# 7. Model error and structural uncertainty

**Gianluca Baio**

Department of Statistical Science | University College London

g.baio@ucl.ac.uk
https://gianluca.statistica.it/
https://egon.stats.ucl.ac.uk/research/statistics-health-economics/
https://github.com/giabaio
https://github.com/StatisticsHealthEconomics
@gianlubaio

Bayesian Methods in Health Economics, Lausanne

# Summary

- Probabilistic sensitivity analysis to structural assumptions
  - Model average vs model comparison
  - Structural PSA using DIC
  - Example

## References


*Bayesian Methods in Health Economics*, chapters 3.6, 4.7.2 Book website (CRC) Book website Code

*The BUGS Book*, chapter 8.6 Book website

# PSA to structural assumptions

## Problem

- All the methods discussed so far assume that, while there may be uncertainty in the current knowledge, the model is "correct"
  - All variables that need to be considered have been considered
  - All distributional assumptions are reasonable (in fact, "correct")
  - ...

# PSA to structural assumptions

## Problem

- All the methods discussed so far assume that, while there may be uncertainty in the current knowledge, the model is “correct”
  – All variables that need to be considered have been considered
  – All distributional assumptions are reasonable (in fact, “correct”)
  – ...

> "All models are wrong, but some are useful" G. Box and N. Draper (1987). Empirical Model Building and Response Surfaces, Wiley & Sons, NY, p. 424

# PSA to structural assumptions

## Problem

- All the methods discussed so far assume that, while there may be uncertainty in the current knowledge, the model is "correct"
  - All variables that need to be considered have been considered
  - All distributional assumptions are reasonable (in fact, "correct")
  - ...

> "All models are wrong, but some are useful" G. Box and N. Draper (1987). Empirical Model Building and Response Surfaces, Wiley & Sons, NY, p. 424

- Possible (potential) solutions
  - Bayesian model averaging
  - Model comparison

# Bayesian model averaging

- "Proper" Bayesian way to deal with the issue
  - Identify a set of *finite and exhausting models* $\mathcal{M} = (\mathcal{M}_1, \ldots, \mathcal{M}_H)$
  - Each of the $H$ models is characterised by a set of parameters $\boldsymbol{\theta}$ and a suitable prior $p(\boldsymbol{\theta} \mid \mathcal{M}_h)$
  - In addition, define a prior $p(\mathcal{M}_h)$ that model $h$ is the "true" one
  - Update these as

$$p(\mathcal{M}_h \mid y) \propto p(\mathcal{M}_h) \int p(y \mid \boldsymbol{\theta}, \mathcal{M}_h) p(\boldsymbol{\theta} \mid \mathcal{M}_h) d\boldsymbol{\theta}$$

and use the posterior probabilities to compute a weighted average for any function of the parameters (eg utilities) over the space of models

# Bayesian model averaging

- "Proper" Bayesian way to deal with the issue
  - Identify a set of *finite and exhausting models* $\mathcal{M} = (\mathcal{M}_1, \ldots, \mathcal{M}_H)$
  - Each of the $H$ models is characterised by a set of parameters $\boldsymbol{\theta}$ and a suitable prior $p(\boldsymbol{\theta} \mid \mathcal{M}_h)$
  - In addition, define a prior $p(\mathcal{M}_h)$ that model $h$ is the "true" one
  - Update these as

$$p(\mathcal{M}_h \mid y) \propto p(\mathcal{M}_h) \int p(y \mid \boldsymbol{\theta}, \mathcal{M}_h) p(\boldsymbol{\theta} \mid \mathcal{M}_h) d\boldsymbol{\theta}$$

  and use the posterior probabilities to compute a weighted average for any function of the parameters (eg utilities) over the space of models

- **NB**: requires that *all* possible models are completely specified and given a prior probability

# Model comparison

- Compare a (not necessarily exhaustive!) set of models in terms of their out-of-sample prediction
  - Quantifies how well the predictive distribution for a given model would fit a replicated dataset based on the observed data
  - **NB**: especially in health economic evaluations, the possible models considered as merely a (rough) approximation to the complex phenomenon under study – so no guarantee that any of those should be the "true" one!

# Model comparison

- Compare a (not necessarily exhaustive!) set of models in terms of their out-of-sample prediction
  - Quantifies how well the predictive distribution for a given model would fit a replicated dataset based on the observed data
  - **NB**: especially in health economic evaluations, the possible models considered as merely a (rough) approximation to the complex phenomenon under study – so no guarantee that any of those should be the "true" one!

- Need to consider a measure of *goodness of fit* for each model
- **One** possibility is the Deviance Information Criterion (DIC)

# Model comparison

- Compare a (not necessarily exhaustive!) set of models in terms of their out-of-sample prediction
  - Quantifies how well the predictive distribution for a given model would fit a replicated dataset based on the observed data
  - **NB**: especially in health economic evaluations, the possible models considered as merely a (rough) approximation to the complex phenomenon under study – so no guarantee that any of those should be the "true" one!

- Need to consider a measure of *goodness of fit* for each model
- **One** possibility is the Deviance Information Criterion (DIC)

## Structural PSA via DIC

- Consider $\mathcal{M} = (\mathcal{M}_1, \ldots, \mathcal{M}_H)$
  1. Compute the value $\text{DIC}_h$ for each of them
  2. Derive the respective weights (model posterior probabilities) by simply re-proportioning them, eg

$$w_h = \frac{\exp(-0.5\Delta\text{DIC}_h)}{\sum_{h=1}^{H}\exp(-0.5\Delta\text{DIC}_h)}, \qquad \Delta\text{DIC}_h = \left|\min_h(\text{DIC}_h) - \text{DIC}_h\right|$$

  3. Use the weights $w_h$ to build an average model accounting for the (limited set) of plausible models

# Example – statins

Review: Statins
Comparison: 91 Placebo-controlled studies: CHD death plus non-fatal MI
Outcome: 01 CHD death plus non-fatal MI

| Study or subcategory | Treatment n/N | Control n/N |
|---|---|---|
| 01 Atorvastatin | | |
| DALI[86] | 0/145 | 1/72 |
| ASCOT-LLA[102] | 100/5168 | 154/5137 |
| Mohler[21] | 7/240 | 3/114 |
| CARDS[103] | 43/1428 | 65/1410 |
| Subtotal (95% CI) | 6981 | 6733 |
| Total events: 150 (treatment), 223 (control) | | |
| Test for heterogeneity: $\chi^2 = 1.32$, df = 3 ($p = 0.973$), $I^2 = 0\%$ | | |
| Test for overall effect: $z = 4.09$ ($p < 0.0001$) | | |
| 02 Fluvastatin | | |
| FLARE[108] | 6/409 | 17/425 |
| LiSA[93] | 2/187 | 5/178 |
| LIPS[110] | 42/844 | 60/833 |
| Subtotal (95% CI) | 1440 | 1436 |
| Total events: 50 (treatment), 82 (control) | | |
| Test for heterogeneity: $\chi^2 = 1.91$, df = 2 ($p = 0.39$), $I^2 = 0\%$ | | |
| Test for overall effect: $z = 2.76$ ($p = 0.006$) | | |
| 03 Pravastatin | | |
| PMSG[96] | 0/530 | 7/532 |
| KAPS[133] | 5/224 | 8/223 |
| WOSCOPS[82] | 174/3302 | 248/3293 |
| CAIUS[107] | 2/151 | 2/154 |
| CARE[111] | 212/2081 | 274/2078 |
| LIPID[112] | 557/4512 | 715/4502 |
| PROSPER[81] | 292/2891 | 356/2913 |
| Subtotal (95% CI) | 13,691 | 13,695 |
| Total events: 1242 (treatment), 1610 (control) | | |
| Test for heterogeneity: $\chi^2 = 4.95$, df = 6 ($p = 0.55$), $I^2 = 0\%$ | | |
| Test for overall effect: $z = 7.19$ ($p < 0.00001$) | | |
| 04 Simvastatin | | |
| 4S[97] | 431/2221 | 622/2223 |
| CIS[98] | 2/129 | 7/125 |
| HPS[74] | 898/10,269 | 1212/10,267 |
| Subtotal (95% CI) | 12,619 | 12,615 |
| Total events: 1331 (treatment), 1841 (control) | | |
| Test for heterogeneity: $\chi^2 = 2.38$, df = 2 ($p = 0.30$), $I^2 = 15.9\%$ | | |
| Test for overall effect: $z = 8.29$ ($p < 0.00001$) | | |
| Total (95% CI) | 34,731 | 34,479 |
| Total events: 2773 (treatment), 3756 (control) | | |
| Test for heterogeneity: $\chi^2 = 15.28$, df = 16 ($p = 0.50$), $I^2 = 0\%$ | | |
| Test for overall effect: $z = 12.94$ ($p < 0.00001$) | | |

- Based on published data on RCTs comparing statins to placebo
- Complex formulation: in particular, model the response for the controls as

$$y_{sj}^{\text{ctr}} \sim \text{Binomial}(\theta_s, n_{sj}^{\text{ctr}})$$
$$\text{logit}(\theta_s) = \alpha_s$$

and then use 2 formulations for the prior on $\alpha_s$

1. **Normal**
   - $\alpha_s \sim \text{Normal}(\mu_\alpha, \tau_\alpha)$
   - $\mu_\alpha \sim \text{Normal}(0, 0.00001)$
   - $\sigma_\alpha = \tau_\alpha^{-2} \sim \text{Uniform}(0, 20)$
2. **Half-Cauchy** (robust alternative)
   - $\alpha_s = \mu_\alpha + \xi\eta_s$
   - $\mu_\alpha \sim \text{Normal}(0, 0.00001)$
   - $\xi \sim \text{Normal}(0, \tau_\xi)$
   - $\eta_s \sim \text{Normal}(0, \tau_\eta)$
   - $\tau_\eta \sim \text{Gamma}(0.5, 0.5)$; $\tau_\xi = 12^{-2}$

# Example – statins

Model $\mathcal{M}_1$

```
> ...
> ## Priors for the hyperparameters
> ## Exchangeable normal prior
> for (s in 1:Nstatins) {
+   alpha[s] ~ dnorm(mu.alpha, tau.alpha)
+ }
> sigma.alpha ~ dunif(0,20)
> tau.alpha <- pow(sigma.alpha,-2)
> mu.alpha ~ dnorm(0,0.0001)
> sigma.alpha <- abs(xi) / sqrt(tau.eta)
> prior.scale <- 12
> mu.alpha ~ dnorm(0,.0001)
> ...
```

Model $\mathcal{M}_2$

```
> ...
> ## Priors for the hyperparameters
> ## Half-Cauchy prior
> for (s in 1:Nstatins) {
+   alpha[s] <- mu.alpha + xi*eta[s]
+   eta[s] ~ dnorm(0,tau.eta)
+ }
> xi ~ dnorm(0,tau.xi)
> tau.xi <- pow(prior.scale,-2)
> tau.eta ~ dgamma(.5,.5)
> sigma.alpha <- abs(xi) / sqrt(tau.eta)
> prior.scale <- 12
> mu.alpha ~ dnorm(0,.0001)
> ...
```

# Example – statins

`print(m1)` → R object containing the MCMC simulations for $\mathcal{M}_1$

```
Inference for Bugs model at "model.txt", fit using  jags , 2  chains, each with 100000
iterations  (first  9500  discarded), n.thin = 181 n.sims = 1000 iterations saved

               mean    sd  2.5%   25%   50%   75%  97.5% Rhat n.eff
cost.hosp[1] 238.7 137.0  91.7 168.6 213.5 278.2  482.0    1   980
cost.hosp[2] 315.6 168.8 124.2 219.2 290.2 372.7  668.5    1  1000
cost.hosp[3] 523.1 451.0 144.0 297.0 427.9 619.9 1358.0    1  1000

DIC info (using the rule, pD = var(deviance)/2)
pD =  71.1  and DIC =  2233.9
DIC is an estimate of expected predictive error (lower deviance is better)
```

`print(m2)` → R object containing the MCMC simulations for $\mathcal{M}_2$

```
Inference for Bugs model at "model.txt", fit using  jags , 2  chains, each with 100000
iterations  (first  9500  discarded), n.thin = 181 n.sims = 1000 iterations saved

               mean    sd  2.5%   25%   50%   75%  97.5% Rhat n.eff
cost.hosp[1] 228.9 108.0  82.9 161.6 209.3 272.9  502.5    1  1000
cost.hosp[2] 301.5 143.0 101.2 211.2 276.6 366.5  621.8    1   960
cost.hosp[3] 469.8 277.1 128.6 280.5 402.1 583.2 1184.6    1  1000

DIC info (using the rule, pD = var(deviance)/2)
pD =  63.9  and DIC =  2226.0
DIC is an estimate of expected predictive error (lower deviance is better)
```

# Example – statins

```
> library(BCEA)
> # Objects containing the MCMC simulations from the posteriors for the two models
> m1.sims <- m1$sims.list;
> m2.sims <- m2$sims.list
>
> # Defines suitable variables of clinical benefits & costs
> e1 <- m1.sims$effect;
> e2 <- m2.sims$effect
> c1 <- m1.sims$cost.tot;
> c2 <- m2.sims$cost.tot
>
> # Runs BCEA's function to do PSA to structural assumptions
> avg <- struct.psa(list(m1,m2),list(e1,e2),list(c1,c2),ref=2)
```

```
> # Weights associated with each model (based on DIC)
> avg$w
```

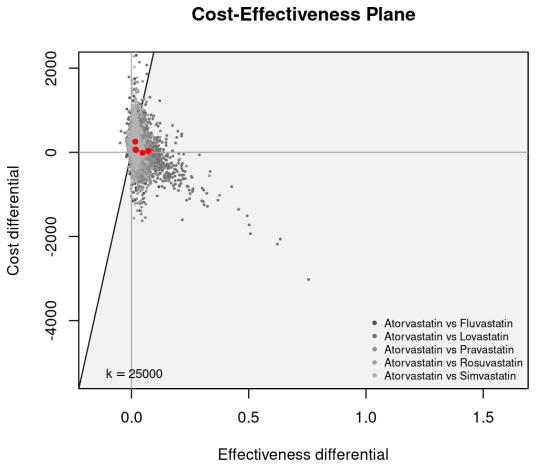
```
[1] 0.01901127 0.98098873
```

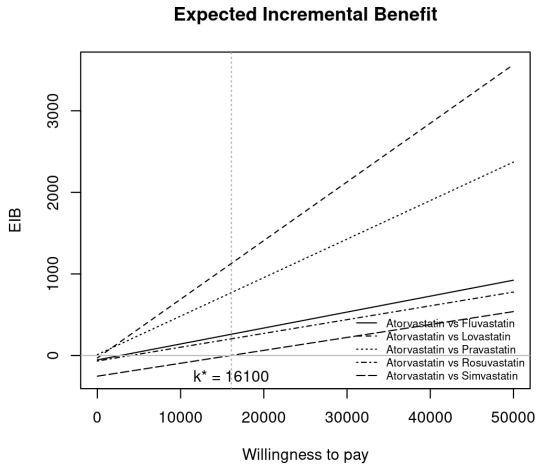
```
> # Actual DIC computed for each model
> avg$DIC
```

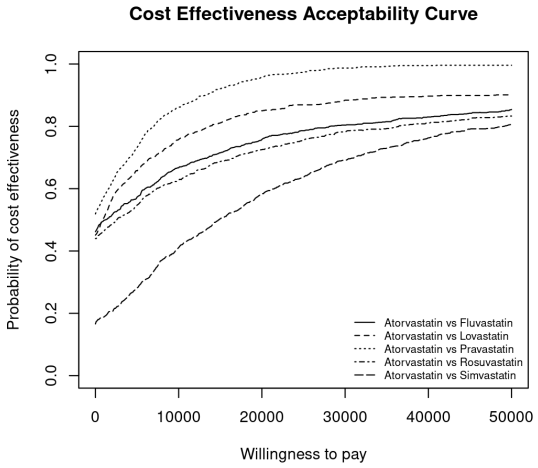
```
[1] 2233.875 2225.988
```

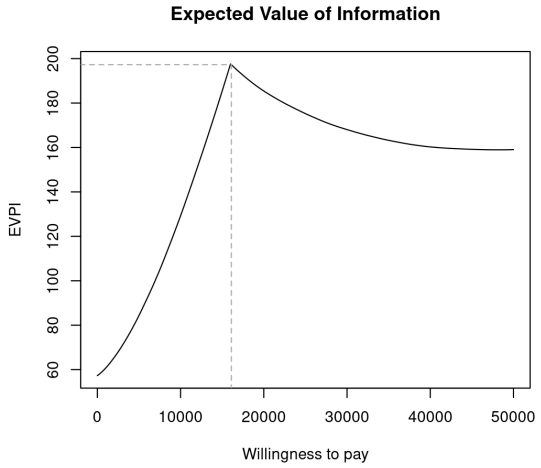
# Example – statins

Model 1 (Normal prior) | Model 2 (HC prior) | Model average



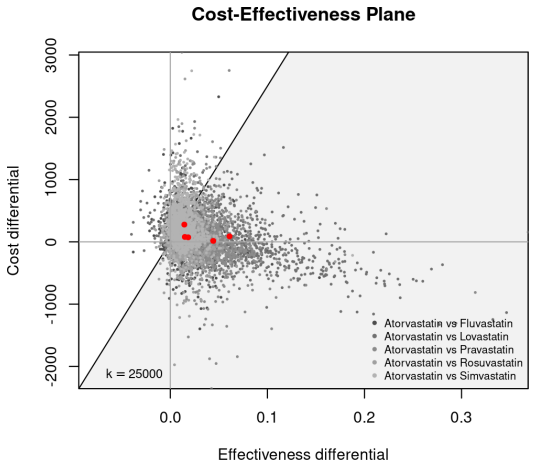



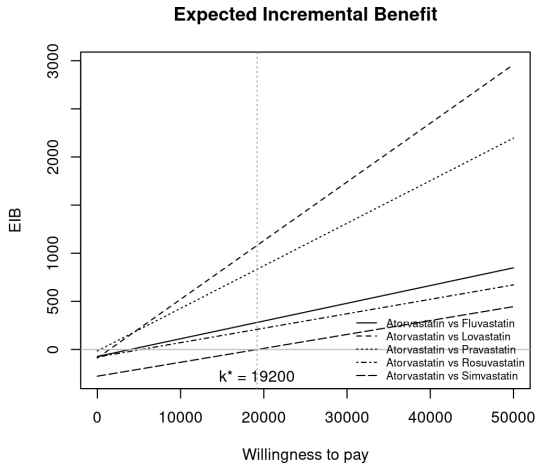



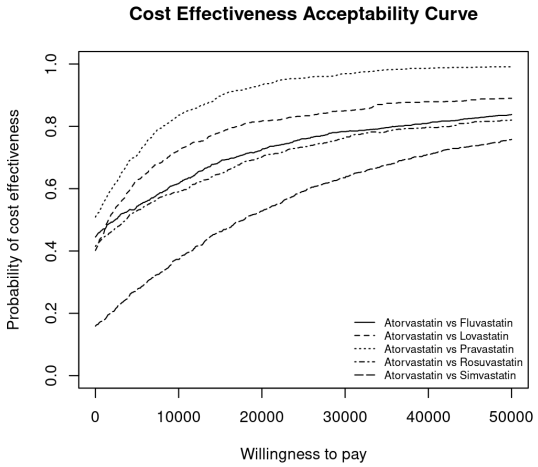

- DIC = 2233.875
- $w_1$ = 0.019

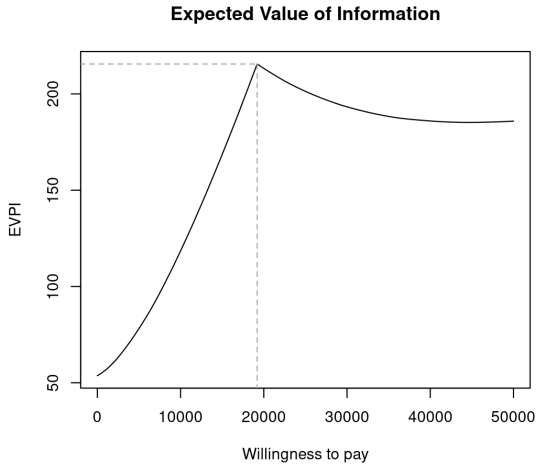
# Example – statins

Model 1 (Normal prior) | **Model 2 (HC prior)** | Model average



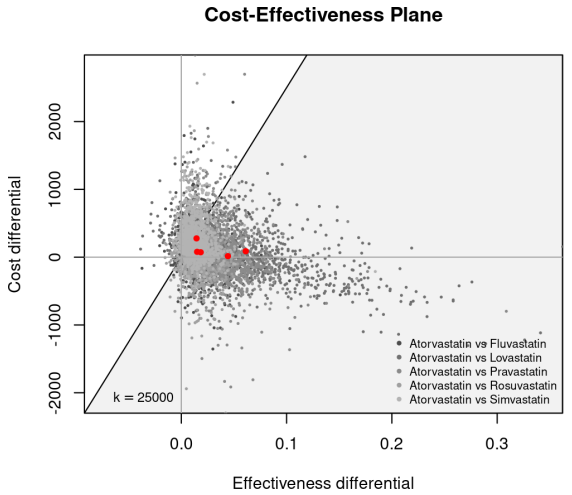



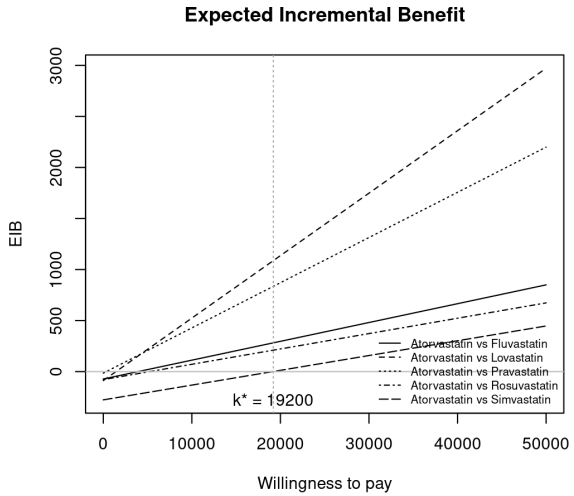



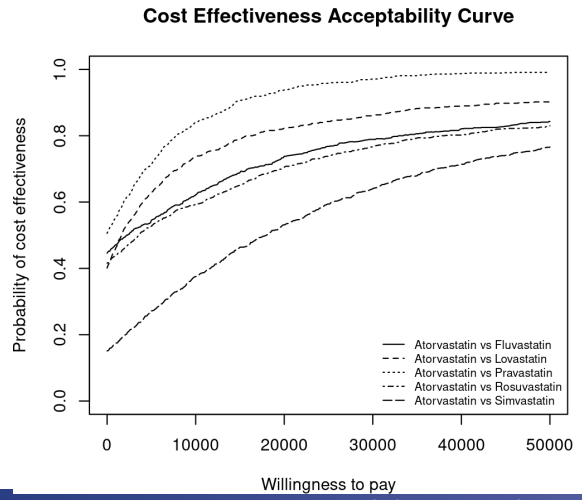

- DIC = 2225.988
- $w_2$ = 0.981

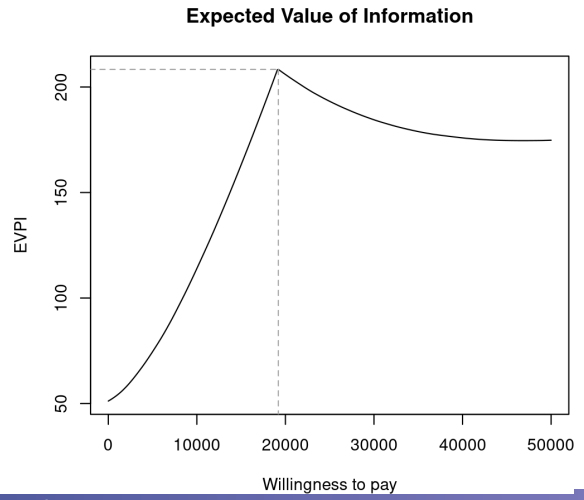
# Example – statins

Model 1 (Normal prior) | Model 2 (HC prior) | **Model average**



# Example – statins

Model $\mathcal{M}_2$

```
Analysis for willingness to pay parameter k = 25000

             Expected utility
Atorvastatin            22824
Fluvastatin             22437
Lovastatin              21389
Pravastatin             21734
Rosuvastatin            22528
Simvastatin             22741

                                  EIB  CEAC     ICER
Atorvastatin vs Fluvastatin   387.804 0.760  3943.45
Atorvastatin vs Lovastatin   1435.830 0.833  1455.71
Atorvastatin vs Pravastatin  1090.236 0.955   348.42
Atorvastatin vs Rosuvastatin  296.181 0.733  5254.11
Atorvastatin vs Simvastatin    83.889 0.591 19195.85

Optimal intervention (max expected utility)
for k = 25000: Atorvastatin


EVPI 201.33
```

Model average

```
Analysis for willingness to pay parameter k = 25000

             Expected utility
Atorvastatin            22823
Fluvastatin             22435
Lovastatin              21381
Pravastatin             21731
Rosuvastatin            22526
Simvastatin             22738

                                  EIB  CEAC     ICER
Atorvastatin vs Fluvastatin   388.695 0.768  3919.60
Atorvastatin vs Lovastatin   1442.154 0.843  1432.44
Atorvastatin vs Pravastatin  1092.152 0.958   336.72
Atorvastatin vs Rosuvastatin  297.304 0.738  5227.07
Atorvastatin vs Simvastatin    85.002 0.594 19129.49

Optimal intervention (max expected utility)
for k = 25000: Atorvastatin


EVPI 193.22
```

Next lecture