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

Lisa Hampson and Franz Koenig

Department of Mathematics & Statistics, Lancaster University
Statistical Innovation Group, Advanced Analytics Centre, AstraZeneca

2017 PSI Conference
17th May, 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 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 Martin Posch, 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:

- Drug development in adults and children is typically staggered.
- 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:

- Drug development in adults and children is typically staggered.
- 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?
Extrapolations from adult efficacy data can range from ‘complete’ to ‘partial’.

Reasonable to assume that children and adults have
- similar disease progression?
- similar response to intervention?

No to either
- Is there a PD measurement that can be used to predict efficacy in children?
  - No
  - Conduct PK studies to establish dosing, then safety and efficacy trials in children (A)
  - Yes
  - Conduct PK/PD studies to establish ER in children for PD measurement
    Conduct PK studies to achieve target concentrations based on ER
    Conduct safety studies at proper dose (B)

Yes to both
- Reasonable to assume similar exposure-response (ER) in children and adults?
  - No
  - Conduct PK studies in children to achieve drug levels similar to adults
  - Conduct safety trials at proper dose (C)
  - Yes
Notation and assumptions

We will use the following notation:

- \( \theta_A \) measures the advantage of a new therapy relative to placebo in adults.
- \( \theta_C \) 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.

\[ P(\theta_A > 0) = 1 - r \]

Power: $1 - \beta$
Type I error rate: $\alpha$
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.
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 $H_{0A}$. 

Prior: $P(\theta_A > 0) = 1 - r$

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

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

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

One can interpret the extrapolation model as follows:

- Full extrapolation implies effects in different age groups are qualitatively similar.
- $s$ reflects our prior *scepticism* about the plausibility of extrapolation.
- Given the impossibility of extrapolation, may still believe there is a chance the treatment is efficacious in children.
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: $\Pr\{\theta_C > 0 \mid \text{Reject } H_{0A}\} = (1 - s) \Pr\{\theta_A > 0 \mid \text{Reject } H_{0A}\} + s(1 - q)$.
Extrapolation model implies we design the paediatric test of $H_{0C} : \theta_C \leq 0$ as follows:

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

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 implies we design the paediatric test of $H_0^C : \theta_C \leq 0$ as follows:

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

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 $H_0^C$ . . .
Extrapolation model implies we design the paediatric test of $H_{0C} : \theta_C \leq 0$ as follows:

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

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

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

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.

$\alpha_{\text{adj}} > \alpha$ leads to a reduction in the paediatric sample size.
Hlavin method can be thought of as a hybrid Bayesian-frequentist approach:

- Paediatric trial is analysed using conventional, frequentist, methods . . .
- . . . but $\alpha$ 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 \textit{hybrid Bayesian-frequentist} approach:

- Paediatric trial is analysed using conventional, frequentist, methods . . .
- . . . but $\alpha$ is adjusted with a \textit{Bayesian interpretation of the data in mind}.
- Working with two point priors implies $\alpha_{\text{adj}}$ 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. \textit{This would lead to an adaptive PIP.}

How would this method compare with taking a fully Bayesian approach?
Below is a Bayesian mixture model representing prior opinion on $\theta_A$ and $\theta_C$.

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

\begin{align*}
L &= E \\text{w.p.} \ (1 - s) \\
1 - \omega &\quad \omega \\
\text{Full extrapolation} &\quad \text{Partial extrapolation} &\quad \text{No extrapolation}
\end{align*}
Below is a Bayesian mixture model representing prior opinion on $\theta_A$ and $\theta_C$.

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

\[
\begin{pmatrix}
\theta_A \\
\theta_C
\end{pmatrix} \sim N\left( \begin{pmatrix}
\mu_A \\
\mu_C, 1
\end{pmatrix}, \begin{pmatrix}
\sigma^2_A \\
\rho \sigma_A \sigma_C, 1 \\
\rho \sigma_A \sigma_C, 1 \\
\sigma^2_C, 1
\end{pmatrix}\right)
\]
Bayesian extrapolation model

Below is a Bayesian mixture model representing prior opinion on $\theta_A$ and $\theta_C$.

No extrapolation: Differences between age groups are such that knowing $\theta_A$ would tell us nothing about a medicine’s effect in children:

$$
\begin{pmatrix}
\theta_A \\
\theta_C
\end{pmatrix} \sim N
\begin{pmatrix}
(\mu_A, \sigma_A^2) \\
(\mu_C, 0)
\end{pmatrix}
\begin{pmatrix}
\sigma_A^2 & 0 \\
0 & \sigma_C^2
\end{pmatrix}
$$
Bayesian extrapolation model

- Priors (for $\theta_A$, $\theta_C$ 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.
Set **pessimistic priors** with mean $-\delta/2$ and variance chosen so that prior probability of 0.995 that standardised effect size in adults is less than 5.

**Full extrapolation**: Treatment effects are qualitatively similar: $\theta_C = 0.8 \theta_A$. 
Set **pessimistic priors** with mean $-\delta/2$ and variance chosen so that prior probability of 0.995 that standardised effect size in adults is less than 5.

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

Set **pessimistic priors** with mean $-\delta/2$ and variance chosen so that prior probability of 0.995 that standardised effect size in adults is less than 5.

\[
L = E_\text{w.p.} \ (1 - s) \\
L = \bar{E}_\text{w.p.} \ s
\]

**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 \(1 - r = 0.22\).
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 $\hat{\theta}_C \geq c^*$) to ensure:

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

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 $\theta_A \sim N(-0.5, 3.8)$. 

Hampson and Koenig
Extrapolation to support the development of paediatric medicines
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 $\hat{\theta}_C \geq c^*$) to ensure:

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

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 $c^* \geq 0$. 
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 $\hat{\theta}_C \geq c^*$) to ensure:

- **Frequentist Power**: $\text{pr}\{\hat{\theta}_C \geq c^* \mid \theta_C = \delta\} \geq 1 - \beta$;
- **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 $\alpha = 0.025$, $1 - \beta = 0.9$, $\delta = 1$, $\theta_A \sim N(-0.5, 3.8)$. 
Bayesian designs depend on the estimate $\hat{\theta}_A$ generated by the adult trial.

At the time the PIP is written, we can calculate the prior predictive distribution of the trial's operating characteristics and report summaries of this (median, max, min).
Results: paediatric sample size

Sample size expressed as ratio of number needed for Neyman-Pearson test of $H_0$ with $\alpha = 0.025$, $1 - \beta = 0.9$, $\delta = 1$.

For $\rho = 0.5$:
- $q = 0.98$
- $q = 0.957$
- $q = 0.933$

Prior probability of impossibility of full extrapolation ($s$)

Sample Size Ratio

Design BD1 BD2 Hlavin TypeI < 0.5 >= 0.5 Summary Median

Extrapolation to support the development of paediatric medicines
Results: paediatric sample size

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

Quantiles calculated with respect to prior predictive distribution of $\hat{\theta}_A$. 

Hampson and Koenig
Extrapolation to support the development of paediatric medicines
Results: paediatric sample size

Sample size expressed as ratio of number needed for Neyman-Pearson test of $H_0\ A$ with $\alpha = 0.025$, $1 - \beta = 0.9$, $\delta = 1$.

Quantiles calculated with respect to prior predictive distribution of $\hat{\theta}_A$. 

Hampson and Koenig
Extrapolation to support the development of paediatric medicines
Results: type I error rate of paediatric trial
Results: type I error rate of paediatric trial

Quantiles calculated with respect to prior predictive distribution of $\hat{\theta}_A$ given $\theta_C = 0$. 

Hampson and Koenig
Extrapolation to support the development of paediatric medicines
Results: type I error rate of paediatric trial

Quantiles calculated with respect to prior predictive distribution of $\hat{\theta}_A$ given $\theta_C = 0$. 

Hampson and Koenig
Extrapolation to support the development of paediatric medicines
Results: frequentist power to detect target difference

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

Hampson and Koenig
Extrapolation to support the development of paediatric medicines
Results: frequentist power curves

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

For each $\theta_C$, summaries calculated with respect to the prior predictive distribution of $\hat{\theta}_A$ given $\theta_C$. 
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 $\rho$).


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