# Investigating Discrepancy in Computer Model Predictions

Jeremy Oakley[1] and Mark Strong[2]

[1]School of Mathematics and Statistics, and [2]School of Health and Related Research, University of Sheffield

# Application: health economic modelling in medical decision making

- A health-care provider wishes to spend budget efficiently
- Considers both treatment costs and clinical effectiveness
- In UK, National Institute for Health and Clinical Excellence (NICE) decides whether some treatments are available on NHS, based on cost-effectiveness
- Treatment efficacy and costs combined into measure of cost-effectiveness (net-benefit), used for decision-making
- Don't directly 'observe' (long-term) cost-effectiveness, so make predictions using health economic models

# Application: health economic modelling in medical decision making

- A health-care provider wishes to spend budget efficiently
- Considers both treatment costs and clinical effectiveness
- In UK, National Institute for Health and Clinical Excellence (NICE) decides whether some treatments are available on NHS, based on cost-effectiveness
- Treatment efficacy and costs combined into measure of cost-effectiveness (net-benefit), used for decision-making
- Don't directly 'observe' (long-term) cost-effectiveness, so make predictions using health economic models

**Model discrepancy**

Model is not perfect. How close is the model prediction to reality?

# Uncertainty in model inputs (ignoring discrepancy)

# Uncertainty in model inputs (ignoring discrepancy)

- Economic model $y = f(x)$, with uncertain true inputs $X$
- Wish to assess uncertainty in $Y = f(X)$

# Uncertainty in model inputs (ignoring discrepancy)

- Economic model $y = f(x)$, with uncertain true inputs $X$
- Wish to assess uncertainty in $Y = f(X)$
- Derive joint probability distribution $p(X)$ given available data
- Sample from $p(X)$ to get sample from $p(Y)$

# Uncertainty in model inputs (ignoring discrepancy)

- Economic model $y = f(x)$, with uncertain true inputs $X$
- Wish to assess uncertainty in $Y = f(X)$
- Derive joint probability distribution $p(X)$ given available data
- Sample from $p(X)$ to get sample from $p(Y)$
- (Almost) required by NICE, referred to as “probabilistic sensitivity analysis” (PSA)

# Uncertainty due to model discrepancy

# Uncertainty due to model discrepancy

- Economic model $y = f(x)$, with uncertain true inputs $X$, and $Y = f(X)$.
- Target quantity $Z$, to be estimated using model.

# Uncertainty due to model discrepancy

- Economic model $y = f(x)$, with uncertain true inputs $X$, and $Y = f(X)$.
- Target quantity $Z$, to be estimated using model.
- If model is not perfect then $Y = f(X) \neq Z$
- PSA quantifies uncertainty about $Y$. How do we quantify uncertainty about $Z$?

# Uncertainty due to model discrepancy

- Economic model $y = f(x)$, with uncertain true inputs $X$, and $Y = f(X)$.
- Target quantity $Z$, to be estimated using model.
- If model is not perfect then $Y = f(X) \neq Z$
- PSA quantifies uncertainty about $Y$. How do we quantify uncertainty about $Z$?
- We define the model discrepancy as $\delta = Z - f(X)$

# Uncertainty due to model discrepancy

- Economic model $y = f(x)$, with uncertain true inputs $X$, and $Y = f(X)$.
- Target quantity $Z$, to be estimated using model.
- If model is not perfect then $Y = f(X) \neq Z$
- PSA quantifies uncertainty about $Y$. How do we quantify uncertainty about $Z$?
- We define the model discrepancy as $\delta = Z - f(X)$
- In health economic modelling, do not have data to learn about $\delta$
- What can we say about $\delta$?

# Assessing model discrepancy without data

$\delta = Z - f(X)$

- Direct elicitation of $p(\delta)$?

# Assessing model discrepancy without data

$\delta = Z - f(X)$

- Direct elicitation of $p(\delta)$?
    - Impossible?

# Assessing model discrepancy without data

$\delta = Z - f(X)$

- Direct elicitation of $p(\delta)$?
    - Impossible? Then why use the model?!

# Assessing model discrepancy without data

$\delta = Z - f(X)$

- Direct elicitation of $p(\delta)$?
  - Impossible? Then why use the model?!
  - Certainly difficult, probably not useful

# Assessing model discrepancy without data

$\delta = Z - f(X)$

- Direct elicitation of $p(\delta)$?
  - Impossible? Then why use the model?!
  - Certainly difficult, probably not useful
- We consider investigating model discrepancy as a tool for model developers
  1. Many sources of discrepancy, which are the most important?
  2. Is it worth building a more complex model?

# Assessing model discrepancy without data

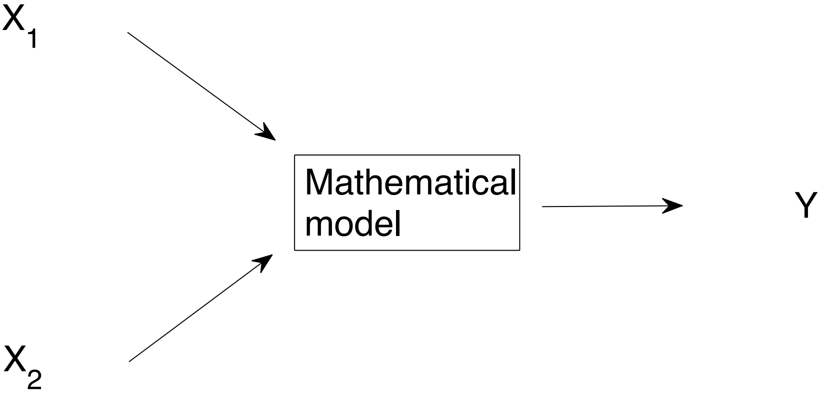
$\delta = Z - f(X)$

- Direct elicitation of $p(\delta)$?
  - Impossible? Then why use the model?!
  - Certainly difficult, probably not useful
- We consider investigating model discrepancy as a tool for model developers
  1. Many sources of discrepancy, which are the most important?
  2. Is it worth building a more complex model?

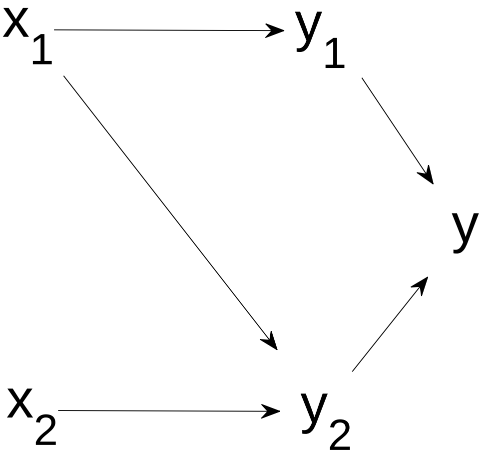
**Our approach**

'Open the black box' and incorporate discrepancy within the model

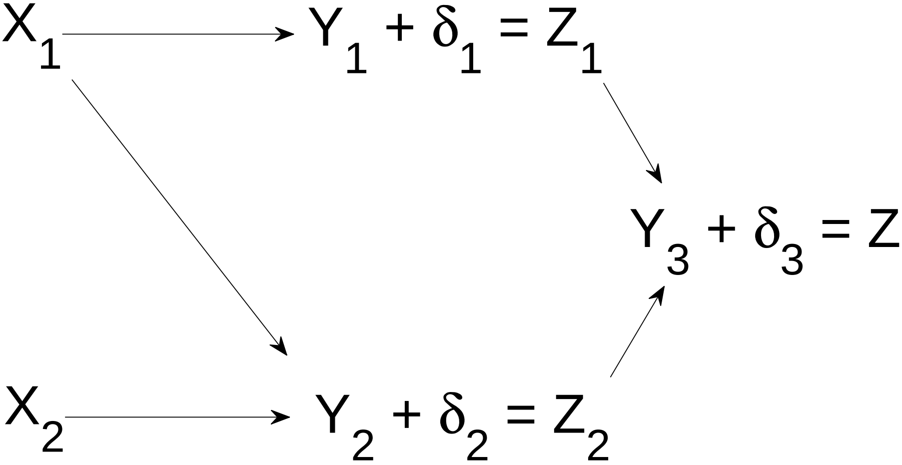
# The model



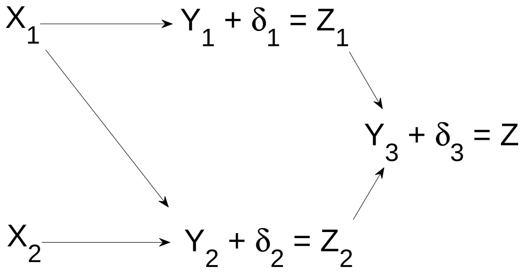
# Opening the black box



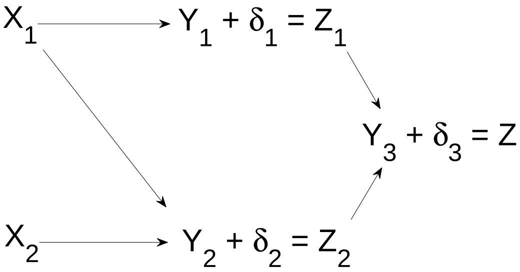
# Linking the model to reality



# In previous work (Strong et al 2011 in JRSS C)

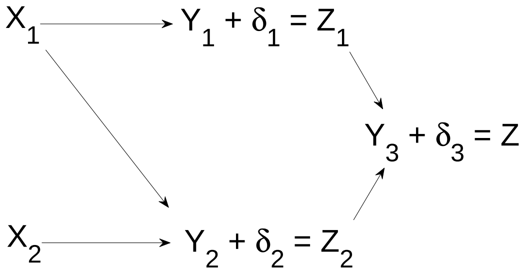


# In previous work (Strong et al 2011 in JRSS C)




- Specify joint distribution for $X_1, X_2, \delta_1, \delta_2, \delta_3$

# In previous work (Strong et al 2011 in JRSS C)




- Specify joint distribution for $X_1, X_2, \delta_1, \delta_2, \delta_3$
- Use variance-based sensitivity analysis to investigate how learning $\delta_i$ would reduce variance of $Z$

$$\frac{Var_{\delta_i}\{E(Z|\delta_i)\}}{Var(Z)}$$

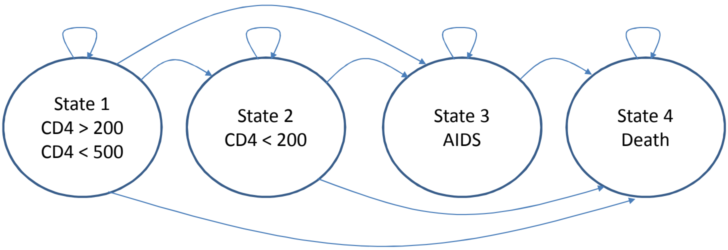
# Analysing discrepancy for decision making

- Suppose model is to be used to guide a decision. Could reducing model discrepancy lead to a different decision?

# Analysing discrepancy for decision making

- Suppose model is to be used to guide a decision. Could reducing model discrepancy lead to a different decision?
- Same framework of introducing discrepancy parameters within a model
- Now use standard decision theory (expected value of perfect information) for quantifying value of learning discrepancy parameters

# Case study: a Markov model




- Model described in Chancellor et al. (1997)
- Predicts costs and health outcomes for zidovudine monotherapy versus zidovudine plus lamivudine combination therapy in HIV patients
- Patients move between states once per year, for 20 years.

- Uncertain transition matrices for monotherapy

$$M_1 = \begin{pmatrix} p_{111} & p_{112} & p_{113} & p_{114} \\ 0 & p_{122} & p_{123} & p_{124} \\ 0 & 0 & p_{133} & p_{134} \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

and for combination therapy: $M_2 =$

$$\begin{pmatrix} 1 - RR\sum_{k=2}^{4} p_{11k} & RR \cdot p_{112} & RR \cdot p_{113} & RR \cdot p_{114} \\ 0 & 1 - RR(p_{123} + p_{124}) & RR \cdot p_{123} & RR \cdot p_{124} \\ 0 & 0 & 1 - RR \cdot p_{133} & RR \cdot p_{134} \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

- Uncertain transition matrices for monotherapy

$$M_1 = \begin{pmatrix} p_{111} & p_{112} & p_{113} & p_{114} \\ 0 & p_{122} & p_{123} & p_{124} \\ 0 & 0 & p_{133} & p_{134} \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

and for combination therapy: $M_2 =$

$$\begin{pmatrix} 1 - RR\sum_{k=2}^{4} p_{11k} & RR \cdot p_{112} & RR \cdot p_{113} & RR \cdot p_{114} \\ 0 & 1 - RR(p_{123} + p_{124}) & RR \cdot p_{123} & RR \cdot p_{124} \\ 0 & 0 & 1 - RR \cdot p_{133} & RR \cdot p_{134} \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

- But constant $M_1$ and $M_2$ for all times steps is a source of model discrepancy

# Putting discrepancy into the Markov model

# Putting discrepancy into the Markov model

- Evolution of $4 \times 1$ state vector $h_{d,t}$ for treatment option $d \in \{1, 2\}$ at time $t$ is

$$h_{d,t} = M'_d h_{d,t-1}.$$

# Putting discrepancy into the Markov model

- Evolution of $4 \times 1$ state vector $h_{d,t}$ for treatment option $d \in \{1,2\}$ at time $t$ is

$$h_{d,t} = M_d' h_{d,t-1}.$$

- Could add discrepancy directly to the state vector

$$h_{d,t} = M_d' h_{d,t-1} + \delta_{d,t},$$

# Putting discrepancy into the Markov model

- Evolution of $4 \times 1$ state vector $h_{d,t}$ for treatment option $d \in \{1, 2\}$ at time $t$ is

$$h_{d,t} = M'_d h_{d,t-1}.$$

- Could add discrepancy directly to the state vector

$$h_{d,t} = M'_d h_{d,t-1} + \delta_{d,t},$$

but more intuitive to add discrepancy to the transition matrix

$$h_{d,t} = (M'_d + \Delta_{d,t}) h_{d,t-1}.$$

# Putting discrepancy into the Markov model

- Evolution of $4 \times 1$ state vector $h_{d,t}$ for treatment option $d \in \{1, 2\}$ at time $t$ is

$$h_{d,t} = M'_d h_{d,t-1}.$$

- Could add discrepancy directly to the state vector

$$h_{d,t} = M'_d h_{d,t-1} + \delta_{d,t},$$

  but more intuitive to add discrepancy to the transition matrix

$$h_{d,t} = (M'_d + \Delta_{d,t}) h_{d,t-1}.$$

- Need to elicit distribution for $\{\Delta_{d,t}\}_{d=1,2;t=1,\ldots,20}$

# Putting discrepancy into the Markov model

$$
\begin{aligned}
h_{d,t} &= (M'_d + \Delta_{d,t})h_{d,t-1}, \\
\Delta_{dt} &= \begin{pmatrix} -(\delta_{d1t} + \delta_{d2t} + \delta_{d3t}) & \delta_{d1t} & \delta_{d2t} & \delta_{d3t} \\ 0 & -(\delta_{d4t} + \delta_{d5t}) & \delta_{d4t} & \delta_{d5t} \\ 0 & 0 & -\delta_{d6t} & \delta_{d6t} \\ 0 & 0 & 0 & 0 \end{pmatrix}.
\end{aligned}
$$

# Putting discrepancy into the Markov model

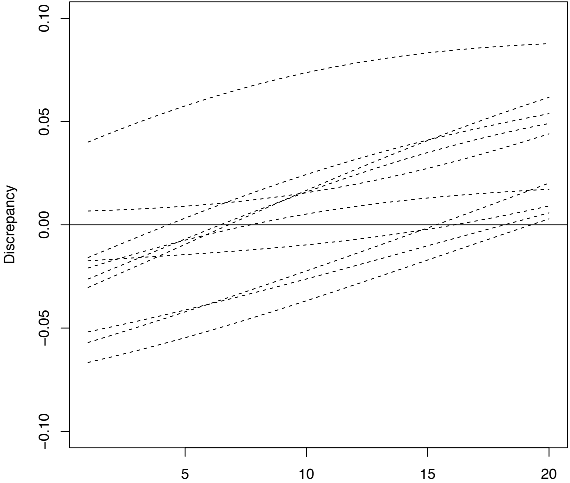
$$
\begin{aligned}
h_{d,t} &= (M'_d + \Delta_{d,t})h_{d,t-1}, \\
\Delta_{dt} &= \begin{pmatrix} -(\delta_{d1t}+\delta_{d2t}+\delta_{d3t}) & \delta_{d1t} & \delta_{d2t} & \delta_{d3t} \\ 0 & -(\delta_{d4t}+\delta_{d5t}) & \delta_{d4t} & \delta_{d5t} \\ 0 & 0 & -\delta_{d6t} & \delta_{d6t} \\ 0 & 0 & 0 & 0 \end{pmatrix}.
\end{aligned}
$$

We consider three scenarios. In each case, multivariate normal distribution assumed for $\delta_{d1t}, \delta_{d2t}, \ldots$
For example

$$
\begin{aligned}
E(\delta_{dit}) &= \beta_{0,di} + \beta_{1,di}t \\
Cov(\delta_{dit}, \delta_{d'i't'}) &= \sigma_{di}\sigma_{d'i'}\psi_{d',d'}\phi_{i,i'} \exp\left\{-\left(\frac{t-t'}{\omega}\right)^2\right\}.
\end{aligned}
$$

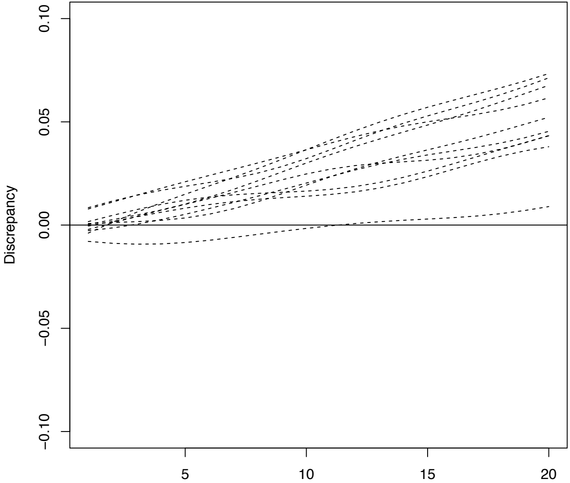
# Scenario 1: probability of death increases over time

Discrepancy $\delta_{dit}$ for a single transition probability over time



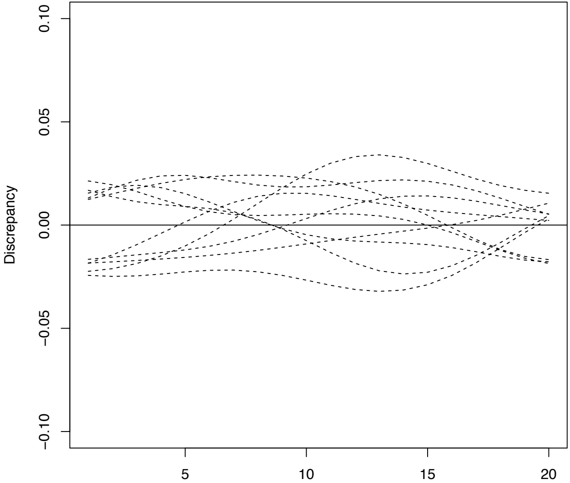
# Scenario 2: efficacy of treatment decreases over time

Discrepancy $\delta_{dit}$ for a single transition probability over time



# Scenario 3: ???

Discrepancy $\delta_{dit}$ for a single transition probability over time



# Measuring input importance for decision making

# Measuring input importance for decision making

- Consider model $Y = f(X_1, \ldots, X_r)$, with uncertain parameters $X_1, \ldots, X_r$
- Is $X_i$ 'important'; is it worth learning $X_i$ before making the decision?

# Measuring input importance for decision making

- Consider model $Y = f(X_1, \ldots, X_r)$, with uncertain parameters $X_1, \ldots, X_r$
- Is $X_i$ 'important'; is it worth learning $X_i$ before making the decision?
- Decision maker's utility for decision $d$ and 'true' model output $Y$ is $U(d, Y)$.

# Measuring input importance for decision making

- Consider model $Y = f(X_1, \ldots, X_r)$, with uncertain parameters $X_1, \ldots, X_r$
- Is $X_i$ 'important'; is it worth learning $X_i$ before making the decision?
- Decision maker's utility for decision $d$ and 'true' model output $Y$ is $U(d, Y)$.
- Utility of optimum decision based on no further information is

$$\max_d E_Y\{U(d, Y)\}$$

# Measuring input importance for decision making

- Consider model $Y = f(X_1, \ldots, X_r)$, with uncertain parameters $X_1, \ldots, X_r$
- Is $X_i$ 'important'; is it worth learning $X_i$ before making the decision?
- Decision maker's utility for decision $d$ and 'true' model output $Y$ is $U(d, Y)$.
- Utility of optimum decision based on no further information is

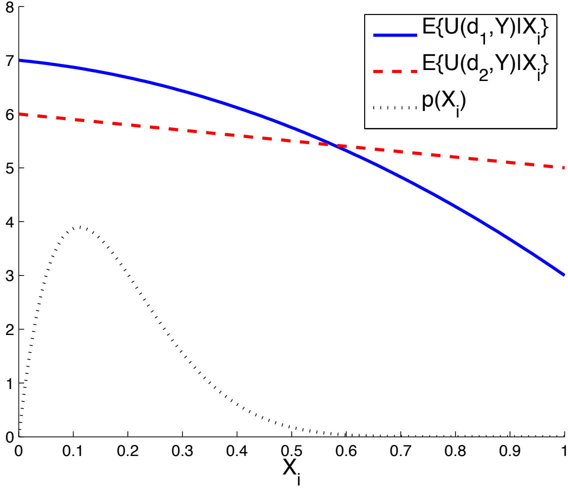
$$\max_d E_Y\{U(d, Y)\}$$

- Expected value of learning $X_i$ before making decision is

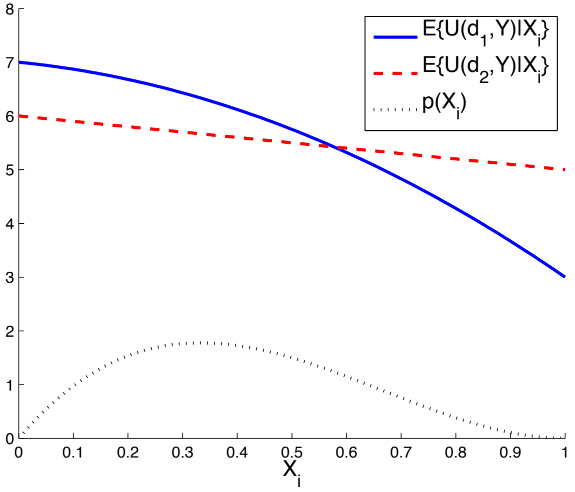
$$E_{X_i}\left[\max_d E\{U(d, Y)|X_i\}\right] - \max_d E_Y\{U(d, Y)\},$$

the **partial EVPI** of $X_i$ (**Expected Value of Perfect Information**)

# partial EVPI example 1



# partial EVPI example 2



# Quantifying the importance of discrepancy parameters

- We calculate partial EVPI jointly for all the discrepancy parameters in each scenario

# Quantifying the importance of discrepancy parameters

- We calculate partial EVPI jointly for all the discrepancy parameters in each scenario
- An ‘expected value of model improvement’...
- ...though improving the model may introduce further uncertain inputs

# Quantifying the importance of discrepancy parameters

- We calculate partial EVPI jointly for all the discrepancy parameters in each scenario
- An ‘expected value of model improvement’...
- ...though improving the model may introduce further uncertain inputs
- EVPIs estimated using Monte Carlo sampling (for expensive models can use emulators to get EVPIs, Oakley, 2009)

# Results

- Partial EVPI: expected value in £ per patient of learning parameters, before choosing treatment option for patient population.
- Utility function is the net benefit
  £$\lambda \times$ mean years of life - mean cost per patient
- For illustration, we choose $\lambda = 12000$

# Results

- Partial EVPI: expected value in £ per patient of learning parameters, before choosing treatment option for patient population.
- Utility function is the net benefit
  £$\lambda \times$ mean years of life - mean cost per patient
- For illustration, we choose $\lambda = 12000$

| Parameter | Partial EVPI | | | |
|---|---|---|---|---|
| | Base case | Scenario 1 | Scenario 2 | Scenario 3 |
| Transition probabilities | £0 | £0 | £0 | £1.17 |
| Relative risk | £169.91 | £193.09 | £64.63 | £164.55 |
| Costs | £194.41 | £201.72 | £65.17 | £167.53 |
| Discrepancy terms | - | £7.86 | £110.21 | £699.06 |
| Overall EVPI | £365.42 | £401.53 | £333.43 | £957.28 |

# Results

- Partial EVPI: expected value in £ per patient of learning parameters, before choosing treatment option for patient population.
- Utility function is the net benefit
  £$\lambda$ $\times$ mean years of life - mean cost per patient
- For illustration, we choose $\lambda = 12000$

| Parameter | Partial EVPI | | | |
|---|---|---|---|---|
| | Base case | Scenario 1 | Scenario 2 | Scenario 3 |
| Transition probabilities | £0 | £0 | £0 | £1.17 |
| Relative risk | £169.91 | £193.09 | £64.63 | £164.55 |
| Costs | £194.41 | £201.72 | £65.17 | £167.53 |
| Discrepancy terms | - | £7.86 | £110.21 | £699.06 |
| | | | | |
| Overall EVPI | £365.42 | £401.53 | £333.43 | £957.28 |

For larger $\lambda$ (20,000-30,000), all partial EVPIs negligible.

# Comments

- Making judgements about discrepancy will always be hard

# Comments

- Making judgements about discrepancy will always be hard (though $\delta \sim N(0, 0)$ is quite a popular judgement)

# Comments

- Making judgements about discrepancy will always be hard (though $\delta \sim N(0, 0)$ is quite a popular judgement)
- Alternatively, could consider how much discrepancy we can 'tolerate'

# Comments

- Making judgements about discrepancy will always be hard (though $\delta \sim N(0, 0)$ is quite a popular judgement)
- Alternatively, could consider how much discrepancy we can 'tolerate'
- Traditional uncertainty and sensitivity analyses of computer models have their role, but we need to pay more attention to reality!

# Comments

- Making judgements about discrepancy will always be hard (though $\delta \sim N(0, 0)$ is quite a popular judgement)
- Alternatively, could consider how much discrepancy we can 'tolerate'
- Traditional uncertainty and sensitivity analyses of computer models have their role, but we need to pay more attention to reality!

## And finally...a quick plug!

- Following on from UCM 2010...
- ...Uncertainty in Computer Models 2012
- Sheffield, July 2-4, 2012
- `www.mucm.ac.uk/UCM2012`