A Bayesian framework for extrapolation using mixture priors: challenges and some recommendations

Clara Dominguez Islas[a], Adrian Mander[a]
Nicky Best[b], Rebecca Turner[c]

[a] MRC Biostatistics Unit, University of Cambridge
[b] GlaxoSmithKline, [c] University College London, Clinical Trials Unit

PSI One Day Meeting: Extrapolation, 22 November 2017

Acknowledgments

GlaxoSmithKline
Naomi Givens
Rimgaile Urbaityte
Sophie Barthel

UCL, Clinical Trials Unit
Ian White

MRC Biostatistics Unit
Sylvia Richardson

Introduction

Paediatric populations present several challenges and opportunities for clinical trial design and analysis:

- Practical and ethical constraints on sample size imply that fully powered efficacy studies are not feasible or justified
- In contrast to other situations of small populations, new treatments and drugs are first tested and licensed for adults.

"It is acknowledged that development of a medicine in adults provides a rich source of data to inform paediatric development and given **reasonable similarity** between adults and children, **extrapolation from adults may reduce paediatric data requirements** to make conclusions for use of the medicine in children".

Extrapolation (working definition)

Extending information and conclusions available from studies in one or more subgroups of the patient population (source population(s)), ... to make inferences for another subgroup of the population (target population), ... thus reducing the need to generate additional information to reach conclusions for the target population.

Motivating example

Pediatric Investigation Plan

- Phase III confirmatory trials in adults to be completed soon.
- PK/PD and other data not sufficient for full extrapolation of efficacy to children (target population).
- Sample size limited by challenging recruitment (small target population).

Possible approaches when sample size is limited (as suggested by EMA):

- Using a larger level for the type 1 error rate than the usual 5%.
- Widening a usually accepted non-inferiority margin.
- Using **Bayesian methods** to either summarise the prior information for the extrapolation concept, or **to explicitly borrow information** (from adult trials, from control groups, from other paediatric clinical trials).

Non inferiority RCT: New treatment vs standard of care

Sample size: 60 patients (randomization ratio 2:1).
Outcome and efficacy endpoint: Binary, difference in proportions.
Regulator's proposal for relaxing data requirements: Extended NI Margin (-20%), relative to NI used for adults (-10%).

Motivating Example

The opportunity :

Extrapolation from adult to pediatric population

Efficacy of the same treatment is being tested in adults (two fully powered trials, N $\approx$ 500). Believed to be very similar for children.

Extrapolation from historical to current controls

The active control treatment (currently recommended therapy) has been evaluated in previous pediatric studies (phase II and phase III RCTs).

Aim: To make use of historical information. Bayesian inference provides a natural framework

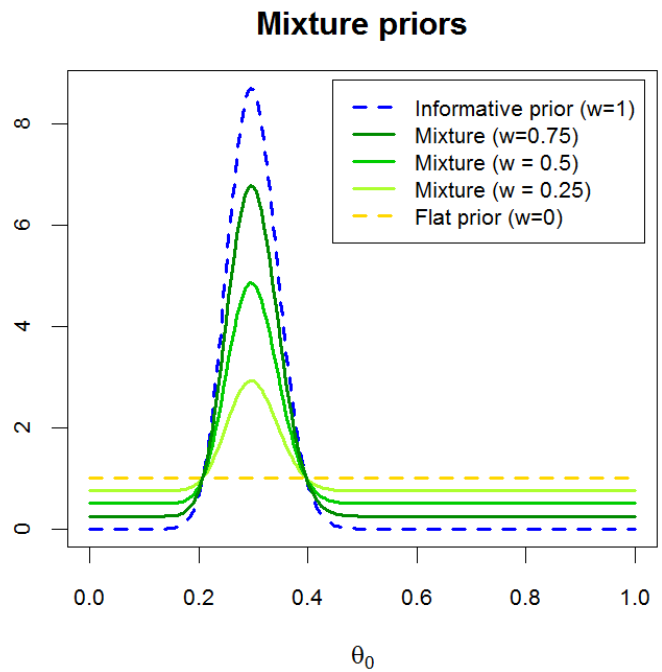
Problem: Although we believe on the similarity between populations, we acknowledge that they are not the same. Important differences, not known a priori, might arise.

Challenge: Informative priors, but with some level of skepticism, to allow for potential conflict.

Mixture priors

Mixture of one **informative component** π_I (from historical sources, for example) with a **less precise (more vague) component** π_V.

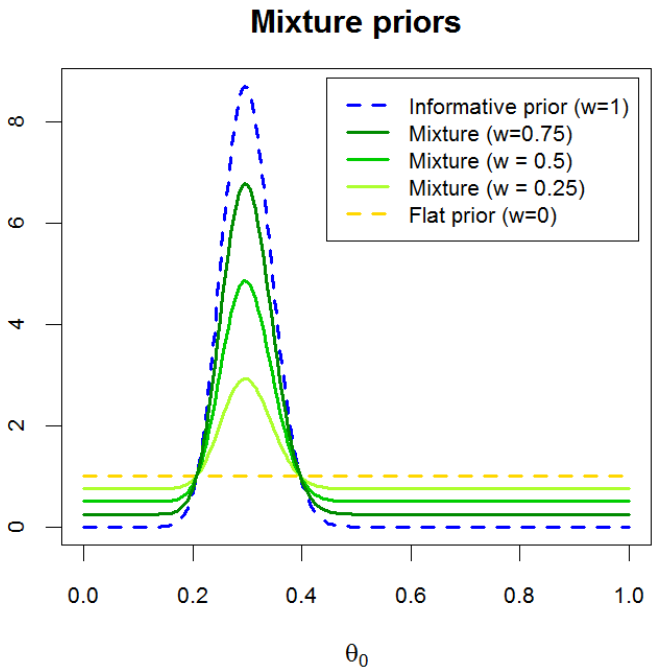
$$\pi(\theta) = w\pi_I(\theta) + (1 - w)\pi_V(\theta)$$

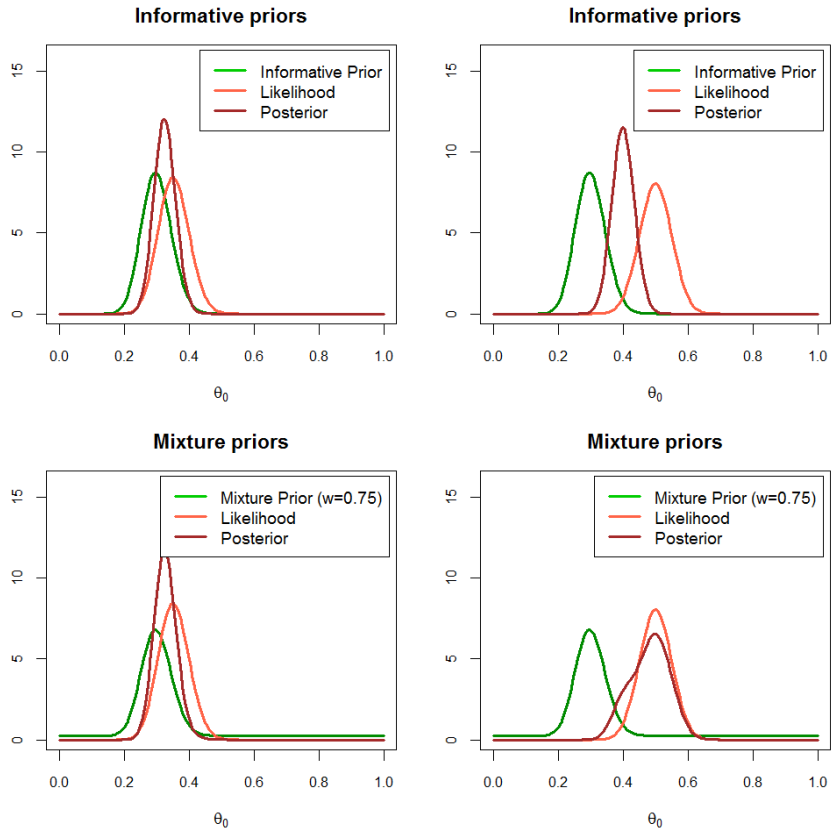


Mixture priors

Mixture of one **informative component** π_I (from historical sources, for example) with a **less precise (more vague) component** π_V.

$$\pi(\theta) = w\pi_I(\theta) + (1 - w)\pi_V(\theta)$$





Mixture priors in the literature

- Possibly the first documented example of inference using a mixture prior:

  Haldane (1932) — A note on inverse probability. *Mathematical Proceedings of the Cambridge Philosophical Society*

- Mixture priors have been proposed in the context bridging studies,

  Hsiao et al. (2007) — Use of prior information for Bayesian evaluation of bridging studies *J Biopharm Stat*

  Gandhi et al. (2012) — A Bayesian approach for inference from a bridging study with binary outcomes *J Biopharm Stat*

- to incorporate information on historical controls,

  Schmidli et al. (2014) — Robust Meta-Analytic-Predictive Priors in Clinical Trials with Historical Control Information, *Biometrics*

- and considered as an option for adults-to-paediatrics extrapolation.

  Wadsworth et al. (2016) — Extrapolation of efficacy and other data to support the development of new medicines for children: A systematic review of methods. *Stat Methods Med Res.*

  Gamalo-Siebers et al. (2017) — Statistical modeling for Bayesian extrapolation of adult clinical trial information in pediatric drug evaluation. *Pharm Stat*

Further research still needed

To reach a better understanding, address some unresolved issues and provide better guidance.

Mixture priors for extrapolation

Objective

Discuss some of the issues and challenges that we have encountered when using mixture priors for extrapolation (using an simplified example).

Outline

- Model and interpretation.
- Understanding the inference.
- Building the mixture prior (choosing its components):
  - the informative prior component,
  - the vague prior component,
  - the weights.
- Evaluating the frequentist operating characteristics.
- Extending from univariate to bivariate mixtures (motivating example).

Mixture priors: model and interpretation

Full Model:

$$
\begin{aligned}
\zeta &\sim \text{Bernoulli}(w) \\
\pi(\theta|\zeta) &= \begin{cases} \pi_I(\theta) \text{ if } \zeta = 1 \\ \pi_V(\theta) \text{ if } \zeta = 0 \end{cases} && \text{Prior} \\
y &\sim f_\theta && \text{Likelihood}
\end{aligned}
$$

In our prior: $P(\zeta = 1) = w$

The prior probability that information in π_I is relevant to the target population

Posterior distribution (also a mixture!):

$$
\begin{aligned}
\pi(\theta|w, y) &= P(\zeta = 1|y)\pi_I(\theta|y) + P(\zeta = 0|y)\pi_V(\theta|y) \\
P(\zeta = 1|y) &= \frac{w f_I(y)}{w f_I(y) + (1 - w) f_V(y),}
\end{aligned}
$$

where $f_I(y)$ and $f_V(y)$ denote the marginal (prior-predictive) distributions.

In our posterior: $P(\zeta = 1|y) = w^*(y)$

The updated (posterior) probability of relevance, given the observed outcome.

Mixture priors: inference

Simplified example using conjugate priors[1]

Binary outcome of interest (θ = probability of success)

- **Mixture prior**
  - $\pi_I(\theta)$ **= Beta(210,90)** (Mean = 0.7, effective sample size = 300)
  - $\pi_V(\theta)$ **= Beta(1,1)** (Mean = 0.5, effective sample size = 2)

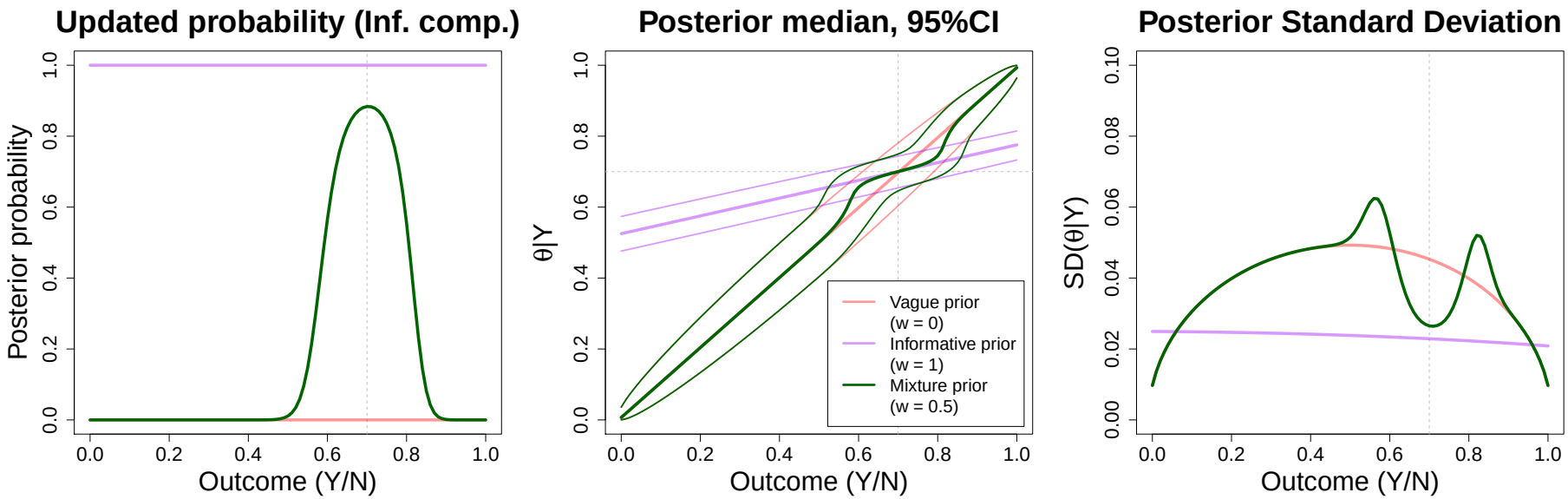
- **Likelihood**
  - Number of successes, **Y** $\sim$ **Binomial(100,** θ**)** (fixed sample size)

- **Posterior, mixture of two beta distributions**
  - $\pi_I(\theta|y)$ **= Beta(210 + y, 90 + 100 - y)**
  - $\pi_V(\theta|y)$ **= Beta(1 + y, 1 + 100 - y)**
  - $w^*(y) = \frac{wf_I(y)}{wf_I(y)+(1-w)f_V(y)}$, with f_I and f_V **beta-binomial**

[1] Not necessary in practice if posteriors and marginals can be estimated

Mixture priors: inference

Mixture prior with informative component centred around (**0.7**) and prior probability **w = 0.5**.

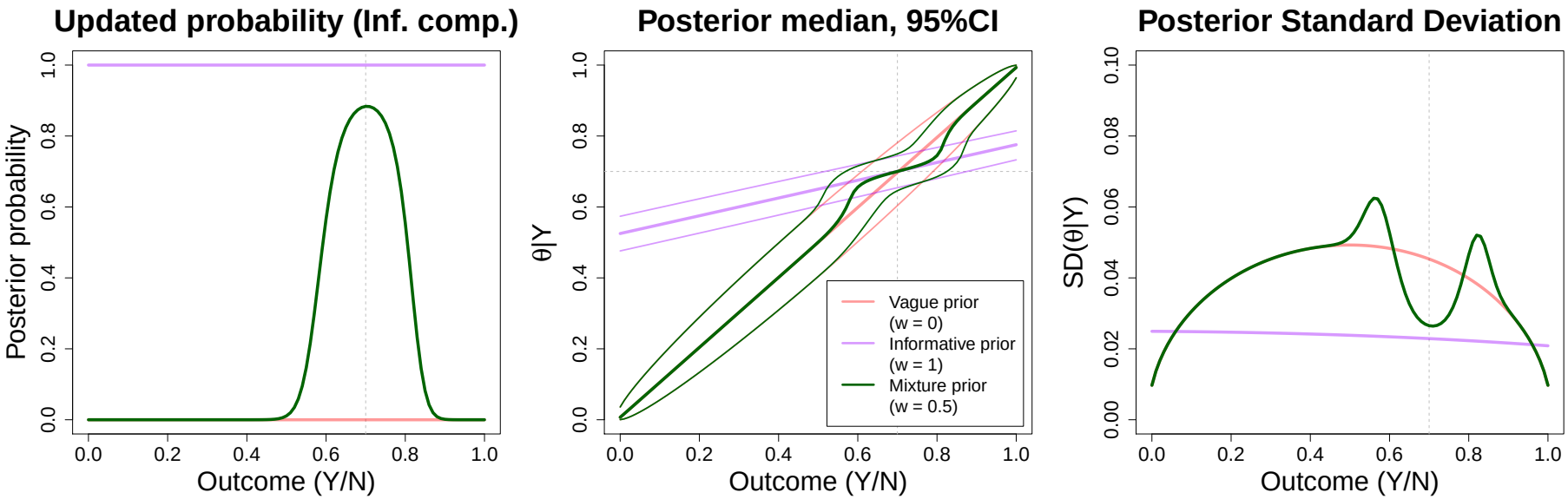


When the outcome is **in agreement** ($Y/N = 0.7$) with informative prior:

- The posterior probability for the informative component is high.
- The posterior is more similar to posterior from informative component.
- The posterior is more precise (SD smaller than with vague prior).

Mixture priors: inference

Mixture prior with informative component centred around (**0.7**) and prior probability **w = 0.5**.

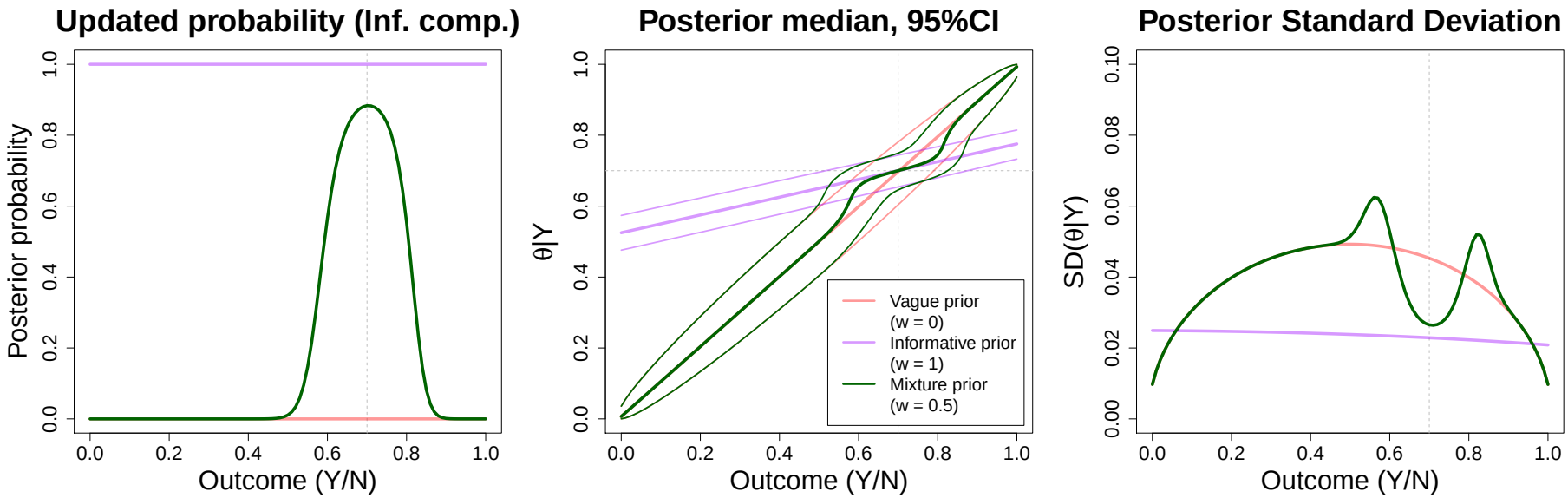


When the outcome is **in strong disagreement** ($^{Y}/_{N} \ll 0.7$ or $^{Y}/_{N} \gg 0.7$) with informative prior:

- The posterior probability for the informative component is close to 0.
- The posterior is more similar to posterior from vague component.
- The posterior SD is similar to that using a vague prior.

Mixture priors: inference

Mixture prior with informative component centred around (**0.7**) and prior probability **w = 0.5**.

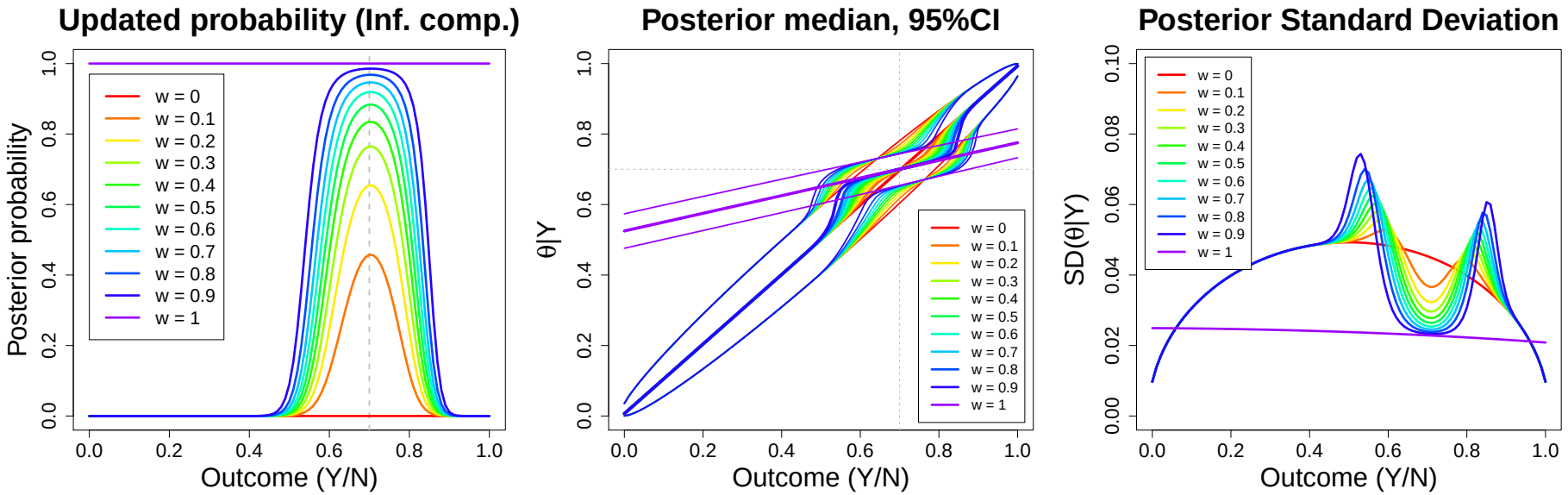


When the outcome is **in moderate conflict** with informative prior:

- The posterior probability of for the informative component is somewhere between 0 and 1.
- The posterior is an average of separate posteriors.
- The variance can be inflated (greater than what would be obtained when using vague prior).

Mixture priors: inference

Varying the prior probability of relevance:

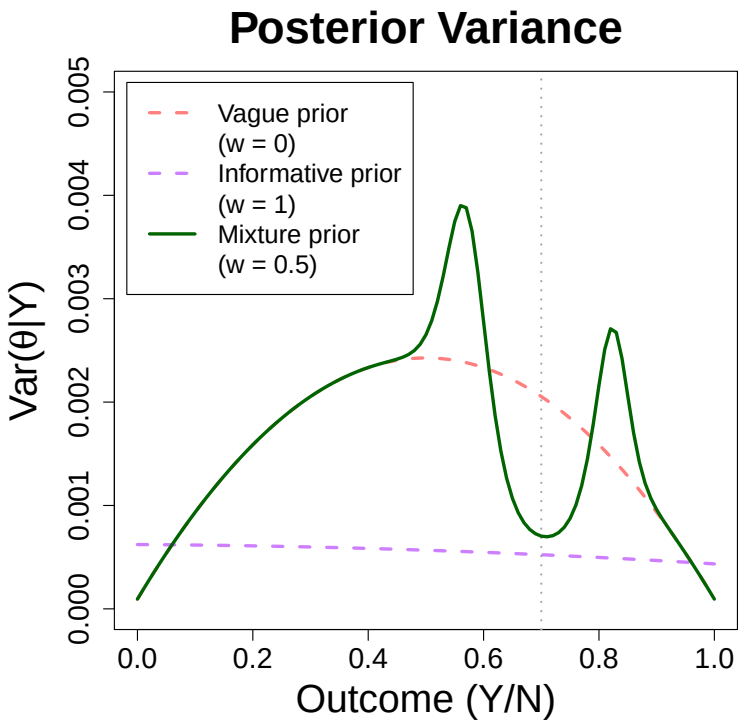


The value given to the prior probability for the informative component will determine:

- An upper bound on the updated probability and how it changes depending on level of disagreement.
- How much the posterior mixture will rely on the informative component.
- The maximum gain in precision (when data is in agreement) and how much the variance will be inflated (when there is moderate disagreement).

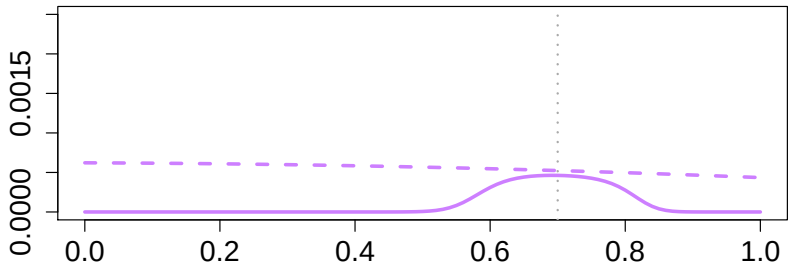
Explaining the variance inflation

The posterior variance can be decomposed in three terms*:

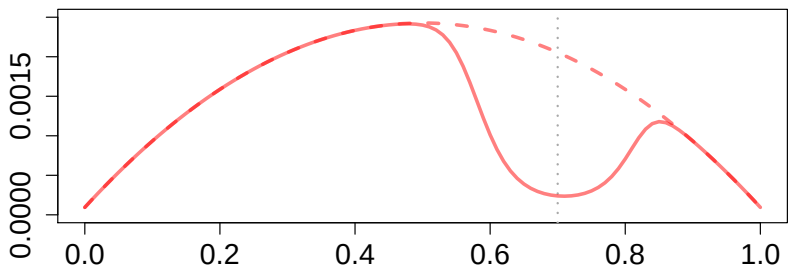


* true for any two-component mixture

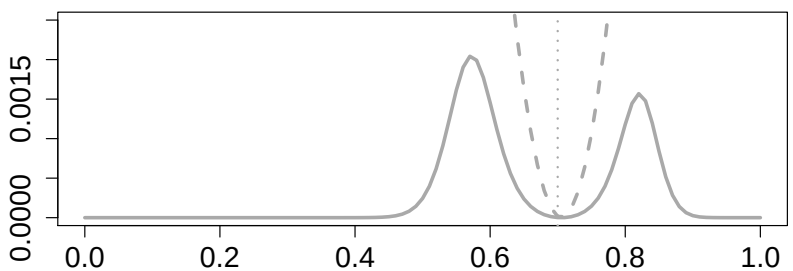
$w^*(y)\mathrm{Var}_I(\theta|y)$



$[1 - w^*(y)]\mathrm{Var}_V(\theta|y)$



$w^*(y)[1 - w^*(y)]\,[\mathrm{E}_I(\theta|y) - \mathrm{E}_V(\theta|y)]^2$



Building a MP: the informative component

Purposes:

1. Summarise the information on the quantity of interest from the source population.
2. Reflect the relationship between studies (or sources of information), within and between populations:
   - Equal
     - Data from a single source (one big trial).
     - Different studies, same population and protocol (FE meta-analysis).
   - Exchangeable
     - Different studies with different designs (RE meta-analysis).[2]
   - Exchangeable within exchangeable populations
     - Different studies from different populations (three-level hierarchical model)[3]

---

[2] Schmidli et al. (2007) - Robust Meta-Analytic-Predictive Priors in Clinical trials with Historical Control Information, *Biometrics*.

[3] U.S. Department of Health and Human Services, Food and Drug Administration (2016), Leveraging Existing Clinical Data for Extrapolation to Pediatric Uses of Medical Devices.

Building a MP: the informative component

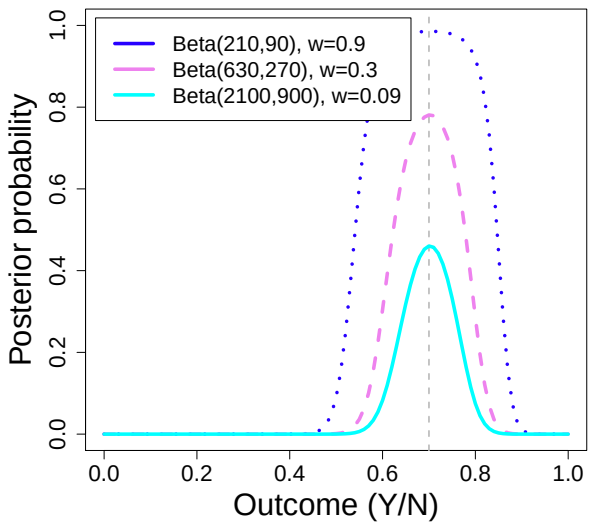
The prior probability of the informative component is not equivalent to the proportion of information that will be borrowed

More precise prior with low probability $\neq$ **Less precise prior with high probability**

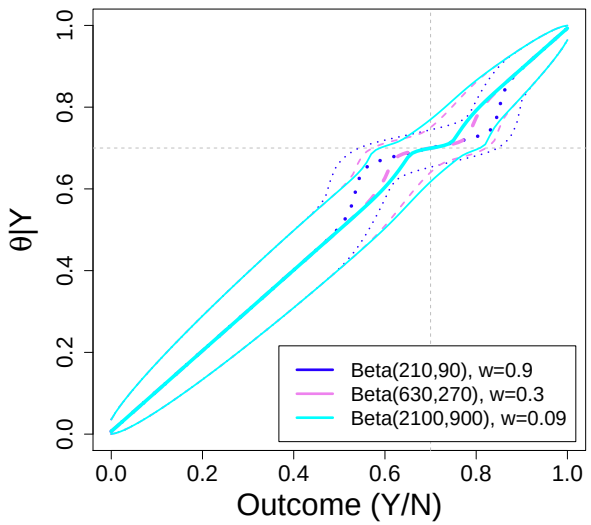
For example, compare the following mixture priors:

Informative Component	ESS(I)	Prior probability (w)	Vague Component
Beta(210,90)	300	0.9	
Beta(630,270)	900	0.3	Beta(1,1)
Beta(2100,900)	3000	0.09	

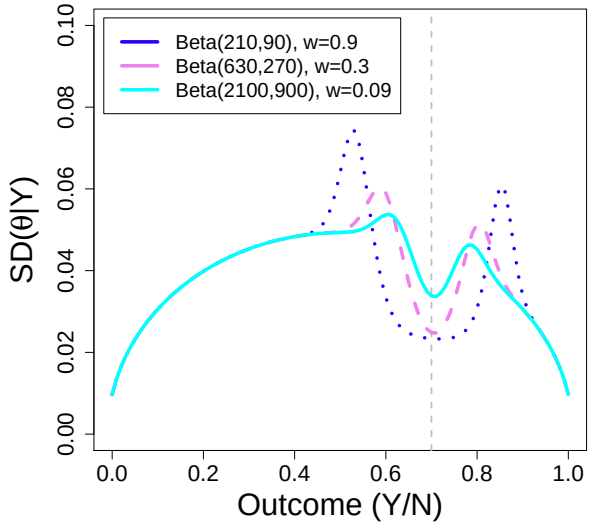
Updated probability (Inf. comp.)



Posterior median, 95%CI



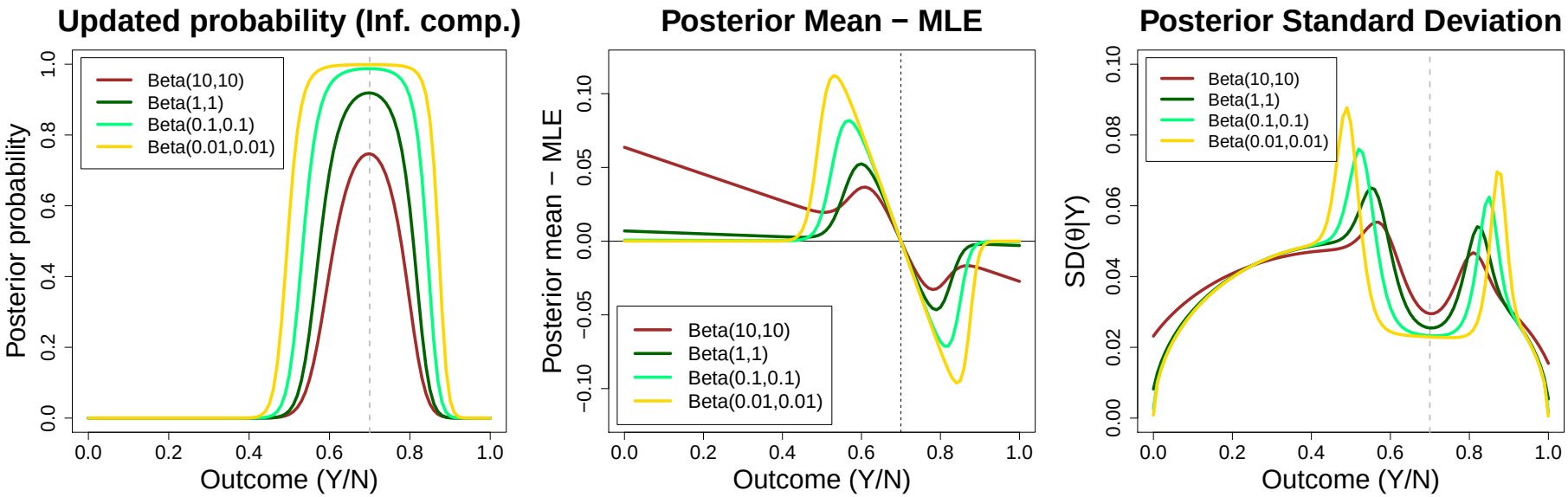
Posterior Standard Deviation



Building a MP: the vague component

How vague is vague? The posterior can be very sensitive to the choice of this component

Comparing mixture priors with same informative component (same probability) but different choices of vague component:



The choice of vague component will influence

- the upper bound of the posterior probability,
- the extent of the “agreement” zone and the influence of the informative component,
- the residual “bias” induced when the informative component is totally discarded.

How vague is vague?

- Shmidli et al.[4] state that "proper" priors are needed (for interpretability) and suggest using Uniform (binary endpoint) or Unit-information priors, but no further justification is given.
- The (little, very little or extremely little) amount information in the vague component should be determined by the purpose of this component.
- But the vague component seems to have a **dual** purpose:

To reflect prior beliefs	It constitutes the alternative model, in which the information from source population is not relevant	More information
To produce robust inference	It "robustifies" inference when prior component is discarded (producing a behaviour close to the frequentist)	Less information

- **Recomendations:**
  - find a balance point by reflecting the beliefs of what is plausible, rather than what is probable,
  - justify the choice in an explicit way,
  - be clear about the consequences of the choice (i.e. bias).

[4] Schmidli et al. (2007) - Robust Meta-Analytic-Predictive Priors in Clinical trials with Historical Control Information, *Biometrics*.

Building a MP: the prior weights

- **Equal probability to both priors (w=0.5)**
  It reflects a state of "equipoise" (common practice in Bayesian Model Averaging).

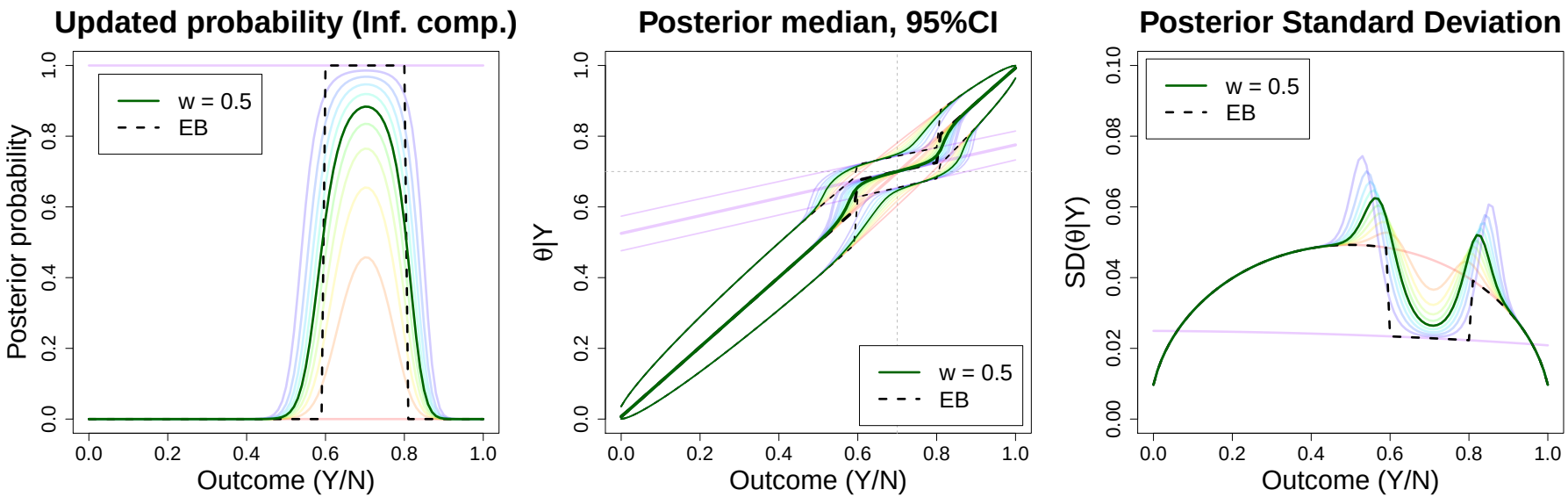
- **Elicitation from experts**
  - A single value:
    Elicitation could be done via a Likert scale.
  - A prior distribution:
    Full Bayesian inference with a prior $\pi(\omega)$ will lead to same inference for the parameter of interest as using a fixed value $E(\omega)$

- **Empirical Bayes (EB) estimation**
  Use the value of ω that maximizes the likelihood of Y.
  - $\omega = 1$ if $f_I(y) > f_V(y)$
  - $\omega = 0$ if $f_I(y) < f_V(y)$

  Similar to a "test and pool" approach.

Building a MP: the prior weights

Comparing the Empirical Bayes (EB) approach to fixed values for the prior probability (of informative component):



The “transition” of the posterior relying on the informative or the vague component is not gradual when using the EB approach

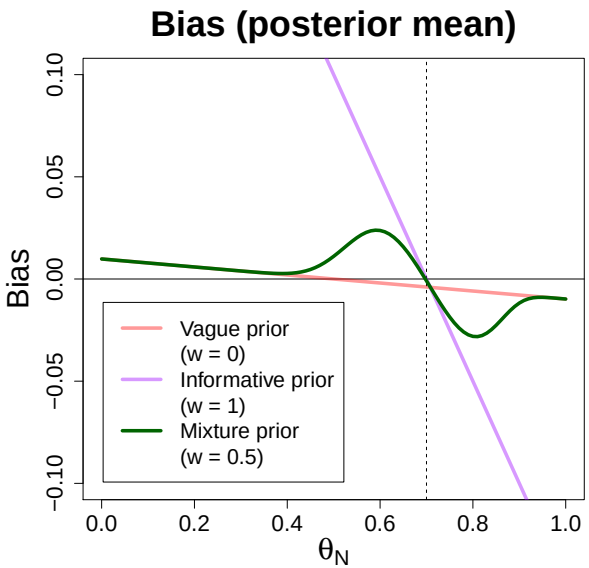
Frequentist Operating Characteristics (FOCs)

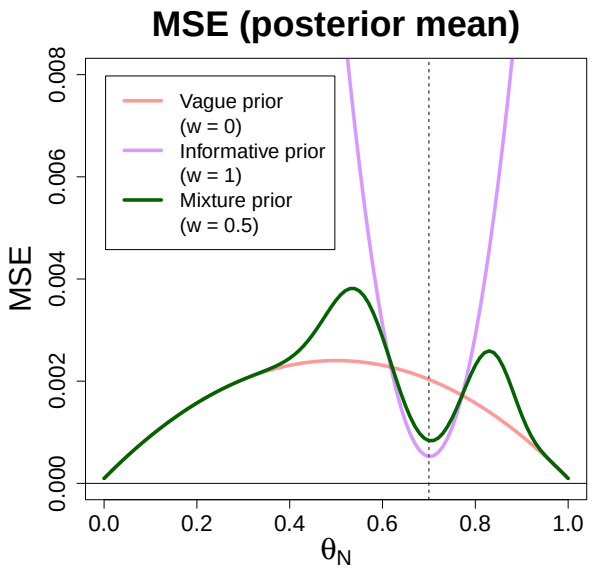
FOCs can be evaluated analytically or via simulation.

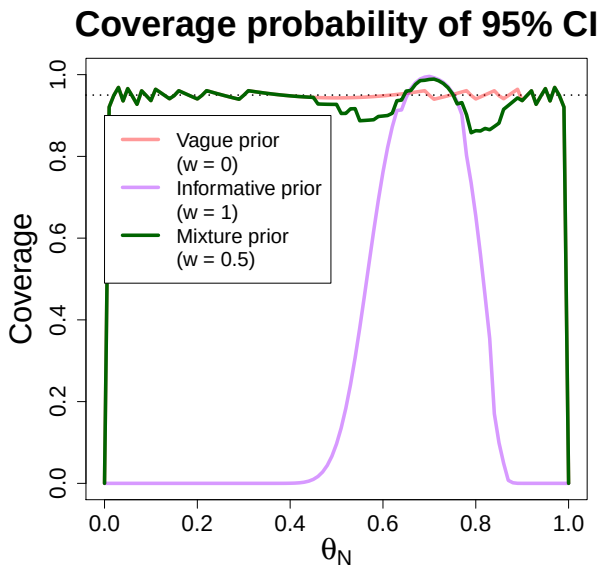
- **Estimation**: bias, MSE, coverage
- **Hypothesis testing**: type I error rate, power.

Mixture priors

Will not have uniformly better FOCs than frequentist methods, but they can limit or bound the "adverse" consequences of using an informative prior

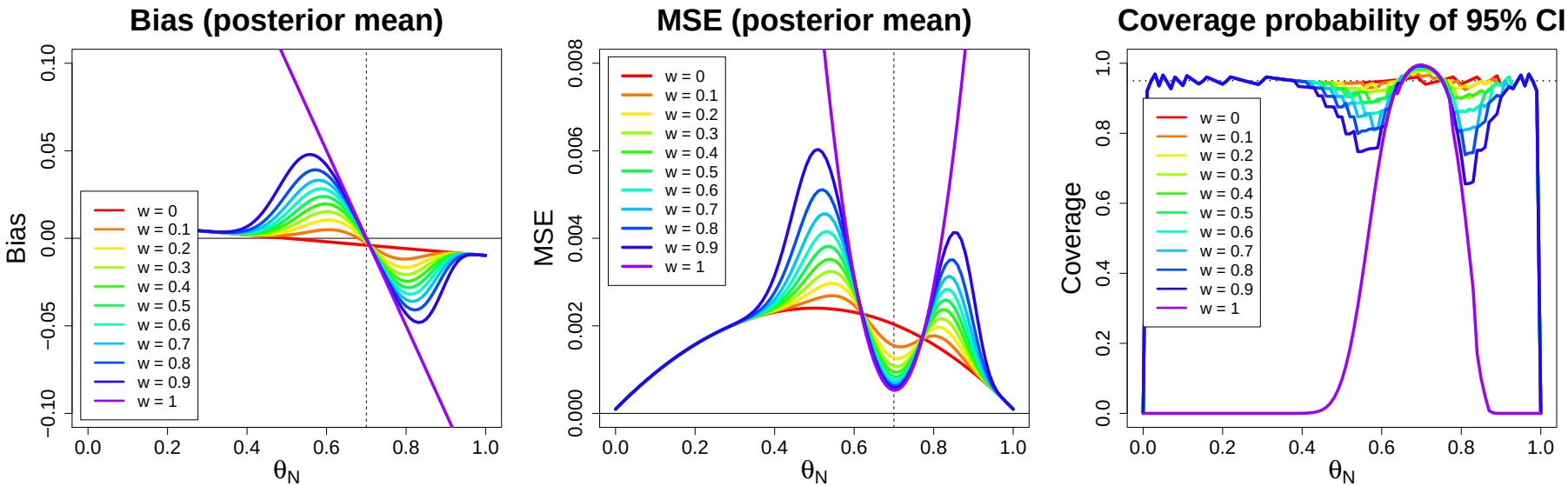






Frequentist Operating Characteristics (FOCs)

Evaluating the FOCs can be helpful as a way to calibrate the choices on the components of the mixture prior.

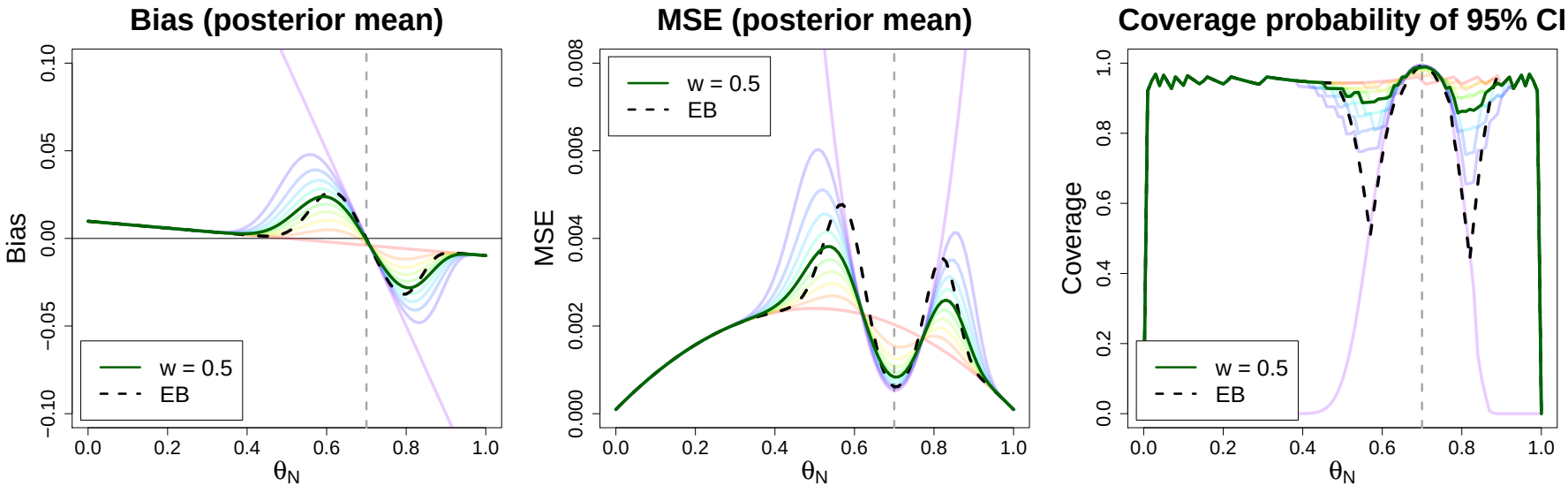


An appropriate value for *w* can be chosen to...

- limit the maximum possible bias,
- limit the maximum inflation on MSE,
- ensure a minimum desired probability of coverage.

Frequentist Operating Characteristics (FOCs)

FOCs can be helpful to compare different strategies:



The Empirical Bayes (EB) approach, compared to using a fixed prior probability ($w = 0.5$), has:

- similar bias,
- similar gains in MSE when data in agreement with prior, but slightly larger inflation when there is moderate conflict,
- a much lower bound for the coverage probability.

Back to motivating example

The opportunity :

Extrapolation from adult to pediatric population

Efficacy of the same treatment is being tested in adults (two fully powered trials, N $\approx$ 500). Believed to be very similar for children.

Extrapolation from historical to current controls

The active control treatment (currently recommended therapy) has been evaluated in previous paediatric studies (phase II and phase III RCTs).

Aim: To make use of historical information. Bayesian inference provides a natural framework

Problem: Although we believe on the similarity between populations, we acknowledge that they are not the same.

Challenge: Informative priors, but with some level of skepticism, to allow for potential conflict.

Bivariate mixture priors

Let: θ_0, θ_1 = probability of response in control and active arms
Δ = difference in probabilities ($\theta_1 - \theta_0$)

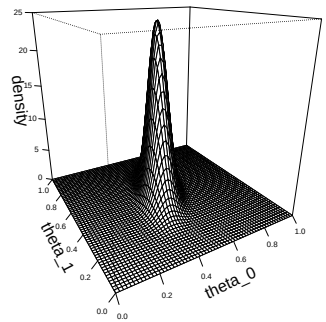
Constructing an informative bivariate prior:

Historical Controls
$\Downarrow$
Informative prior on θ_0
$\text{Beta}(\alpha, \beta)$

+

Efficacy in Adults
$\Downarrow$
Informative prior on $\Delta|\theta_0$
$\text{TN}_{[-\theta_0, 1-\theta_0]}(\mu, \sigma^2)$

=



Informative prior θ_0 and θ_1.

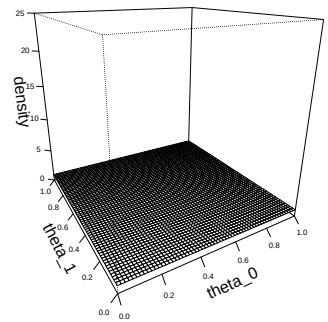
Constructing a vague bivariate prior:

Uniform prior on Θ_0
$\text{Beta}(1, 1)$

+

Vague prior on $\Delta|\theta_0$
$\text{TN}_{[-\theta_0, 1-\theta_0]}(\mu, 100 \times \sigma^2)$

=



Vague prior for θ_0 and θ_1.

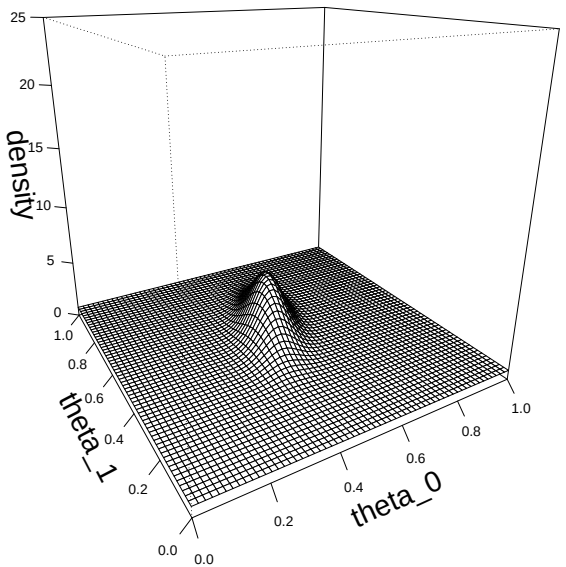
Bivariate mixture priors

Simple extension from univariate case:

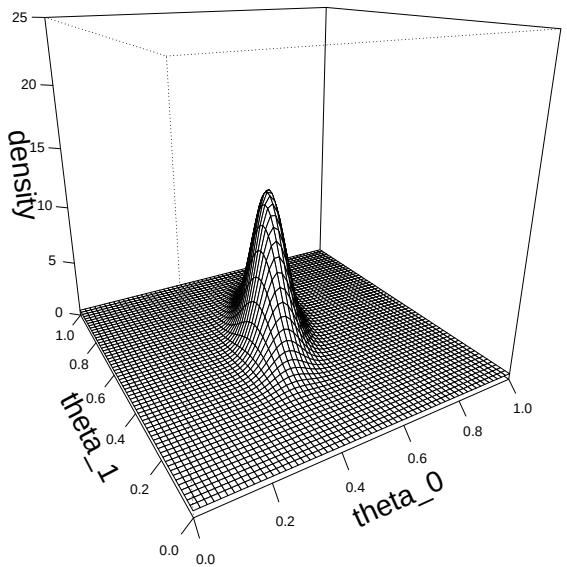
- Component 1: Informative bivariate distribution for both parameters (information from historical sources relevant for both θ_0 and Δ).
- Component 2: Vague bivariate distribution (none of the information relevant).

Robust prior: $\pi^R(\theta_0, \theta_1) = w\pi^I(\theta_0, \theta_1) + (1 - w)\pi^V(\theta_0, \theta_1)$

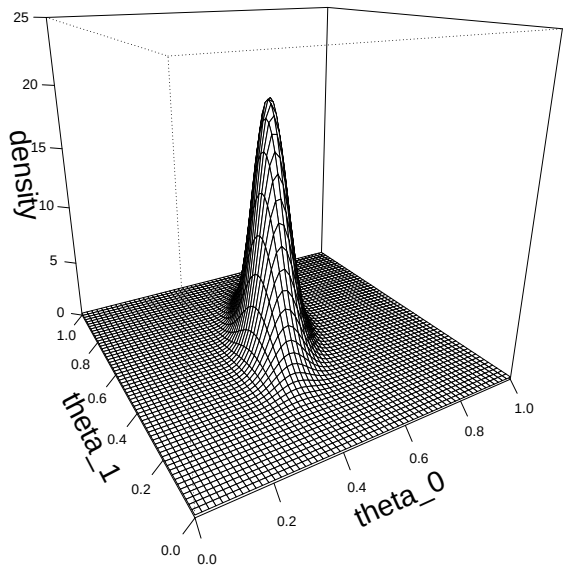
$w = 0.2$



$w = 0.5$



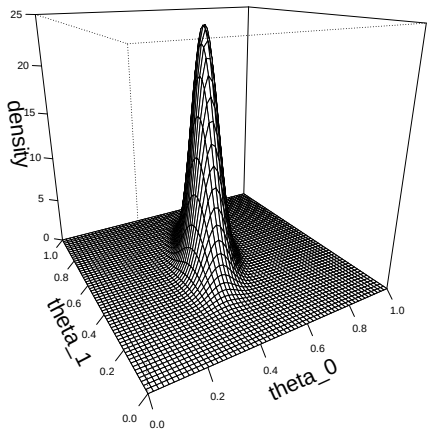
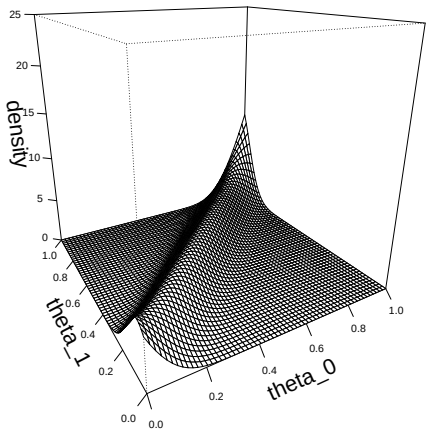
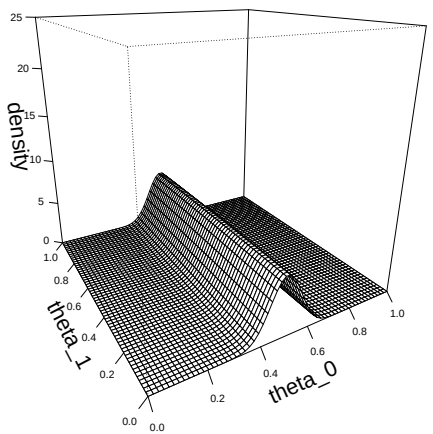
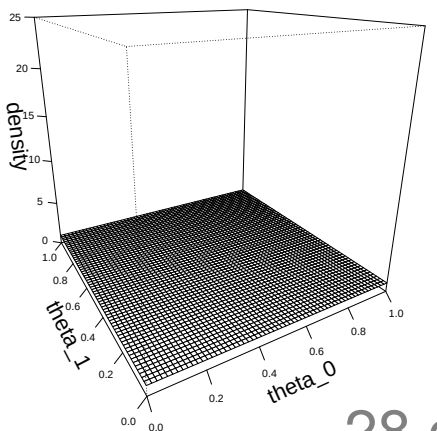
$w = 0.8$



Bivariate mixture priors

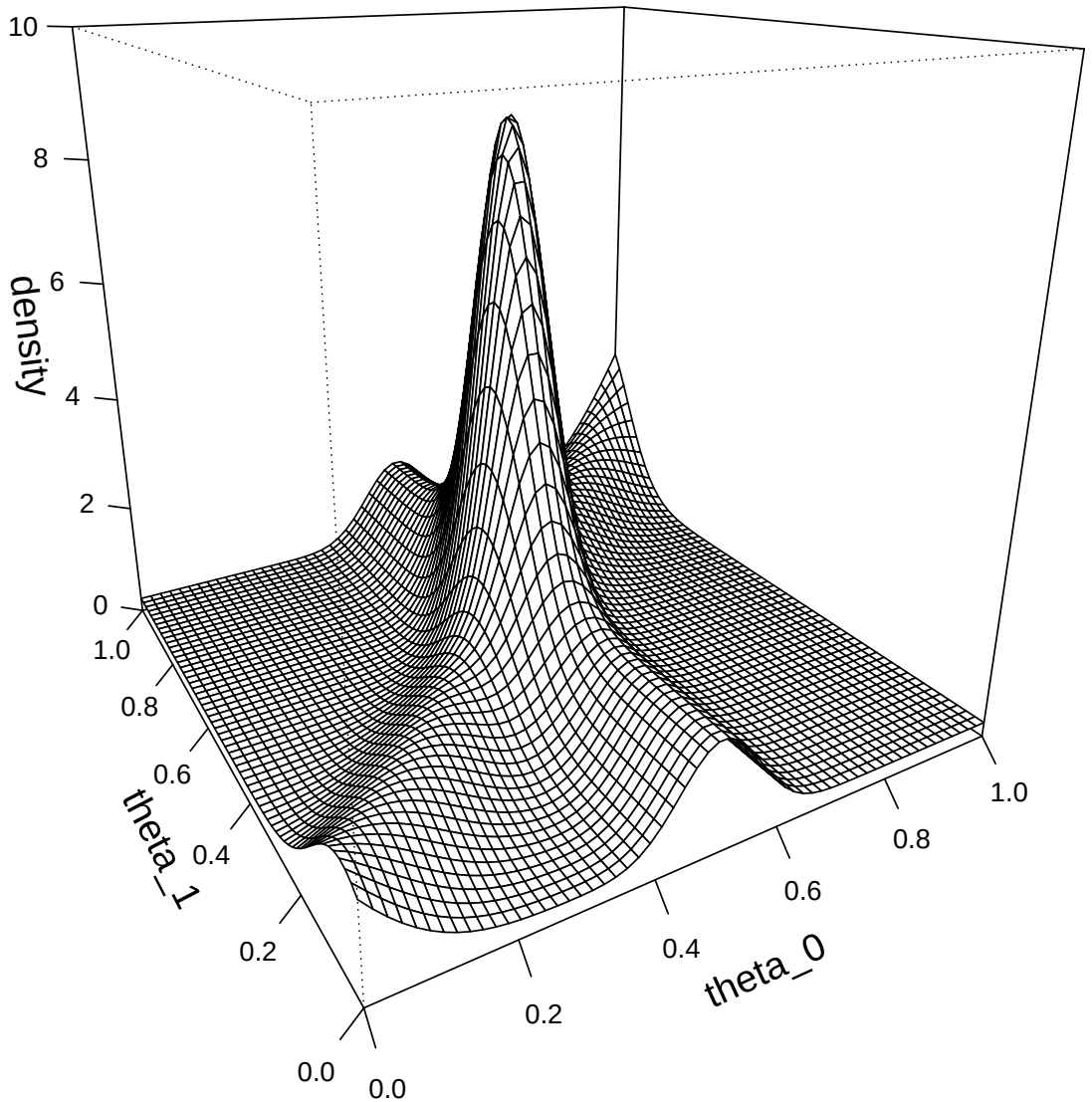
"Two way" mixture: four components

Mixture prior on Δ

Mixture prior on Θ_0	Informative component (w_Δ)	Vague component (1- w_Δ)
Informative component (w_0)		
Vague component ($1 - w_0$)		

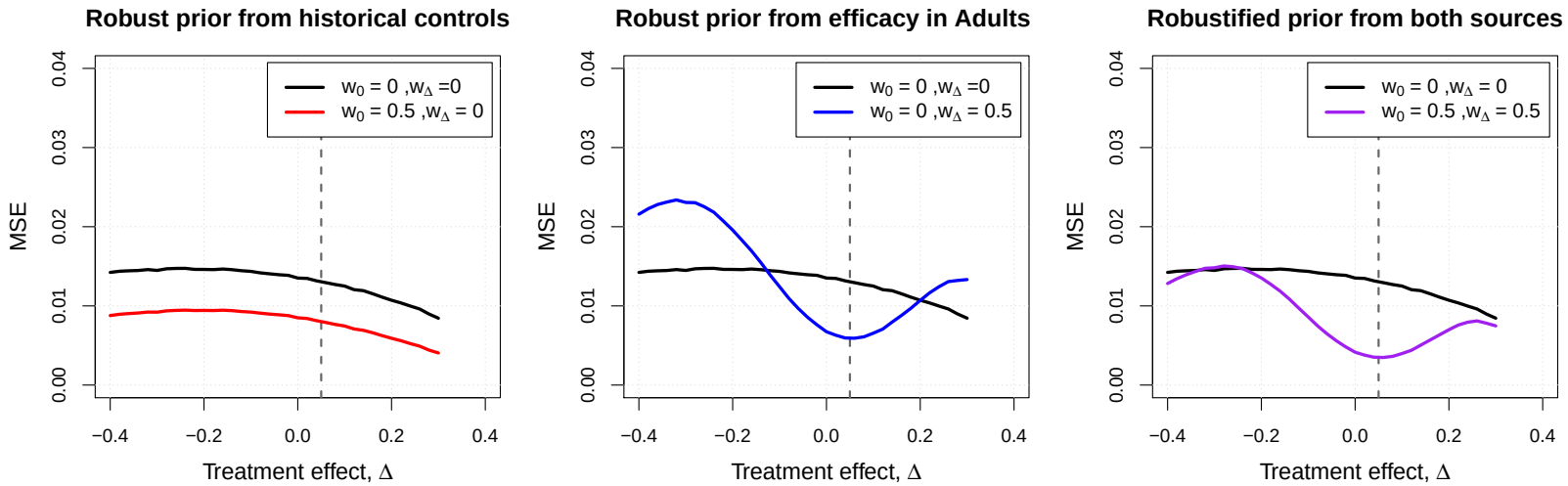
Bivariate mixture priors

With $w_0 = 0.5$, $w_\Delta = 0.5$:

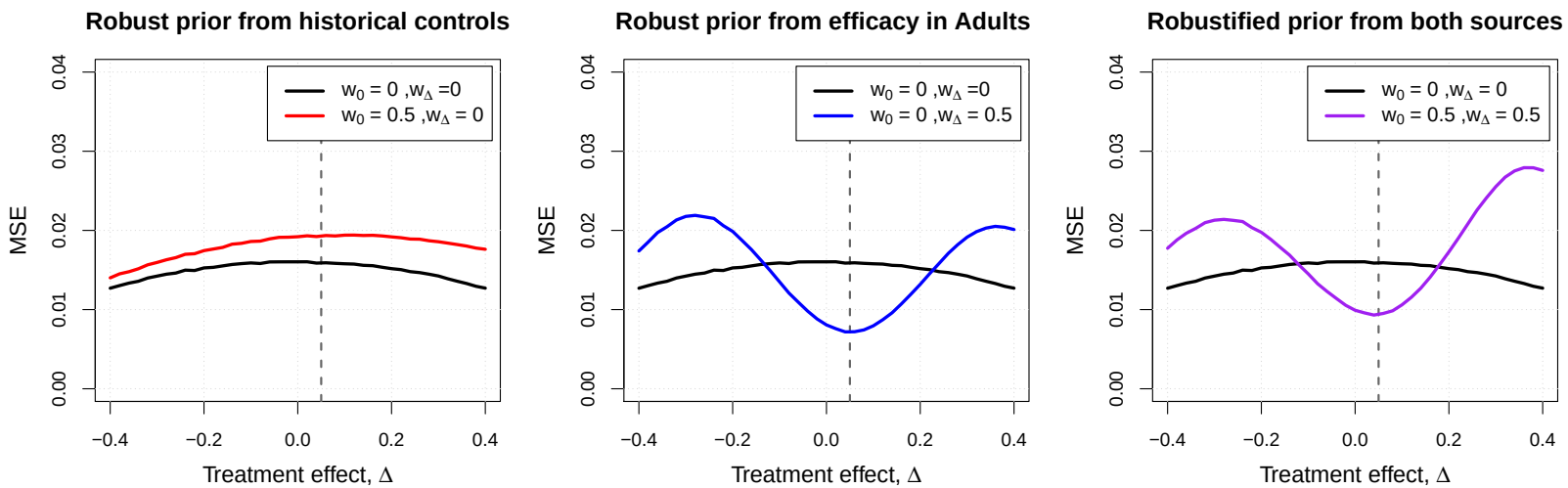


Bivariate mixture priors - FOCs

MSE - Prior-data agreement on control rate ($\Theta_0^N = 0.7 = E(\Theta_0^H)$)



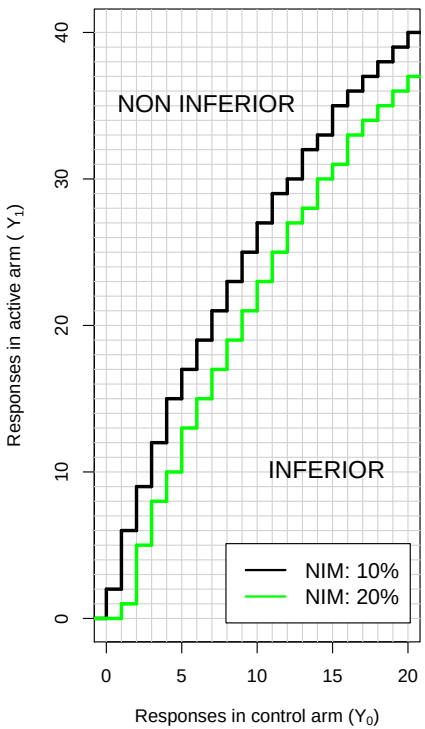
MSE - Prior-data conflict on control rate ($\Theta_0^N = 0.5; E(\Theta_0^H) = 0.7$)



Testing for non-inferiority: decision rule

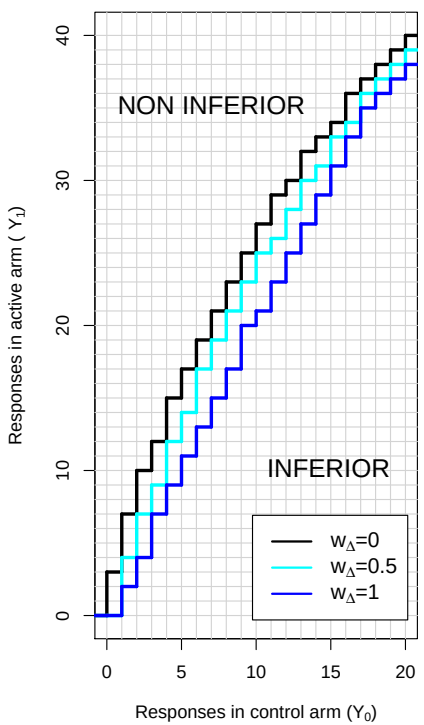
Frequentist

Relaxing NI Margin from -10% to -20%

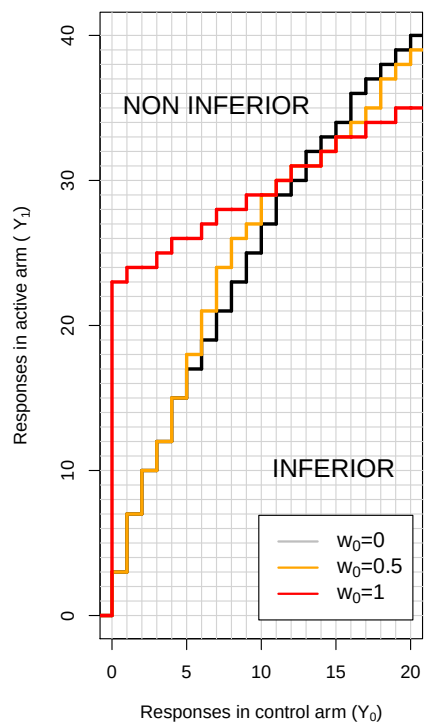


Bayesian (Mixture priors)

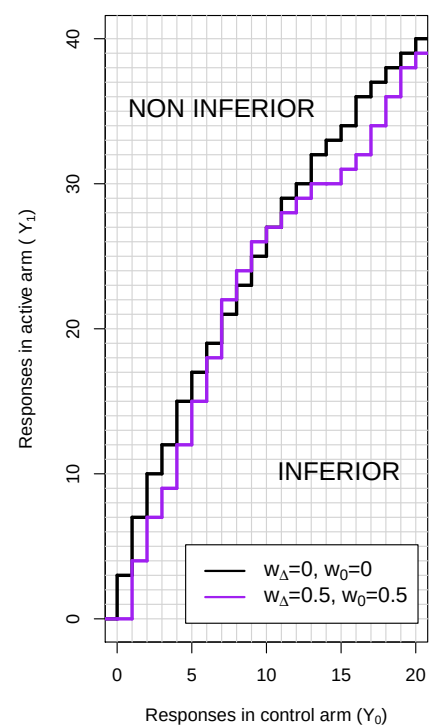
MP to extrapolate efficacy from adults



MP to extrapolate from historical controls



Two-way mixture prior (adults and hist. controls)



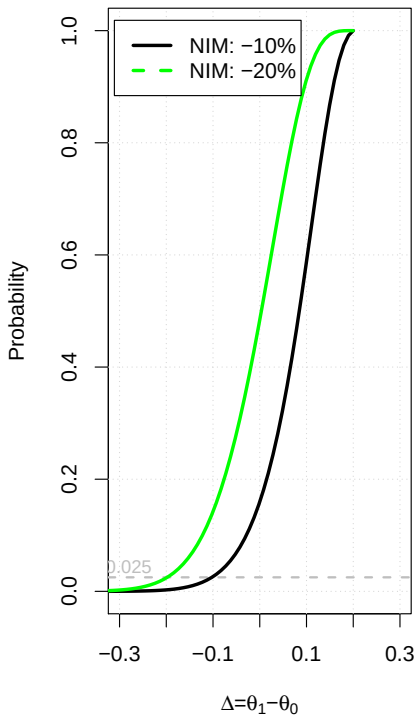
Bivariate mixture priors - FOCs

Testing for non-inferiority: probability of rejecting inferiority

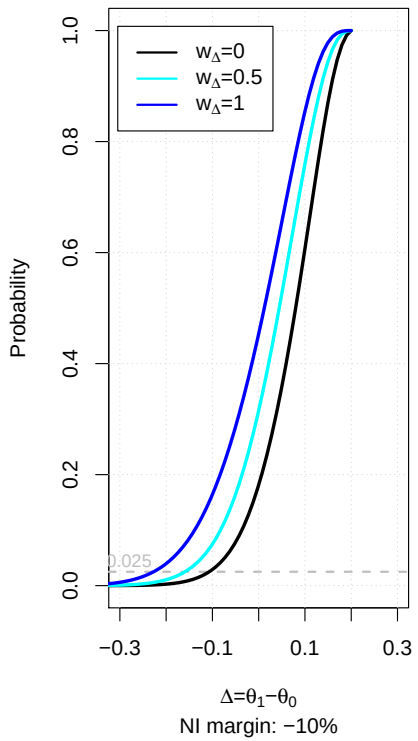
Frequentist

Bayesian (Mixture priors)

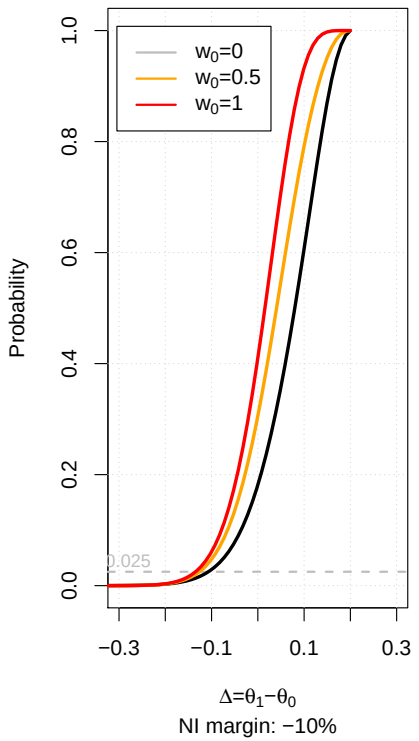
Relaxing NI Margin from -10% to -20%



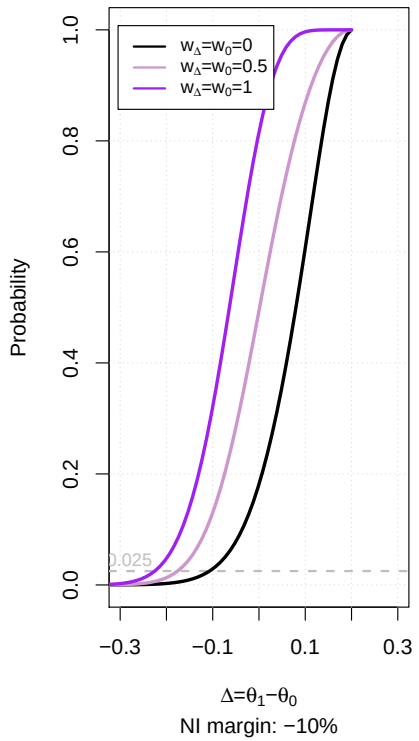
MP to extrapolate efficacy from adults



MP to extrapolate from historical controls



Two-way mixture prior (adults and hist. controls)



Increased power, at expense of increased probability "type I" error

Bivariate mixture priors - FOCs

Probability of rejecting inferiority (at $\theta_0 = 0.8$)

	Relaxing NI margin	**Extrapolation via a two-way mixture prior** $w_\Delta = 0.5, w_0 = 0.5$
Inferior treatment		
$\Delta = -0.2$	0.024	0.013
$\Delta = -0.1$	0.14	0.13
Non inferior treatment		
$\Delta = 0$	0.48	0.49
Superior treatment		
$\Delta = 0.1$	0.91	0.86

	Simple strategy, but may lack clinical justification	Complex strategy (requiring work and though around justifying data sources) but makes transparent and explicit assumptions

Conclusions

Mixture prior distributions for extrapolation, pros:

- **Model Averaging:** Allow to provide inference as an average of two models (in one of which the information from historical sources is relevant).
- **Interpretability:** Provide a measure of prior-data agreement (updated probability of relevance).
- **Robustness:** Provide gains in precision while limiting the extent of adverse consequences of using informative priors.
- **Testing:** Induce a decision rule that is built upon historical information and its probability of relevance for the current population(s).

Cons:

- Require an initial statement about the similarity between populations (prior probability).
- Need careful consideration for choosing the different components.
- Will have penalties (in the form of bias, MSE, type I error rate) when there is moderate conflict.