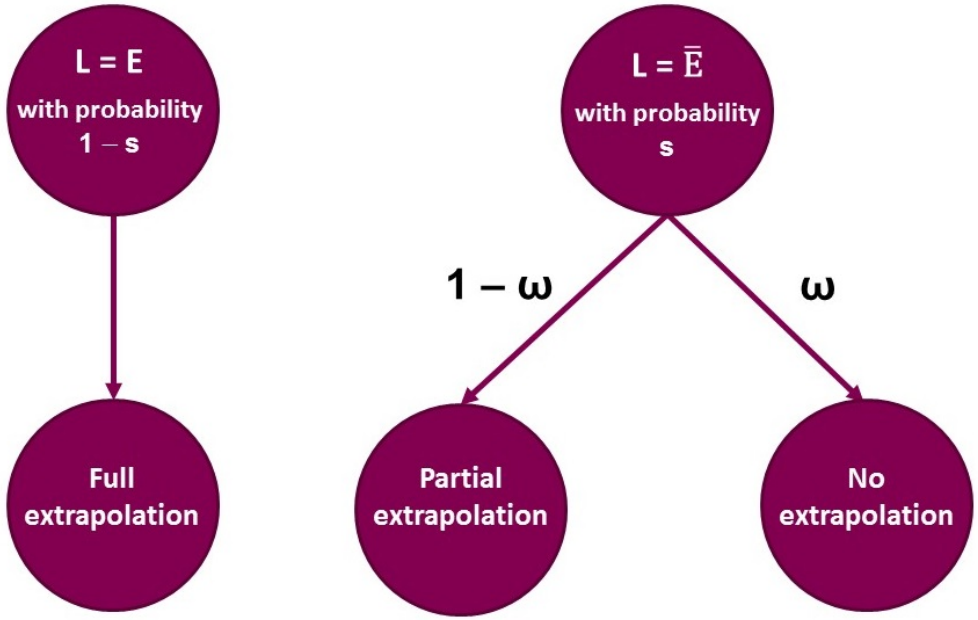
Below is a Bayesian mixture model representing prior opinion on θ_A and θ_C.
In all cases, $\theta_A \sim N(\mu_A, \sigma_A^2)$.



Full extrapolation: Treatment effects are qualitatively similar: $\theta_C = b\,\theta_A$ with $b > 0$.

Bayesian extrapolation model

Below is a Bayesian mixture model representing prior opinion on θ_A and θ_C.
In all cases, $\theta_A \sim N(\mu_A, \sigma_A^2)$.

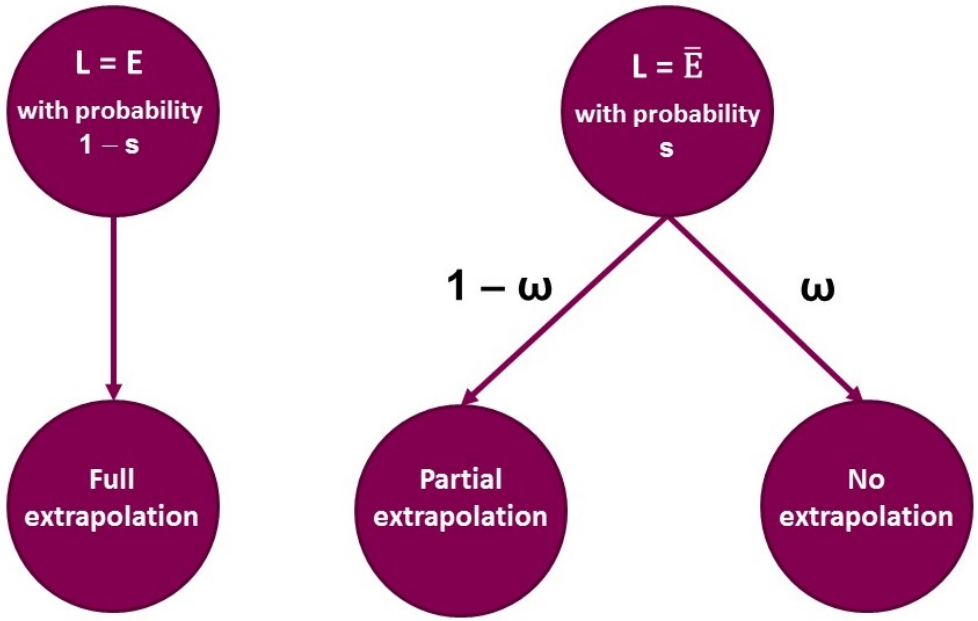


Partial extrapolation: Could be qualitative *and* quantitative differences between effects.
$\rho > 0$ represents opinion on degree of similarity.

$$\begin{pmatrix} \theta_A \\ \theta_C \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_A \\ \mu_{C,1} \end{pmatrix}, \begin{pmatrix} \sigma_A^2 & \rho\sigma_A\sigma_{C,1} \\ \rho\sigma_A\sigma_{C,1} & \sigma_{C,1}^2 \end{pmatrix} \right)$$

Bayesian extrapolation model

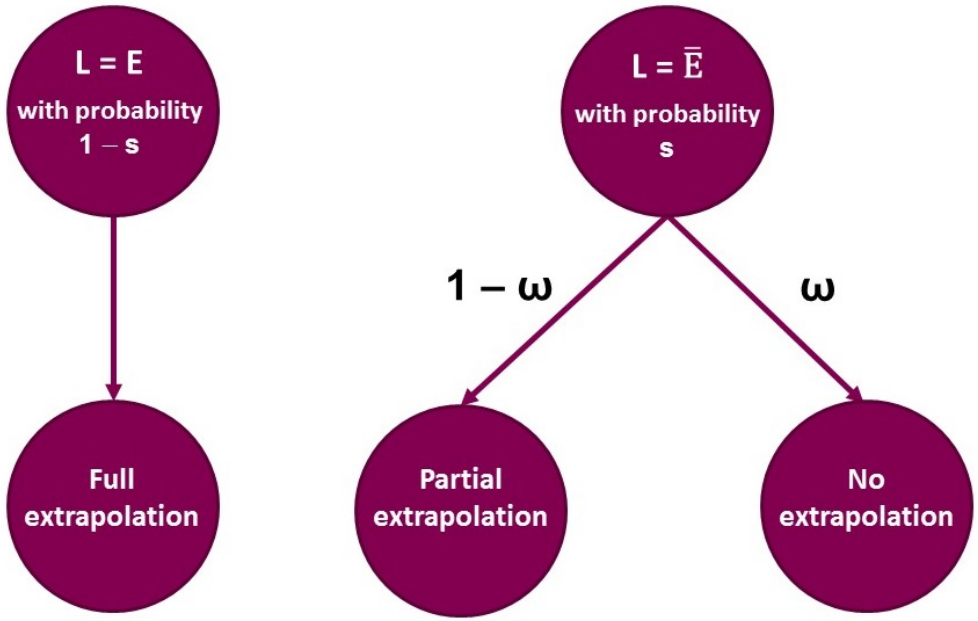
Below is a Bayesian mixture model representing prior opinion on θ_A and θ_C.
In all cases, $\theta_A \sim N(\mu_A, \sigma_A^2)$.



No extrapolation: Differences between age groups are such that knowing θ_A would tell us nothing about a medicine's effect in children:

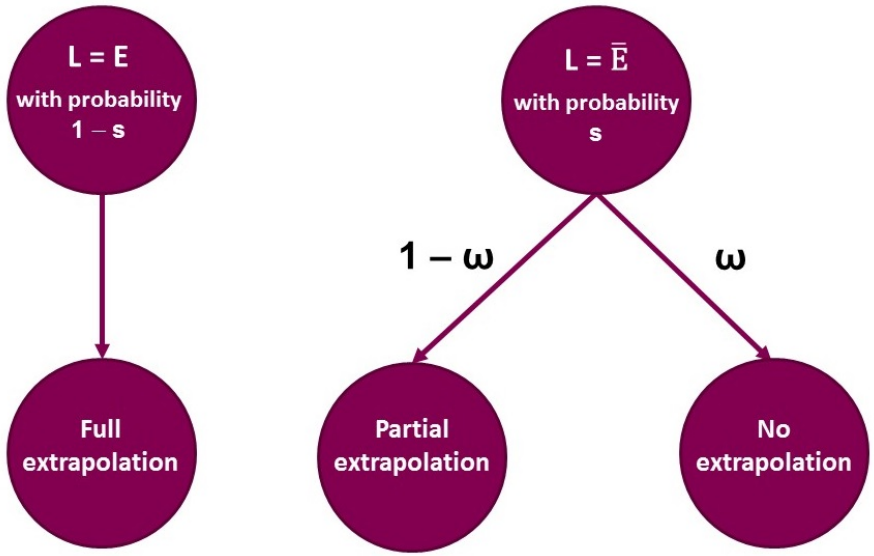
$$\begin{pmatrix} \theta_A \\ \theta_C \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_A \\ \mu_{C,2} \end{pmatrix}, \begin{pmatrix} \sigma_A^2 & 0 \\ 0 & \sigma_{C,2}^2 \end{pmatrix} \right)$$

Bayesian extrapolation model

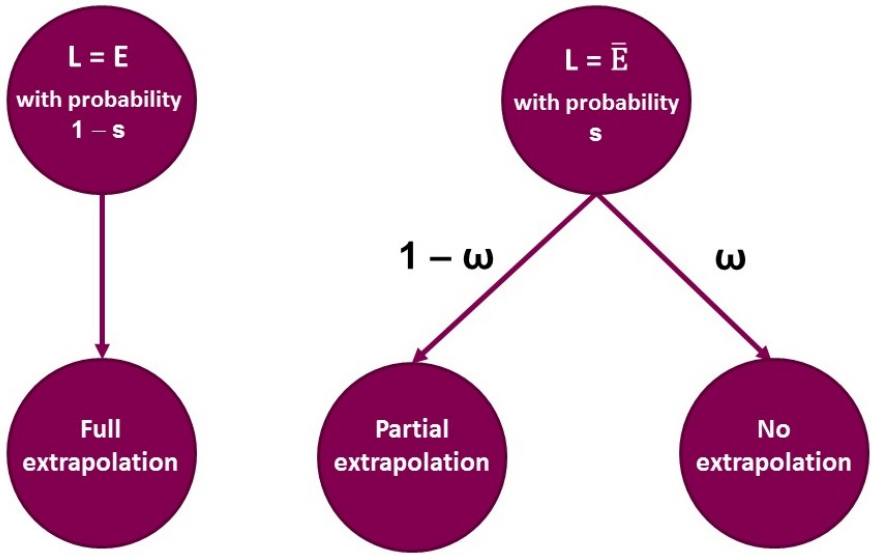


- Priors (for θ_A, θ_C and extrapolation scenarios) specified before adult trial.
- Beliefs on effect sizes are updated as data accumulate.
- Once data are available on both adults and children, opinion on the plausibility of three extrapolation scenarios will be updated.

Prior distributions

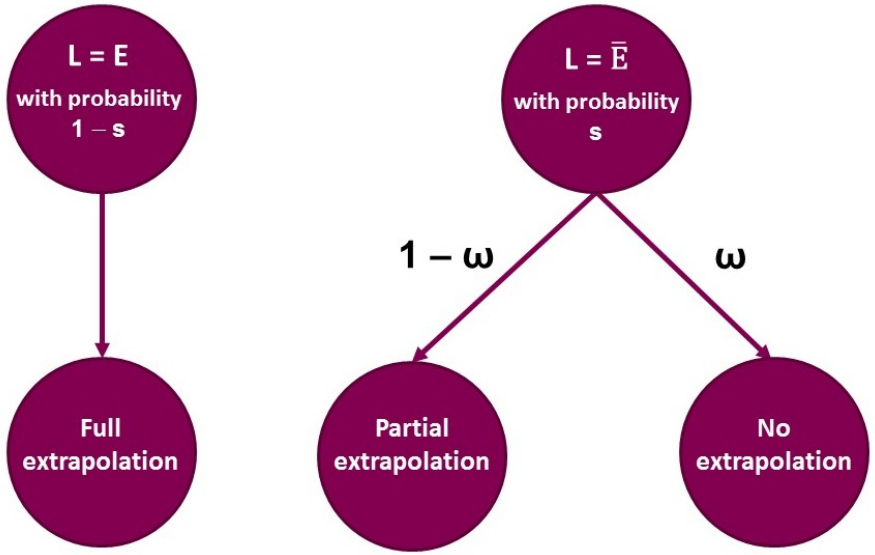


Prior distributions



Full extrapolation: Treatment effects are qualitatively similar: $\theta_C = 0.8\,\theta_A$, with $\theta_A \sim N(-0.5,\, 3.8)$.

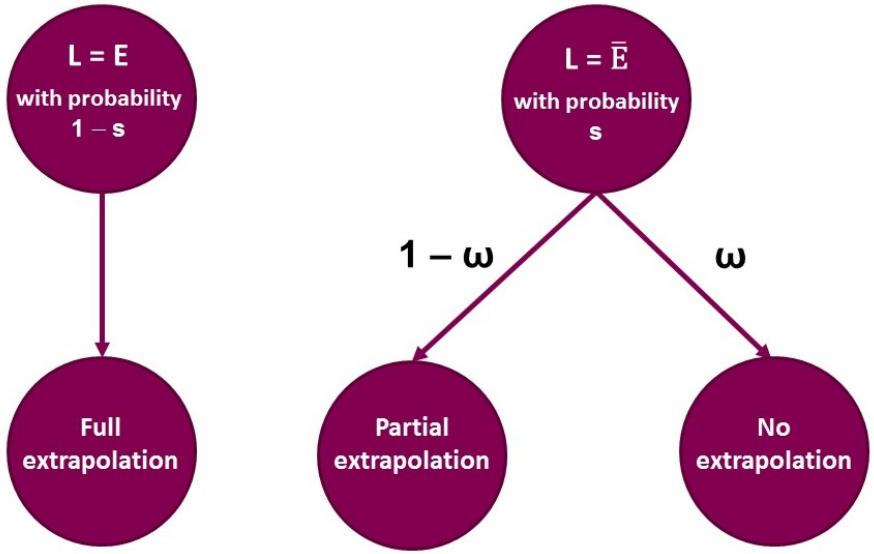
Prior distributions



Partial extrapolation: Set

$$\begin{pmatrix} \theta_A \\ \theta_C \end{pmatrix} \sim N\left(\begin{pmatrix} -0.5 \\ -0.5 \end{pmatrix}, \begin{pmatrix} 3.8 & 3.8 \times \rho \\ 3.8 \times \rho & 3.8 \end{pmatrix} \right)$$

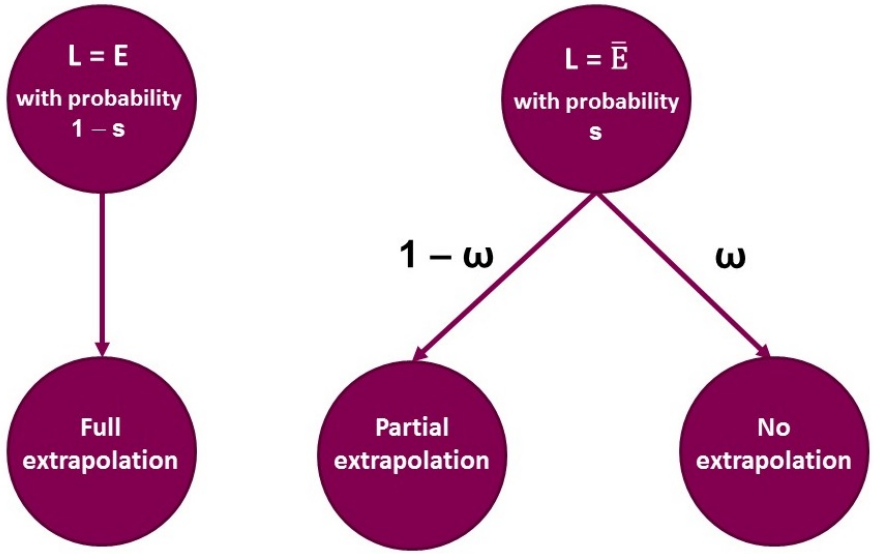
Prior distributions



No extrapolation: Set

$$\begin{pmatrix}\theta_A \\ \theta_C\end{pmatrix} \sim N\left(\begin{pmatrix}-0.5 \\ -0.5\end{pmatrix}, \begin{pmatrix}3.8 & 0 \\ 0 & 3.8\end{pmatrix}\right)$$

Prior distributions



No extrapolation: Set

$$\begin{pmatrix}\theta_A \\ \theta_C\end{pmatrix} \sim N\left(\begin{pmatrix}-0.5 \\ -0.5\end{pmatrix}, \begin{pmatrix}3.8 & 0 \\ 0 & 3.8\end{pmatrix}\right)$$

Under chosen priors, $\text{pr}\{\theta_A > 0 \mid \text{Reject } H_{0A}\} = 1 - \gamma = 0.998$. For consistency, we find Hlavin designs setting $1 - r = 0.22$.

Designing the paediatric trial: Bayesian Design 1 (BD1)

Objective: Control the **average** positive predictive value of decision to reject H_{0C}

Given the adult effect estimate, $\hat{\theta}_A$, choose the sample size (n_C) and success criterion (Reject H_{0C} if $\hat{\theta}_C \geq c^\star$) to ensure:

- Frequentist Power: $\text{pr}\{\hat{\theta}_C \geq c^\star \mid \theta_C = \delta\} \geq 1 - \beta$;
- Average +ve predictive value: $\text{pr}\{\theta_C > 0 \mid \hat{\theta}_C > c^\star, \hat{\theta}_A\} \geq 0.998$;

where 0.998 is the average positive predictive value of a significant Neyman-Pearson test of H_{0A} designed with $\alpha = 0.025$, $1 - \beta = 0.9$, $\delta = 1$ assuming $\theta_A \sim N(-0.5, 3.8)$.

Designing the paediatric trial: Bayesian Design 1 (BD1)

Objective: Control the **average** positive predictive value of decision to reject H_{0C}

Given the adult effect estimate, $\hat{\theta}_A$, choose the sample size (n_C) and success criterion (Reject H_{0C} if $\hat{\theta}_C \geq c^\star$) to ensure:

- Frequentist Power: $\text{pr}\{\hat{\theta}_C \geq c^\star \mid \theta_C = \delta\} \geq 1 - \beta$;
- Average +ve predictive value: $\text{pr}\{\theta_C > 0 \mid \hat{\theta}_C > c^\star, \hat{\theta}_A\} \geq 0.998$;

where 0.998 is the average positive predictive value of a significant Neyman-Pearson test of H_{0A} designed with $\alpha = 0.025$, $1 - \beta = 0.9$, $\delta = 1$ assuming $\theta_A \sim N(-0.5, 3.8)$.

Notes:

- If $\text{pr}\{\theta_C > 0 \mid \hat{\theta}_A\} \geq 0.998$, set $n_C = 0$ and reject H_{0C}.
- Constrain $c^\star \geq 0$.
- Search over $n_C \geq 1$ to find the smallest total sample size for which the paediatric trial has at least power $1 - \beta$.

Bayesian Design 2 (BD2)

Objective: Control the **minimum** positive predictive value of a decision to reject H_{0C}

Choose sample size (n_C) and success criterion (Reject H_{0C} if $\hat{\theta}_C \geq c^\star$) to ensure:

- Frequentist Power: $\text{pr}\{\hat{\theta}_C \geq c^\star \mid \theta_C = \delta\} \geq 1 - \beta$;
- Minimum +ve predictive value: Reject H_{0C} if $\text{pr}\{\theta_C > 0 \mid \hat{\theta}_C, \hat{\theta}_A\} \geq \eta$

where $\eta = 0.97$ is the smallest positive predictive value consistent with a significant test of H_{0A} setting $\alpha = 0.025$, $1 - \beta = 0.9$, $\delta = 1$, $\theta_A \sim N(-0.5, 3.8)$.

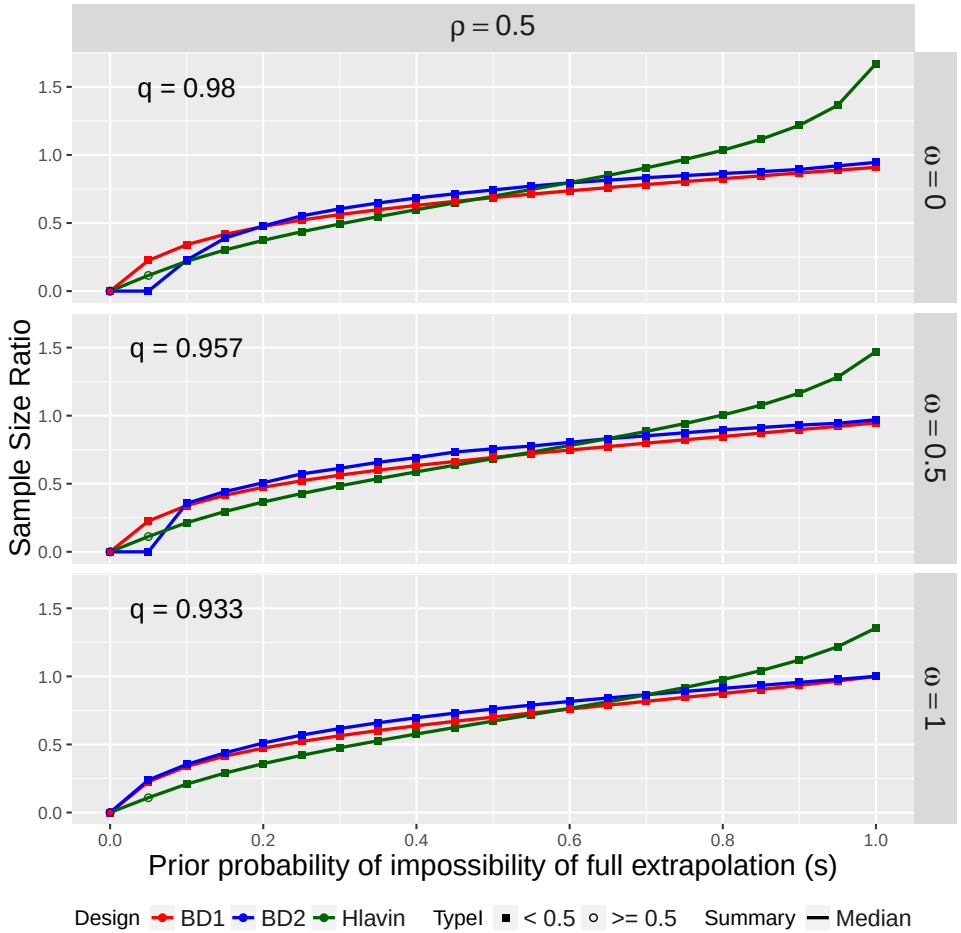
Reporting Bayesian designs

Bayesian designs depend on the estimate $\hat{\theta}_A$ generated by the adult trial.

At the time the PIP is written $\hat{\theta}_A$ will likely be unknown. In this case, we can calculate the prior predictive distribution of the trial's operating characteristics and report summaries of this.

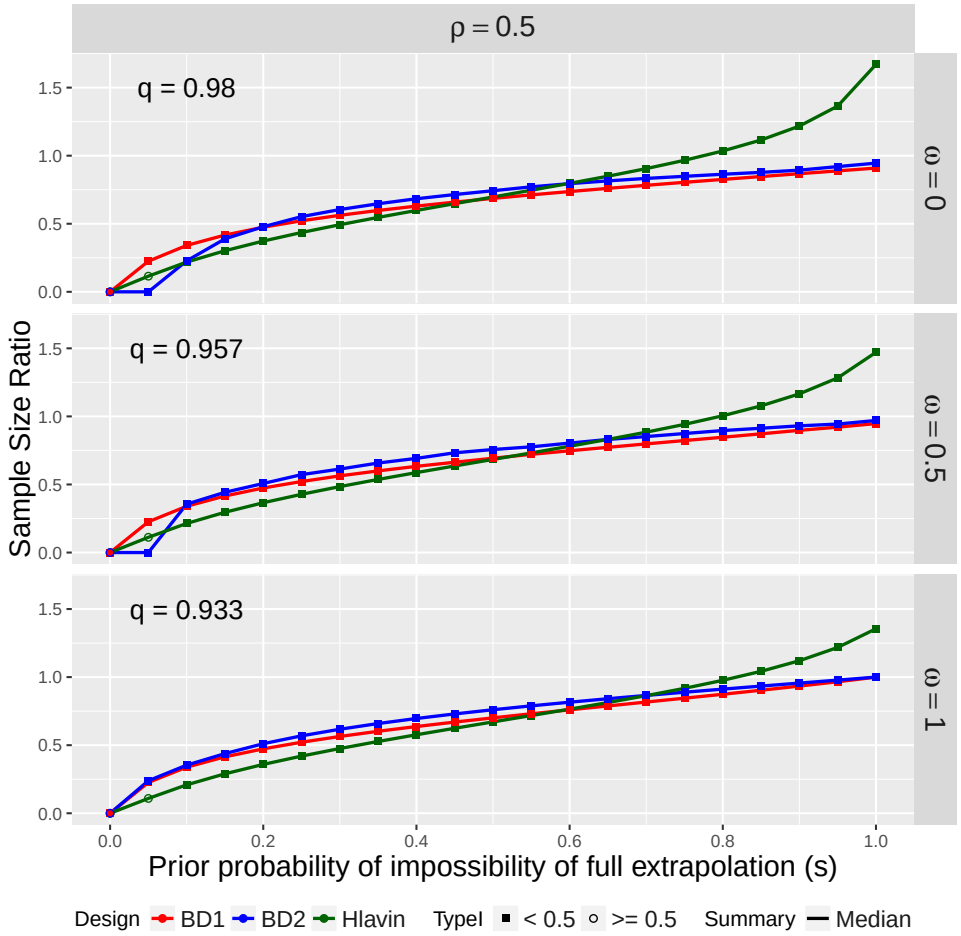
Results: paediatric sample size

Sample size expressed as ratio of number needed for Neyman-Pearson test of H_{0C} with $\alpha = 0.025$, $1 - \beta = 0.9$, $\delta = 1$.



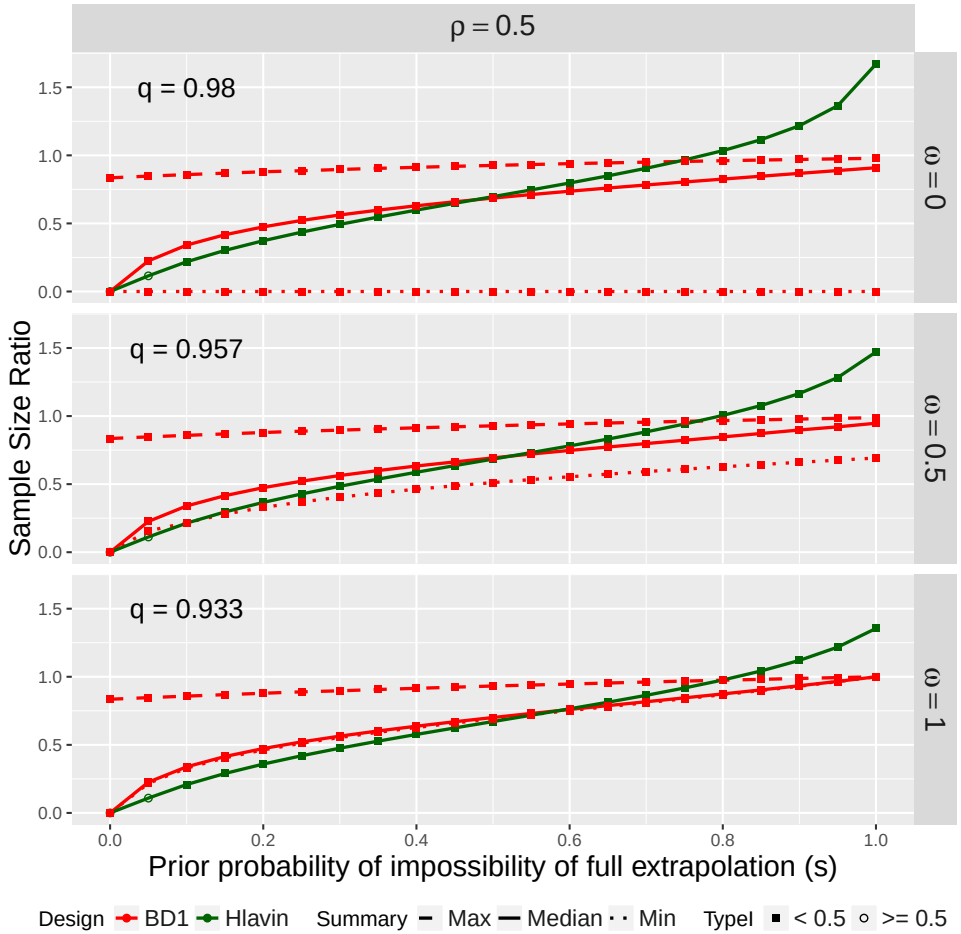
Results: paediatric sample size

For BD1 and BD2, present median sample size averaging over prior predictive distribution of $\hat{\theta}_A$.



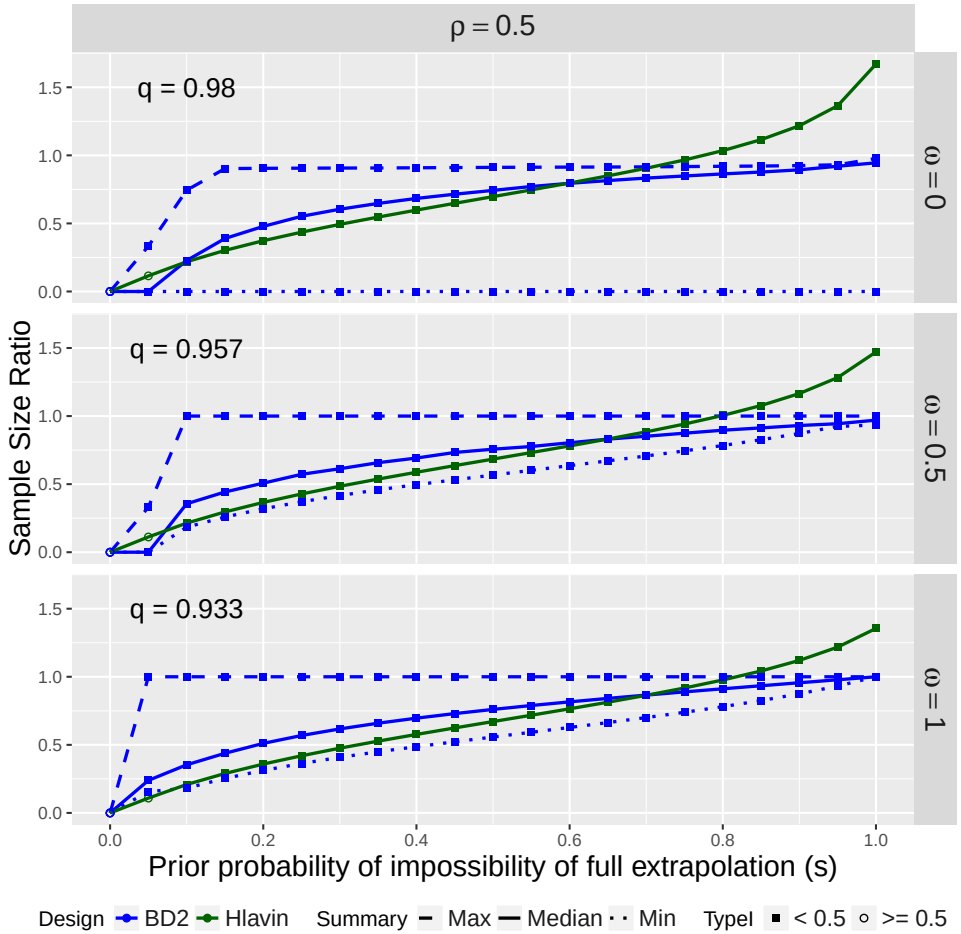
Results: paediatric sample size

Comparing BD1 and Hlavin designs (adding max and min):



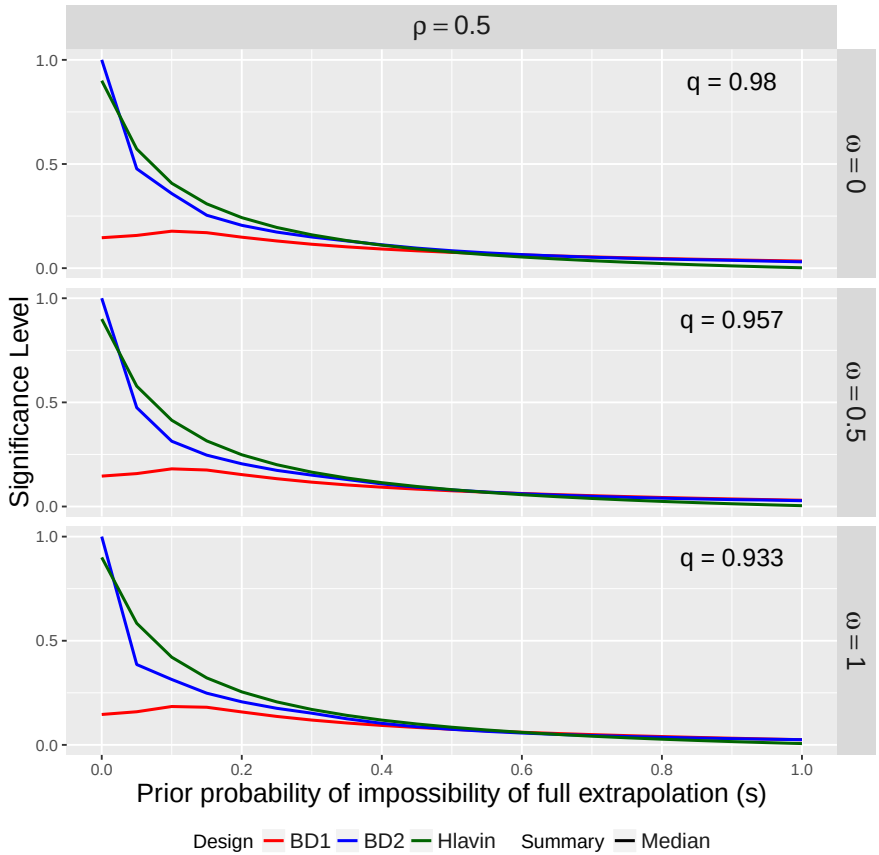
Results: paediatric sample size

Comparing BD2 and Hlavin designs (adding max and min):



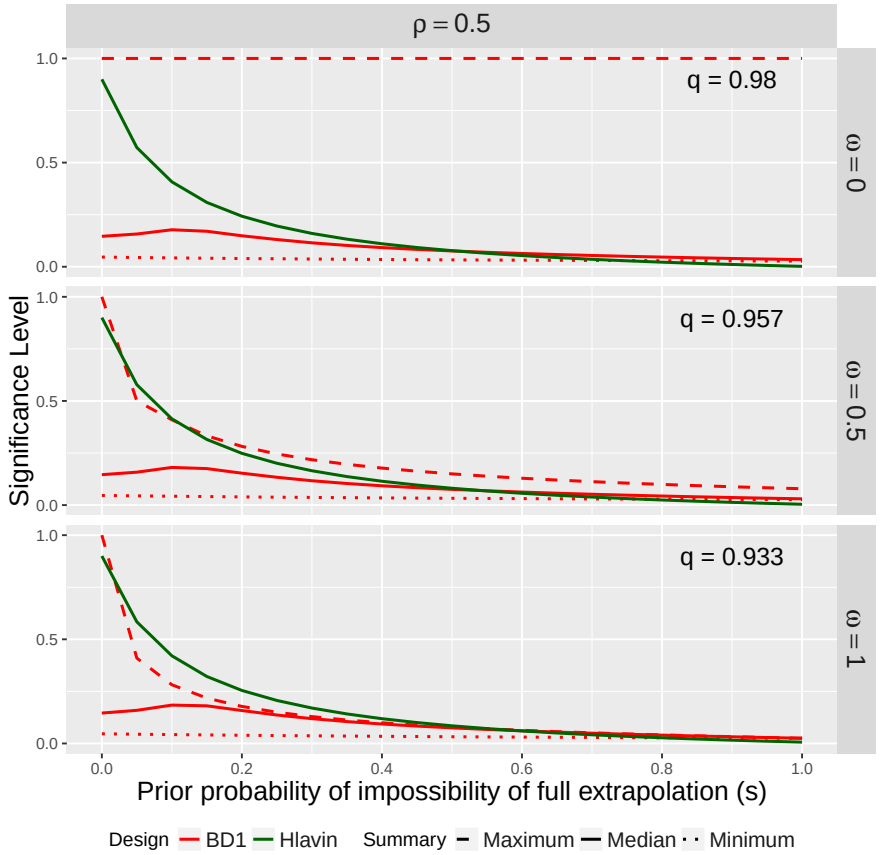
Results: type I error rate of paediatric trial

For BD1 and BD2, present median type I error rate averaging over prior predictive distribution of $\hat{\theta}_A$.



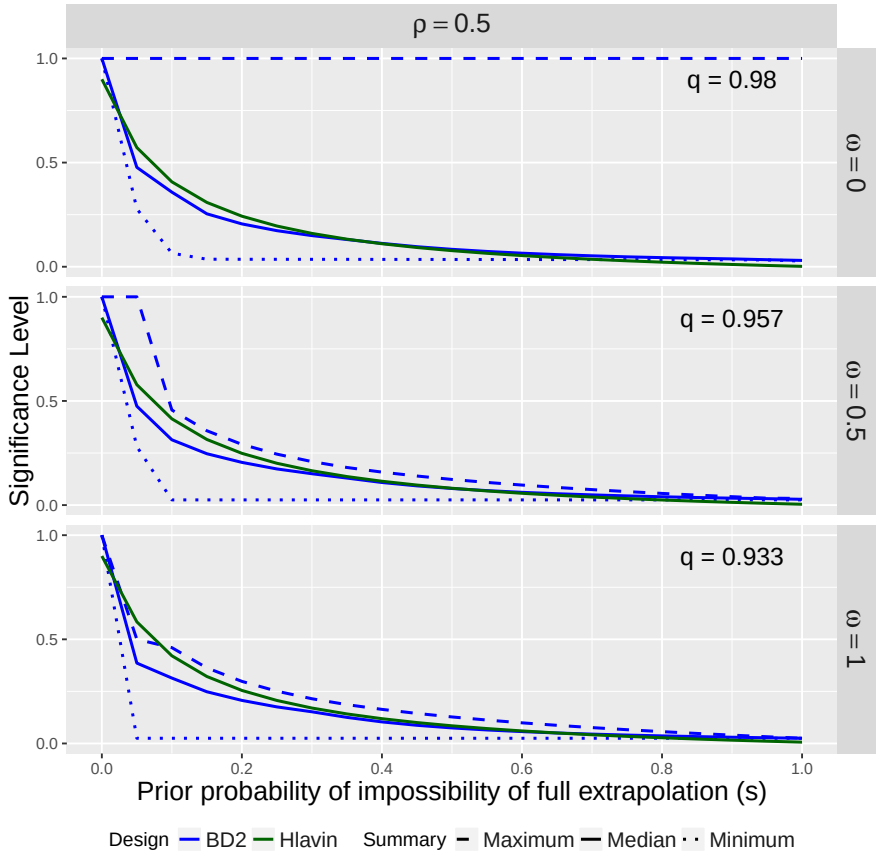
Results: type I error rate of paediatric trial

Comparing BD1 and Hlavin designs (adding max and min):



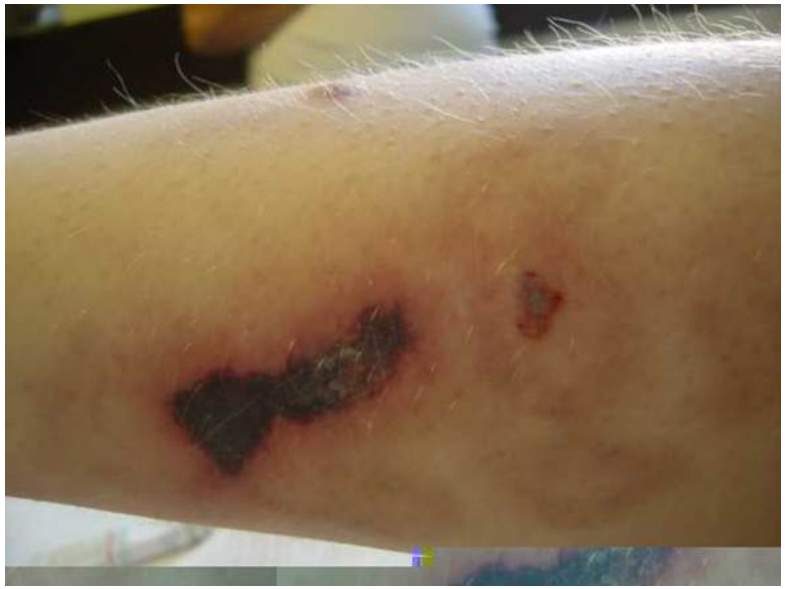
Results: type I error rate of paediatric trial

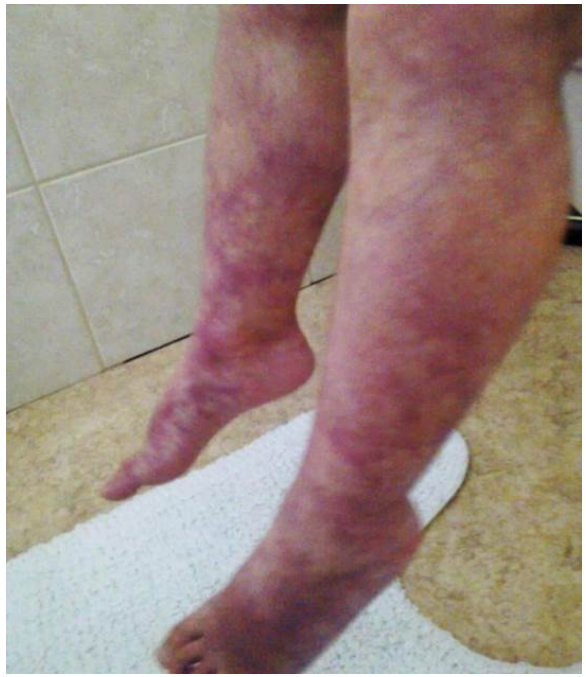
Comparing BD2 and Hlavin designs (adding max and min):



Example: The MYPAN trial

Childhood polyarteritis nodosa (PAN) is a serious inflammatory blood vessel disease which affects around 1 per million children.





MYPAN will compare Mycophenolate mofetil (MMF) versus cyclophosphamide (CYC) for the treatment of PAN in children.

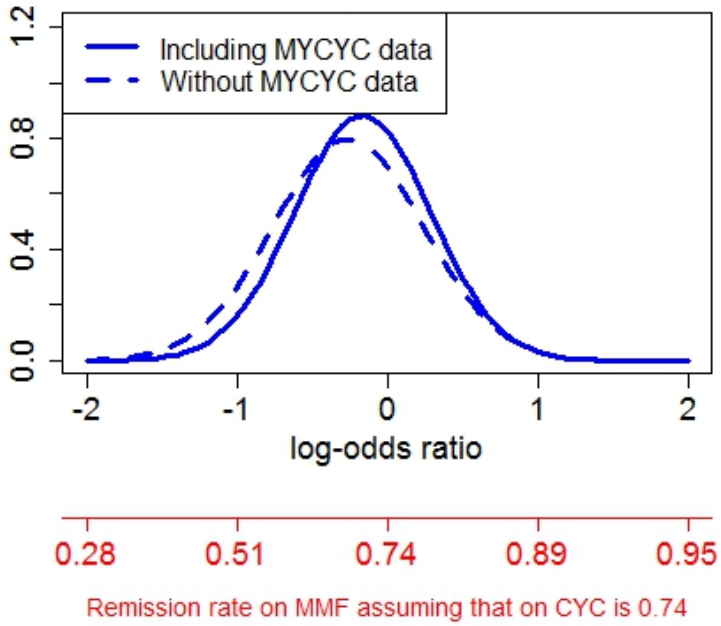
Primary endpoint is remission within 6-months. We measure the advantage of MMF over CYC in children using the log-odds ratio (θ_C).

Eliciting opinion for MYPAN trial

Hampson et al. (2014)

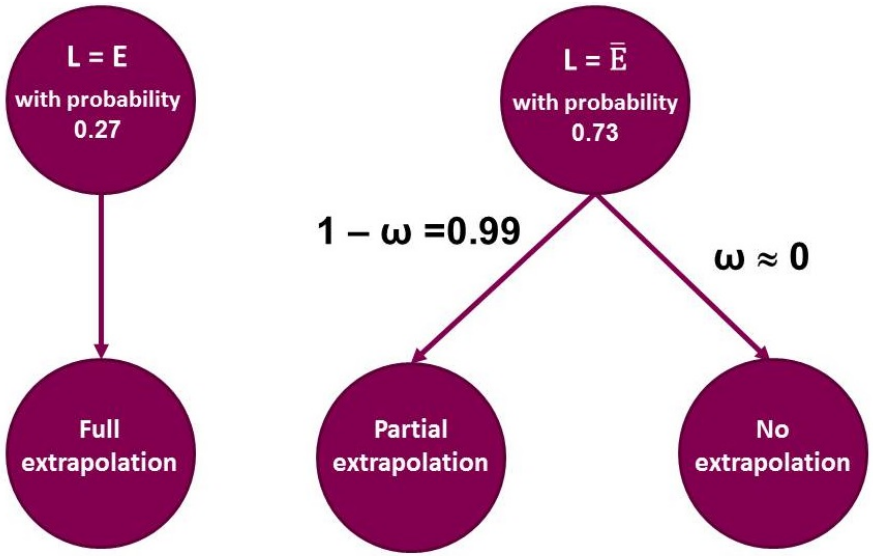
Prior opinion on MMF and CYC was recorded at a 2-day meeting:

- On Day 1, we elicited opinion on θ_C.
- On Day 2, we combined consensus opinion from Day 1 with related data.
- MYCYC trial involved 132 adults and 8 children with a condition related to PAN.
- MYCYC results: 52/70 remissions on CYC; 51/70 remissions on MMF.
- Let θ_A denote the log-odds ratio comparing MMF versus CYC in adults.



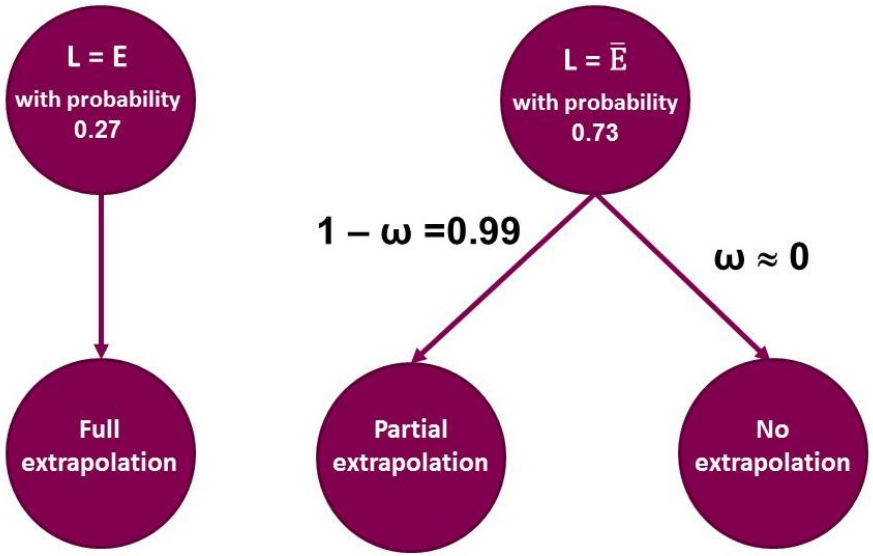
Bayesian mixture model

We calibrate the Bayesian mixture model to represent the experts' elicited opinions.
Shift treatment effects so we can consider MYPAN and MYCYC as superiority trials.



Bayesian mixture model

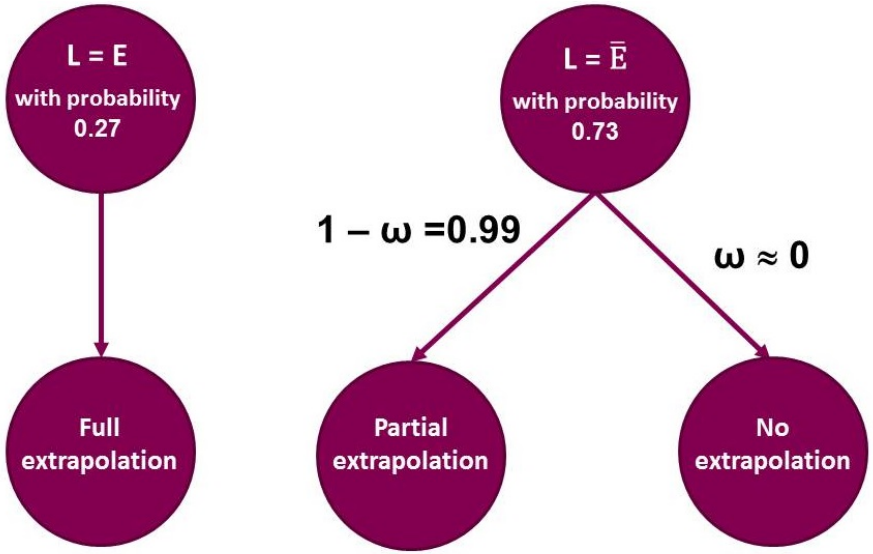
We calibrate the Bayesian mixture model to represent the experts' elicited opinions. Shift treatment effects so we can consider MYPAN and MYCYC as superiority trials.



Full extrapolation: Treatment effects are qualitatively similar: $\theta_C = 0.8\,\theta_A$, with $\theta_A \sim N(0.22,\, 0.57)$.

Bayesian mixture model

We calibrate the Bayesian mixture model to represent the experts' elicited opinions. Shift treatment effects so we can consider MYPAN and MYCYC as superiority trials.



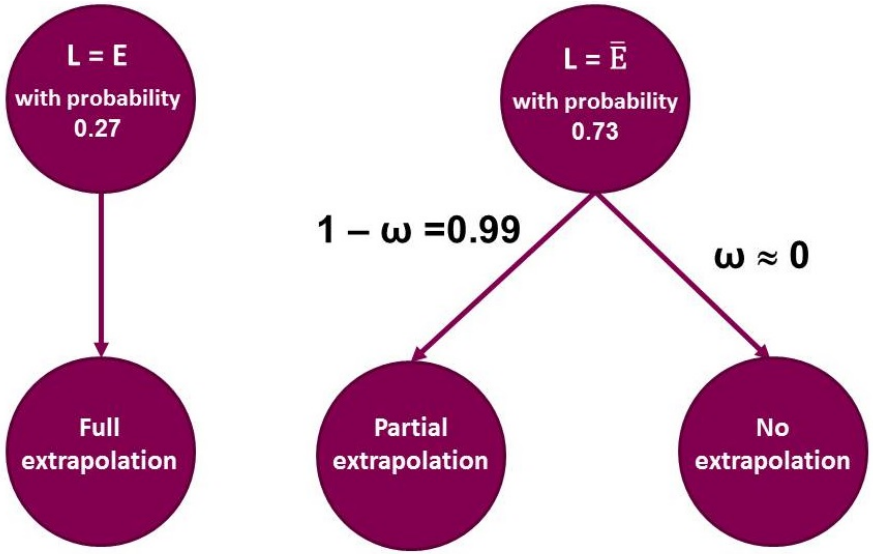
Partial extrapolation: Set

$$\begin{pmatrix}\theta_A\\ \theta_C\end{pmatrix} \sim N\left(\begin{pmatrix}0.22\\ 0.18\end{pmatrix}, \begin{pmatrix}0.57 & \sqrt{(0.57\times 0.25)\times\rho}\\ \sqrt{(0.57\times 0.25)\times\rho} & 0.25\end{pmatrix}\right)$$

where we set $\rho = 0.51$.

Bayesian mixture model

We calibrate the Bayesian mixture model to represent the experts' elicited opinions. Shift treatment effects so we can consider MYPAN and MYCYC as superiority trials.

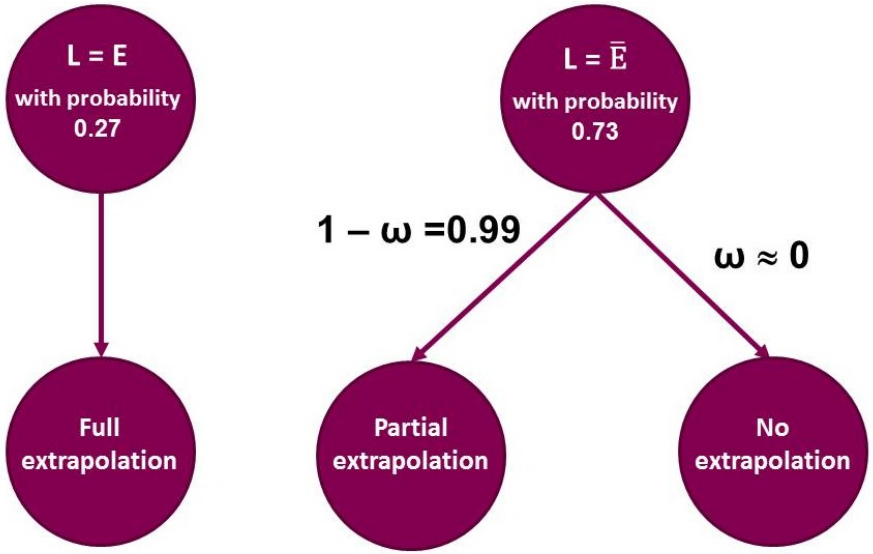


No extrapolation: Set

$$\begin{pmatrix}\theta_A \\ \theta_C\end{pmatrix} \sim N\left(\begin{pmatrix}0.22 \\ 0.18\end{pmatrix}, \begin{pmatrix}0.57 & 0 \\ 0 & 0.25\end{pmatrix}\right)$$

Bayesian mixture model

We calibrate the Bayesian mixture model to represent the experts' elicited opinions. Shift treatment effects so we can consider MYPAN and MYCYC as superiority trials.



No extrapolation: Set

$$\begin{pmatrix}\theta_A \\ \theta_C\end{pmatrix} \sim N\left(\begin{pmatrix}0.22 \\ 0.18\end{pmatrix}, \begin{pmatrix}0.57 & 0 \\ 0 & 0.25\end{pmatrix}\right)$$

Under these priors, $\text{pr}\{\theta_A > 0 \mid \text{Reject } H_{0A}\} = 1 - \gamma = 0.97$. For consistency with the continuous priors, we find Hlavin designs setting $1 - r = 0.384$ and $q = 0.84$.

MYPAN trial designs

Frame MYCYC as superiority trial with $\alpha = 0.15$ and power $1 - \beta = 0.85$ to detect a target effect of $\delta = 0.44$. In this case, we observe $\hat{\theta}_A = 0.37$ and $\mathcal{I}_A = 8.0$.

Set the target effect for MYPAN as $\delta = 0.44$.

Standalone paediatric trial:

- Sample size: 140
- Adjusted frequentist significance level: 0.15

Recommended design under Hlavin framework:

- Sample size: 212
- Adjusted frequentist significance level: 0.09

MYPAN trial designs

Frame MYCYC as superiority trial with $\alpha = 0.15$ and power $1 - \beta = 0.85$ to detect a target effect of $\delta = 0.44$. In this case, we observe $\hat{\theta}_A = 0.37$ and $\mathcal{I}_A = 8.0$.

Set the target effect for MYPAN as $\delta = 0.44$.

Standalone paediatric trial:

- Sample size: 140
- Adjusted frequentist significance level: 0.15

Recommended design under Hlavin framework:

- Sample size: 212
- Adjusted frequentist significance level: 0.09

Bayesian Design 1 (with average +ve predictive value = 0.97):

- Sample size: 172
- Adjusted frequentist significance level: 0.12

MYPAN trial designs

Frame MYCYC as superiority trial with $\alpha = 0.15$ and power $1 - \beta = 0.85$ to detect a target effect of $\delta = 0.44$. In this case, we observe $\hat{\theta}_A = 0.37$ and $\mathcal{I}_A = 8.0$.

Set the target effect for MYPAN as $\delta = 0.44$.

Standalone paediatric trial:

- Sample size: 140
- Adjusted frequentist significance level: 0.15

Recommended design under Hlavin framework:

- Sample size: 212
- Adjusted frequentist significance level: 0.09

Bayesian Design 1 (with average +ve predictive value = 0.97):

- Sample size: 172
- Adjusted frequentist significance level: 0.12

Bayesian Design 2 (maintain minimum +ve predictive value = 0.86):

- Sample size: 106
- Adjusted frequentist significance level: 0.19

Conclusions

We have proposed a number of Bayesian designs for paediatric efficacy trials which can offer sample size savings without lowering the evidence threshold:

- Comparing BD1 and BD2 with Hlavin designs, we see that the sample size required by the fully Bayesian designs is less sensitive to the choice of scepticism factor.
- Further work will focus on developing a framework for eliciting prior distributions for the Bayesian model parameters.
- Bayesian extrapolation model can be extended in several ways (set priors on b and ρ).

Bibliography

Bauer P, Koenig F. Adaptive paediatric investigation plans, a small step to improve regulatory decision making in drug development for children? *Pharmaceutical Statistics* 2016; **15**:384.

Dunne J, Rodriguez WJ, Murphy MD. Extrapolation of adult data and other data in pediatric drug-development programs. *Pediatrics* 2011; **128**:e1242.

European Medicines Agency. Concept paper on extrapolation of efficacy and safety in medicine development. Final 2012.

Hampson LV, Whitehead J, Eleftheriou D, Brogan P. Bayesian methods for the design and interpretation of clinical trials in very rare diseases. *Statistics in Medicine* 2014; **33**:4186

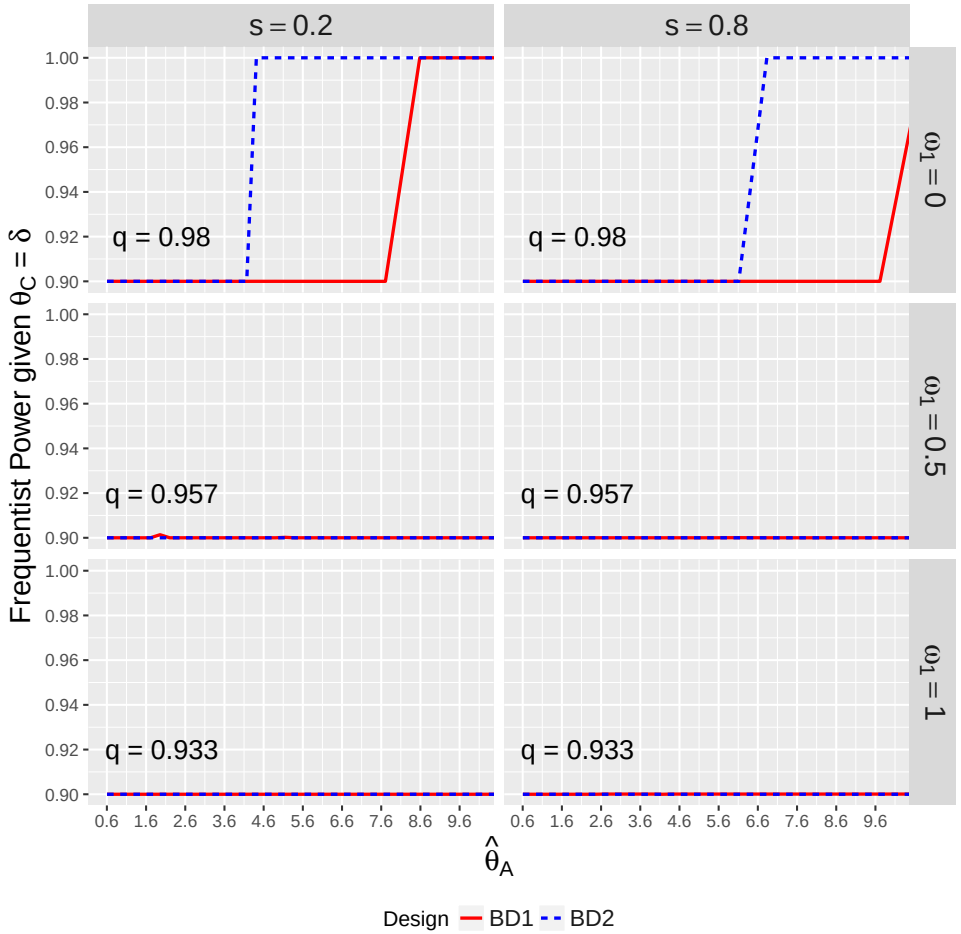
Hlavin G, Koenig F, Male C, Posch M, Bauer P. Evidence, eminence and extrapolation. *Statistics in Medicine* 2016; **35**:2117. See: `http://www.ideal.rwth-aachen.de/wp-content/uploads/2016/10/2016_10_20_Webinar_IDEAL_Koenig.pdf`

Simon R. Comment on 'Clinical trials and sample size considerations: another perspective'. *Statistical Science* 2000; **15**:95

Whitehead J, Valdes-Marquez E, Johnson P, Graham G. Bayesian sample size for exploratory clinical trials incorporating historical data. *Statistics in Medicine* 2008; **27**:2307

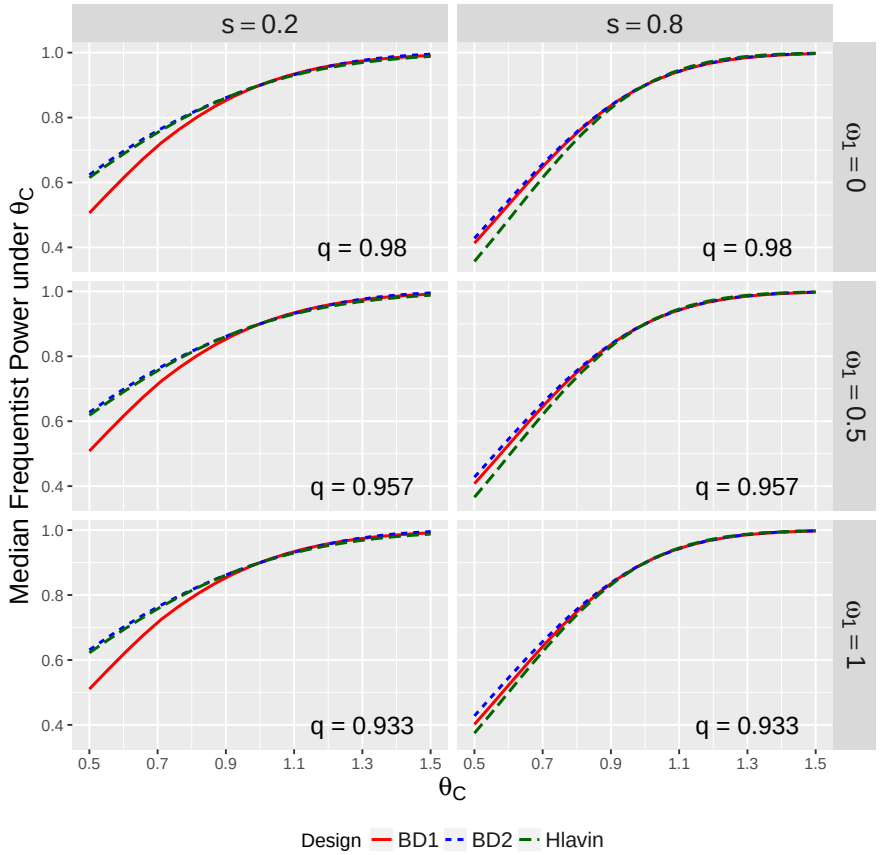
Results: frequentist power to detect target difference

Compare the power of Bayesian designs as a function of $\hat{\theta}_A$ when $\delta = 1$, $\rho = 0.5$.



Results: frequentist power curves

Compare the frequentist power curves of Bayesian and Hlavin designs when $\rho = 0.5$.



For each θ_C, average power over prior predictive distribution of $\hat{\theta}_A$ given θ_C.