Aspects of Decision Making in Cost-effectiveness Modelling

Nathan Green (n.green@ucl.ac.uk)
(with thanks to Gianluca Baio, Chris Jackson, Nicky J. Welton, Mark Strong, Anna Heath)

24th November 2022

I

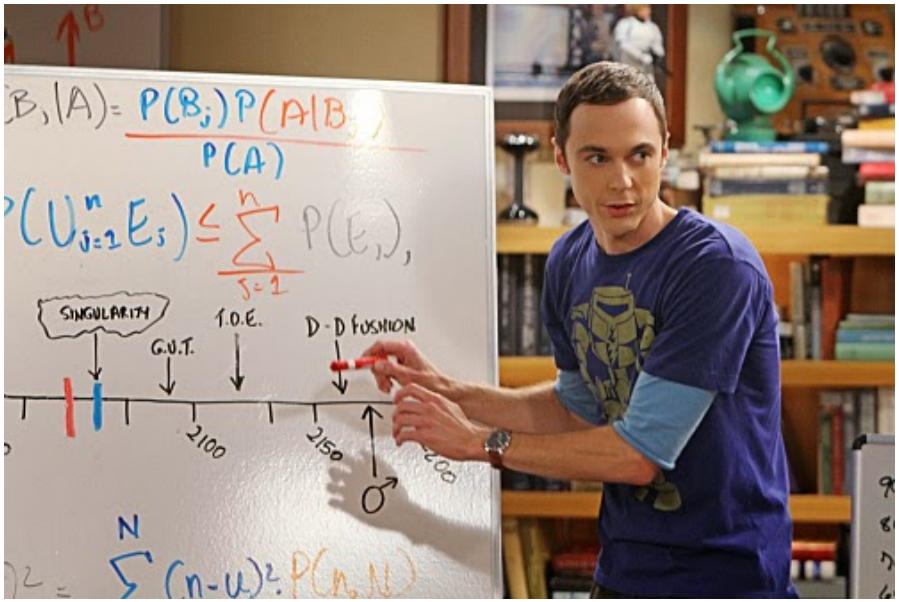
Preliminaries

University College London

- UCL was rated 2nd in the UK for research power in the Research Excellence Framework 2021
- UCL is ranked 8th in the 2022 QS World University Rankings
- The Department of Statistical Science has played a major role in the development of the subject ever since its foundation in 1911 as the Department of Applied Statistics

Objectives

Lectures



- Introduction to *Health economics modelling*
  - Decision trees
  - Markov models
- Introduction to sensitivity analyses
  - Deterministic
    - ★ One-way & multi-way
    - ★ Scenario
  - Probabilistic

Objectives

Computer practicals

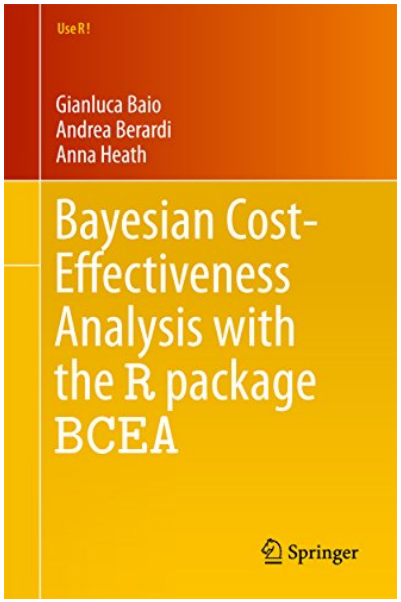
- Emphasis on practical examples
  - Decision tree and Markov models
  - using R programming language

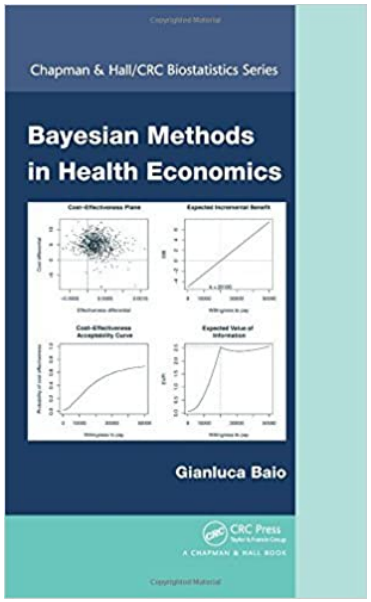
Timetable

- 0:00-1:00 Health Economics modelling lecture
- 1:00 - 1:45 Decision tree and Markov model practical
- BREAK
- 1:50 - 2:20 Sensitivity analysis
- 2:20-3:00 Sensitivity analysis practical

More Bayesian Health Economics...

Books





- This course is only a small part of an *annual week-long summer school*
  - usually in Florence, Italy
- Several books available
- Edition two of BCEA book in the pipeline and a Health Economic in R book close to being finished!

Lecture 2

Introduction to health economic evaluations

Summary

- Health economic evaluation
  - What is health economics?
  - Why do we need health economics?
- A framework for health economic evaluation
  - Statistical modelling
  - Economic modelling
  - Decision analysis
  - Uncertainty analysis
- Standard vs Bayesian HTA
  - Two-stage vs integrated approach
- Decision-making
  - Cost-effectiveness plane
  - ICER
  - EIB

References

Bayesian Methods in Health Economics, chapter 1.
Baio et al (2017). *Bayesian Cost-Effectiveness Analysis with the* `R` *package* `BCEA`

Health technology assessment (HTA)

Objective: Combine costs & benefits of a given intervention into a rational scheme for allocating resources

Health technology assessment (HTA)

Objective: Combine costs & benefits of a given intervention into a rational scheme for allocating resources

Statistical model

- Estimates relevant **population** parameters $\boldsymbol{\theta}$
- Varies with the type of available data (& statistical approach!)

Health technology assessment (HTA)

Objective: Combine costs & benefits of a given intervention into a rational scheme for allocating resources

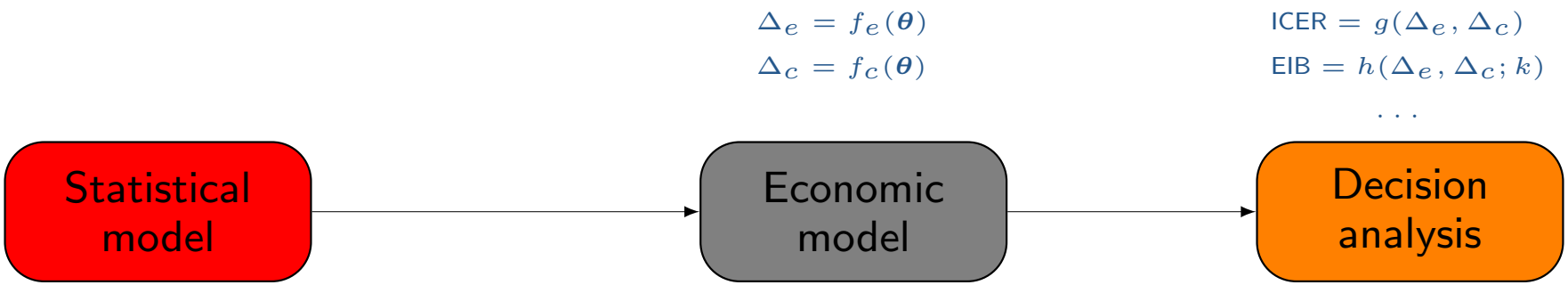
$$\Delta_e = f_e(\boldsymbol{\theta})$$
$$\Delta_c = f_c(\boldsymbol{\theta})$$

- Estimates relevant **population** parameters $\boldsymbol{\theta}$
- Varies with the type of available data (& statistical approach!)

- Combines the parameters to obtain a population average measure for costs and clinical benefits
- Varies with the type of available data & statistical model used

Health technology assessment (HTA)

Objective: Combine costs & benefits of a given intervention into a rational scheme for allocating resources



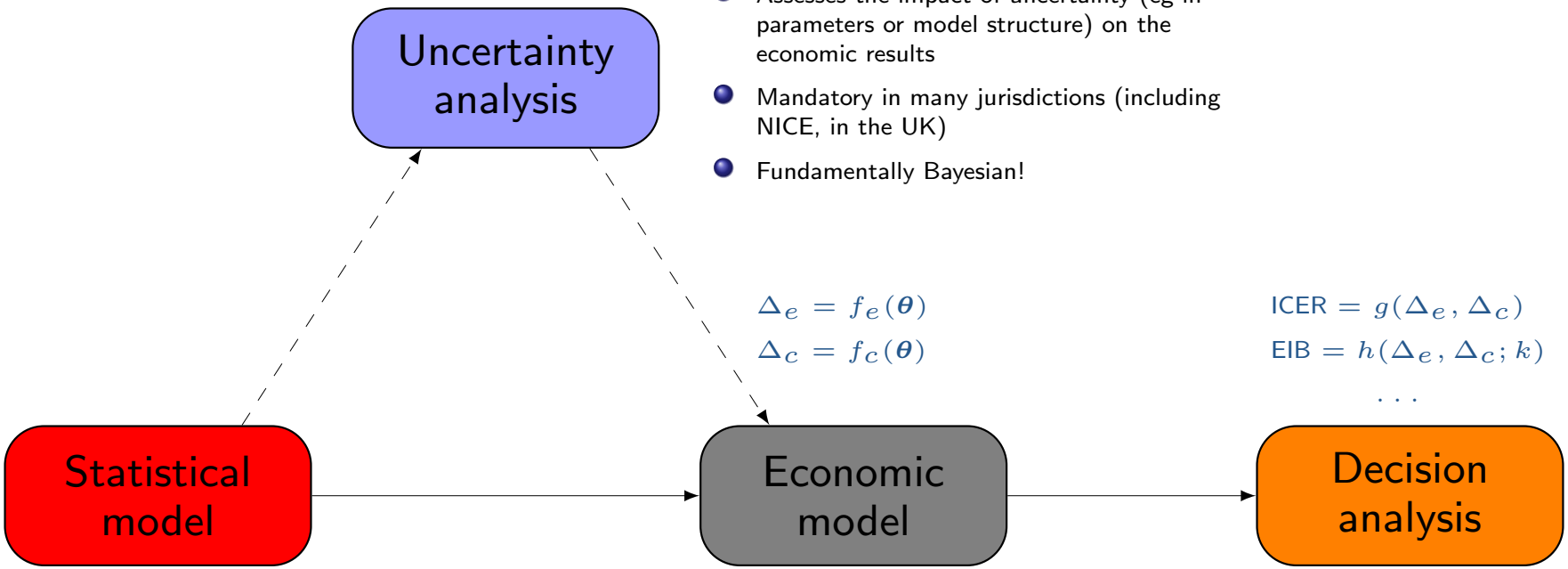
- Estimates relevant **population** parameters $\boldsymbol{\theta}$
- Varies with the type of available data (& statistical approach!)

- Combines the parameters to obtain a population average measure for costs and clinical benefits
- Varies with the type of available data & statistical model used

- Summarises the economic model by computing suitable measures of "cost-effectiveness"
- Dictates the best course of actions, given current evidence
- Standardised process

Health technology assessment (HTA)

Objective: Combine costs & benefits of a given intervention into a rational scheme for allocating resources



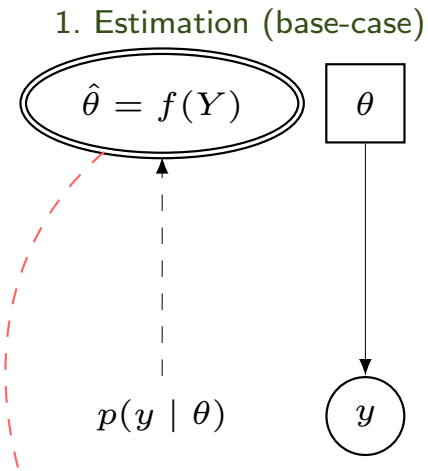
- Assesses the impact of uncertainty (eg in parameters or model structure) on the economic results
- Mandatory in many jurisdictions (including NICE, in the UK)
- Fundamentally Bayesian!

- Estimates relevant **population** parameters $\boldsymbol{\theta}$
- Varies with the type of available data (& statistical approach!)

- Combines the parameters to obtain a population average measure for costs and clinical benefits
- Varies with the type of available data & statistical model used

- Summarises the economic model by computing suitable measures of "cost-effectiveness"
- Dictates the best course of actions, given current evidence
- Standardised process

"Standard" approach to HTA — "Two-stage"



Uncertainty analysis

- Assesses the impact of uncertainty (eg in parameters or model structure) on the economic results
- Mandatory in many jurisdictions (including NICE, in the UK)
- Fundamentally Bayesian!

Statistical model

- Estimates relevant **population** parameters $\boldsymbol{\theta}$
- Varies with the type of available data (& statistical approach!)

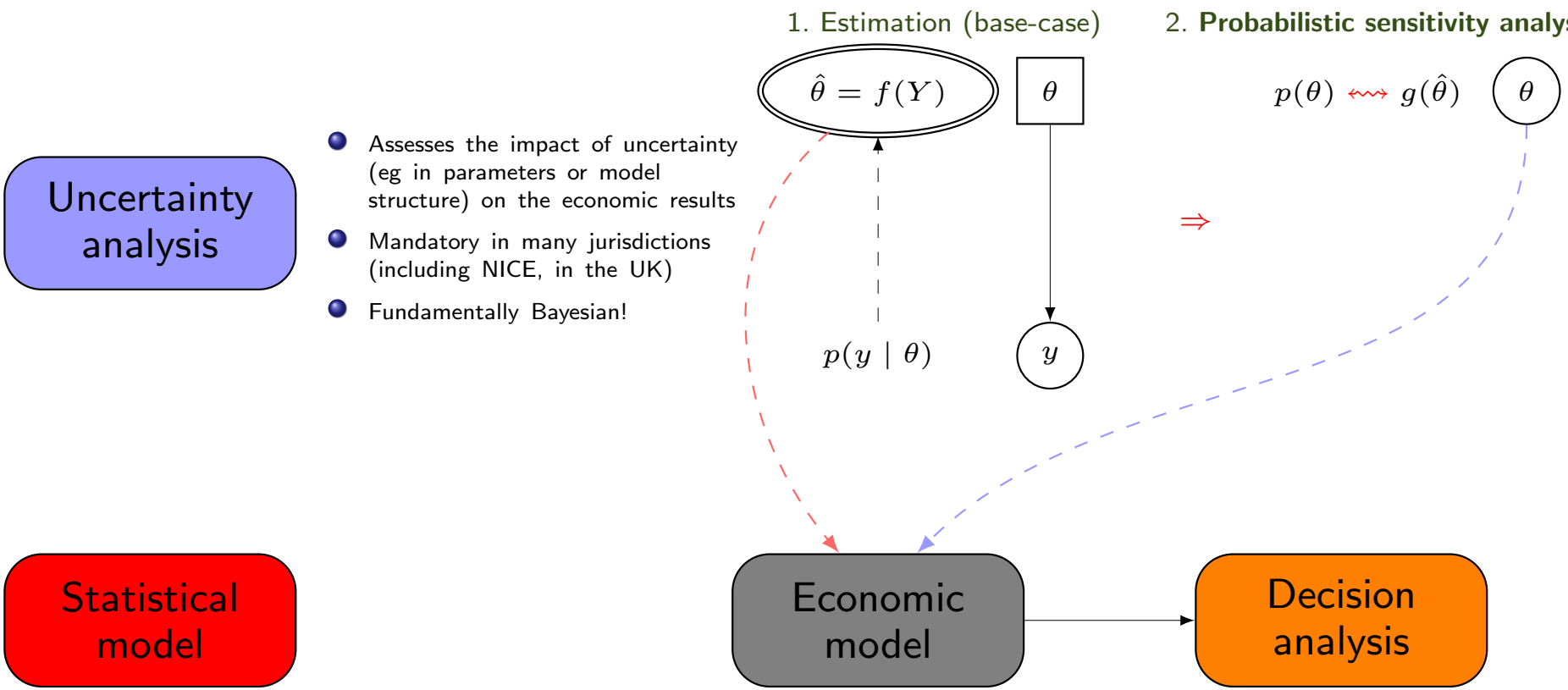
Economic model

- Combines the parameters to obtain a population average measure for costs and clinical benefits
- Varies with the type of available data & statistical model used

Decision analysis

- Summarises the economic model by computing suitable measures of "cost-effectiveness"
- Dictates the best course of actions, given current evidence
- Standardised process

"Standard" approach to HTA — "Two-stage"



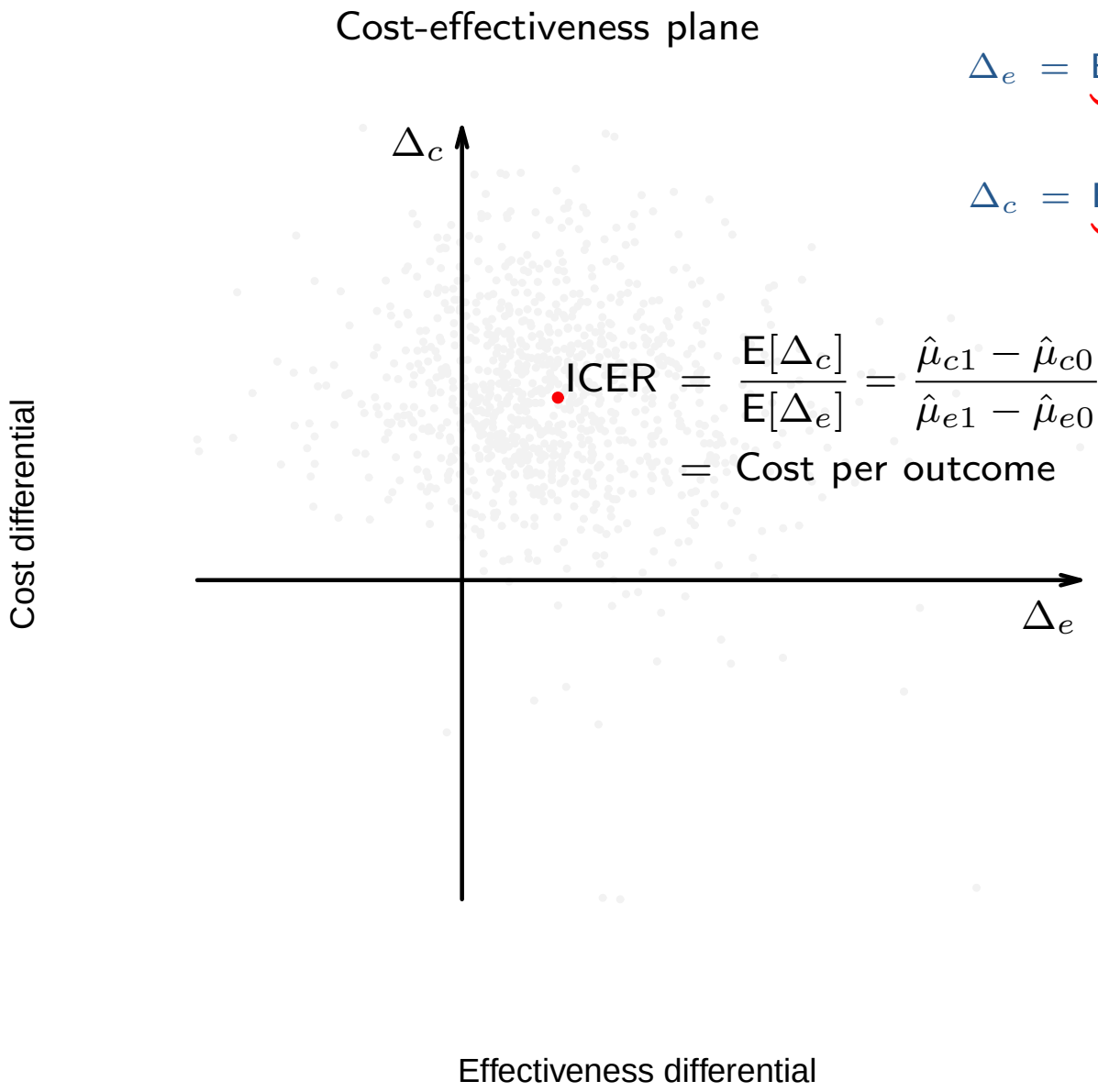
- Assesses the impact of uncertainty (eg in parameters or model structure) on the economic results
- Mandatory in many jurisdictions (including NICE, in the UK)
- Fundamentally Bayesian!

- Estimates relevant **population** parameters $\boldsymbol{\theta}$
- Varies with the type of available data (& statistical approach!)

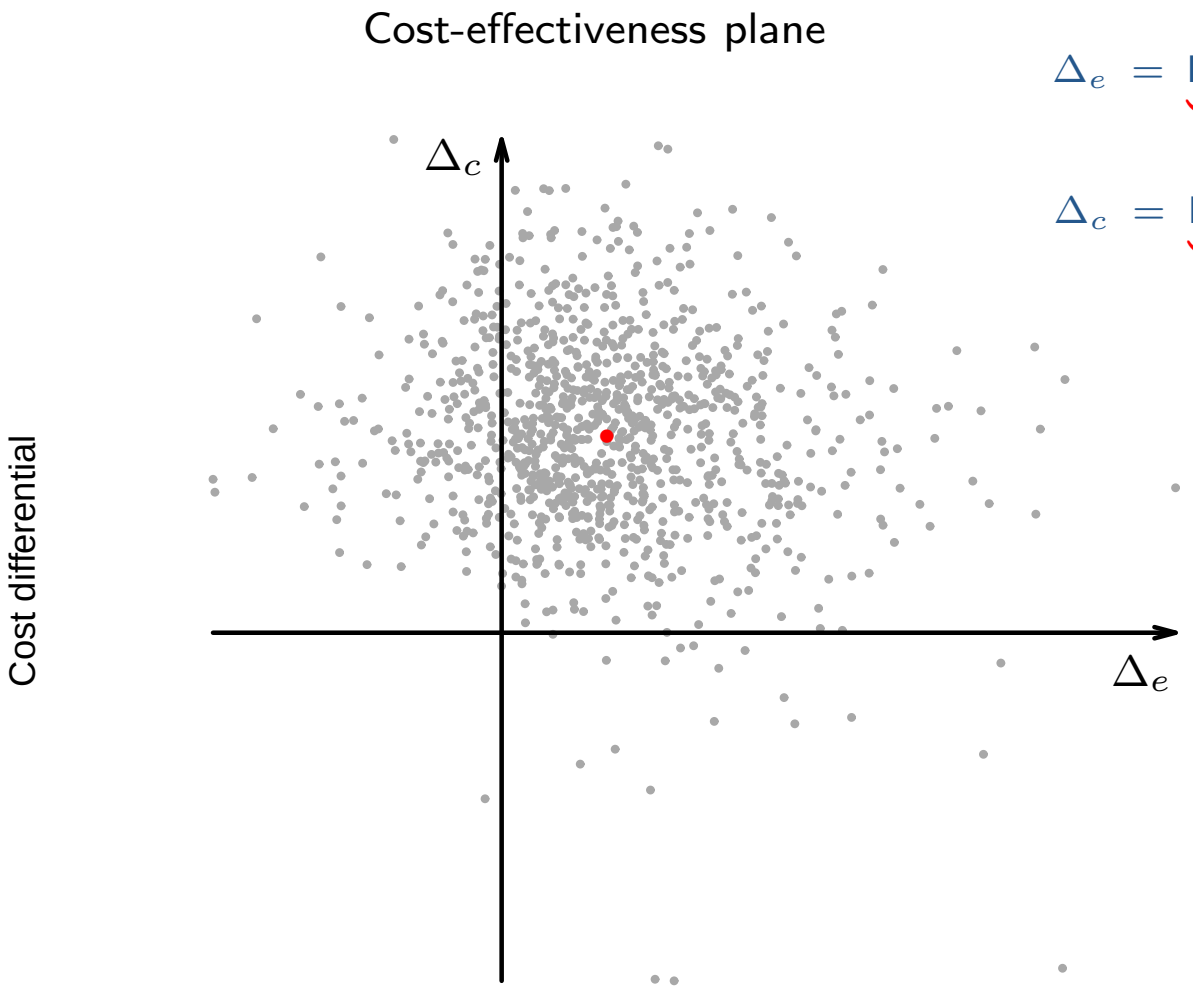
- Combines the parameters to obtain a population average measure for costs and clinical benefits
- Varies with the type of available data & statistical model used

- Summarises the economic model by computing suitable measures of "cost-effectiveness"
- Dictates the best course of actions, given current evidence
- Standardised process

2./3. Economic modelling+Decision analysis



2./3./4. ...+Uncertainty analysis*



$$\Delta_e = \underbrace{\mathrm{E}[e \mid \boldsymbol{\theta}_1]}_{\mu_{e1}} - \underbrace{\mathrm{E}[e \mid \boldsymbol{\theta}_0]}_{\mu_{e0}}$$

$$\Delta_c = \underbrace{\mathrm{E}[c \mid \boldsymbol{\theta}_1]}_{\mu_{c1}} - \underbrace{\mathrm{E}[c \mid \boldsymbol{\theta}_0]}_{\mu_{c0}}$$

*Induced by $g(\hat{\boldsymbol{\theta}}_0), g(\hat{\boldsymbol{\theta}}_1)$

What's wrong with this?...

- Potential correlation between costs & clinical benefits **[Individual Level + Aggregated Level Data]**
  - Strong positive correlation — effective treatments are innovative and result from intensive and lengthy research $\Rightarrow$ are associated with higher unit costs
  - Negative correlation — more effective treatments may reduce total care pathway costs e.g. by reducing hospitalisations, side effects, etc.
  - Because of the way in which standard models are set up, bootstrapping generally only approximates the underlying level of correlation — **MCMC does a better job!**

What's wrong with this?...

- Potential correlation between costs & clinical benefits [Individual Level + Aggregated Level Data]
  - Strong positive correlation — effective treatments are innovative and result from intensive and lengthy research $\Rightarrow$ are associated with higher unit costs
  - Negative correlation — more effective treatments may reduce total care pathway costs e.g. by reducing hospitalisations, side effects, etc.
  - Because of the way in which standard models are set up, bootstrapping generally only approximates the underlying level of correlation — **MCMC does a better job!**

- Joint/marginal normality not realistic [Mainly ILD]
  - Costs usually skewed and benefits may be bounded in $[0; 1]$
  - Can use transformation (e.g. logs) — but care is needed when back transforming to the natural scale
  - Should use more suitable models (e.g. Beta, Gamma or log-Normal) — **generally easier under a Bayesian framework**
  - Particularly relevant in presence of partially observed data — more on this later!

What's wrong with this?...

- Potential correlation between costs & clinical benefits [Individual Level + Aggregated Level Data]
  - Strong positive correlation — effective treatments are innovative and result from intensive and lengthy research $\Rightarrow$ are associated with higher unit costs
  - Negative correlation — more effective treatments may reduce total care pathway costs e.g. by reducing hospitalisations, side effects, etc.
  - Because of the way in which standard models are set up, bootstrapping generally only approximates the underlying level of correlation — **MCMC does a better job!**

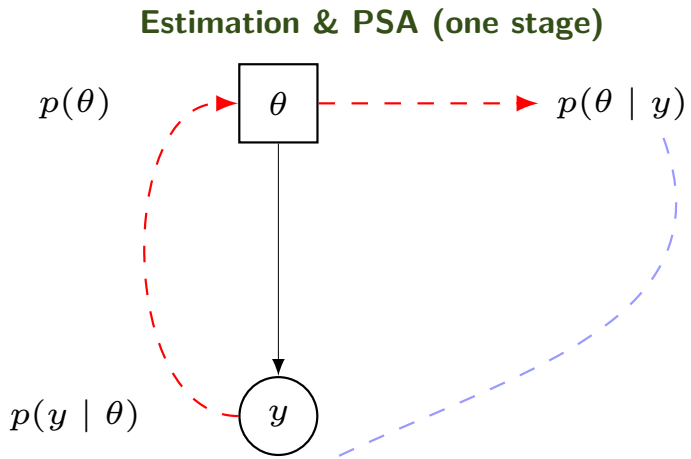
- Joint/marginal normality not realistic [Mainly ILD]
  - Costs usually skewed and benefits may be bounded in $[0; 1]$
  - Can use transformation (e.g. logs) — but care is needed when back transforming to the natural scale
  - Should use more suitable models (e.g. Beta, Gamma or log-Normal) — **generally easier under a Bayesian framework**
  - Particularly relevant in presence of partially observed data — more on this later!

- Particularly as the focus is on decision-making (rather than just inference), we need to use **all** available evidence to fully characterise current uncertainty on the model parameters and outcomes [Mainly ALD]
  - A Bayesian approach is helpful in combining different sources of information
  - **Propagating uncertainty is a fundamentally Bayesian operation!**

Bayesian approach to HTA — "Integrated"

Uncertainty analysis

- Assesses the impact of uncertainty (eg in parameters or model structure) on the economic results
- Mandatory in many jurisdictions (including NICE, in the UK)
- Fundamentally Bayesian!



Statistical model

- Estimates relevant **population** parameters $\boldsymbol{\theta}$
- Varies with the type of available data (& statistical approach!)

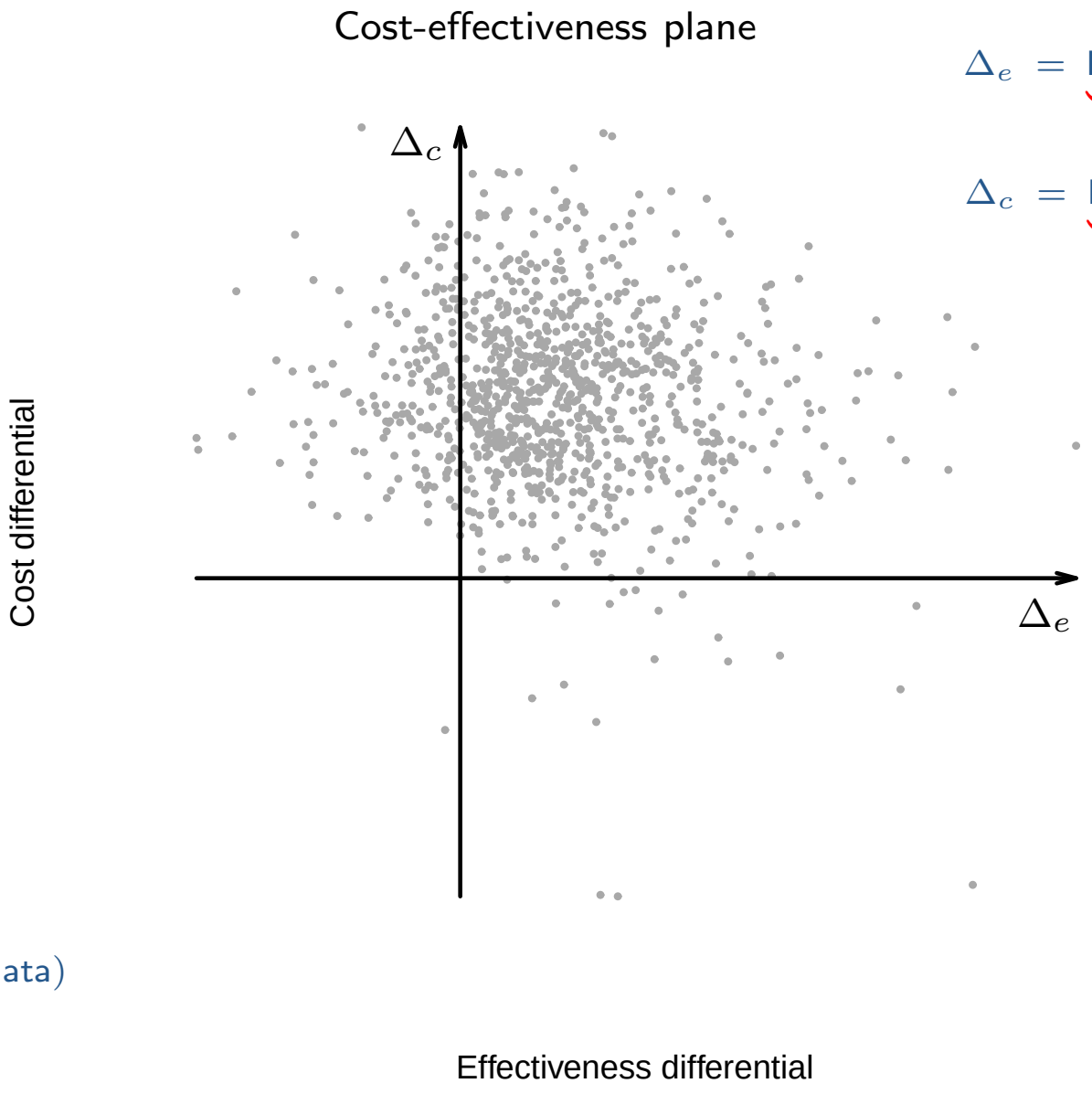
Economic model

- Combines the parameters to obtain a population average measure for costs and clinical benefits
- Varies with the type of available data & statistical model used

Decision analysis

- Summarises the economic model by computing suitable measures of "cost-effectiveness"
- Dictates the best course of actions, given current evidence
- Standardised process

2./4. Economic modelling+Uncertainty analysis*



$$\Delta_e = \underbrace{\mathsf{E}[e \mid \boldsymbol{\theta}_1]}_{\mu_{e1}} - \underbrace{\mathsf{E}[e \mid \boldsymbol{\theta}_0]}_{\mu_{e0}}$$

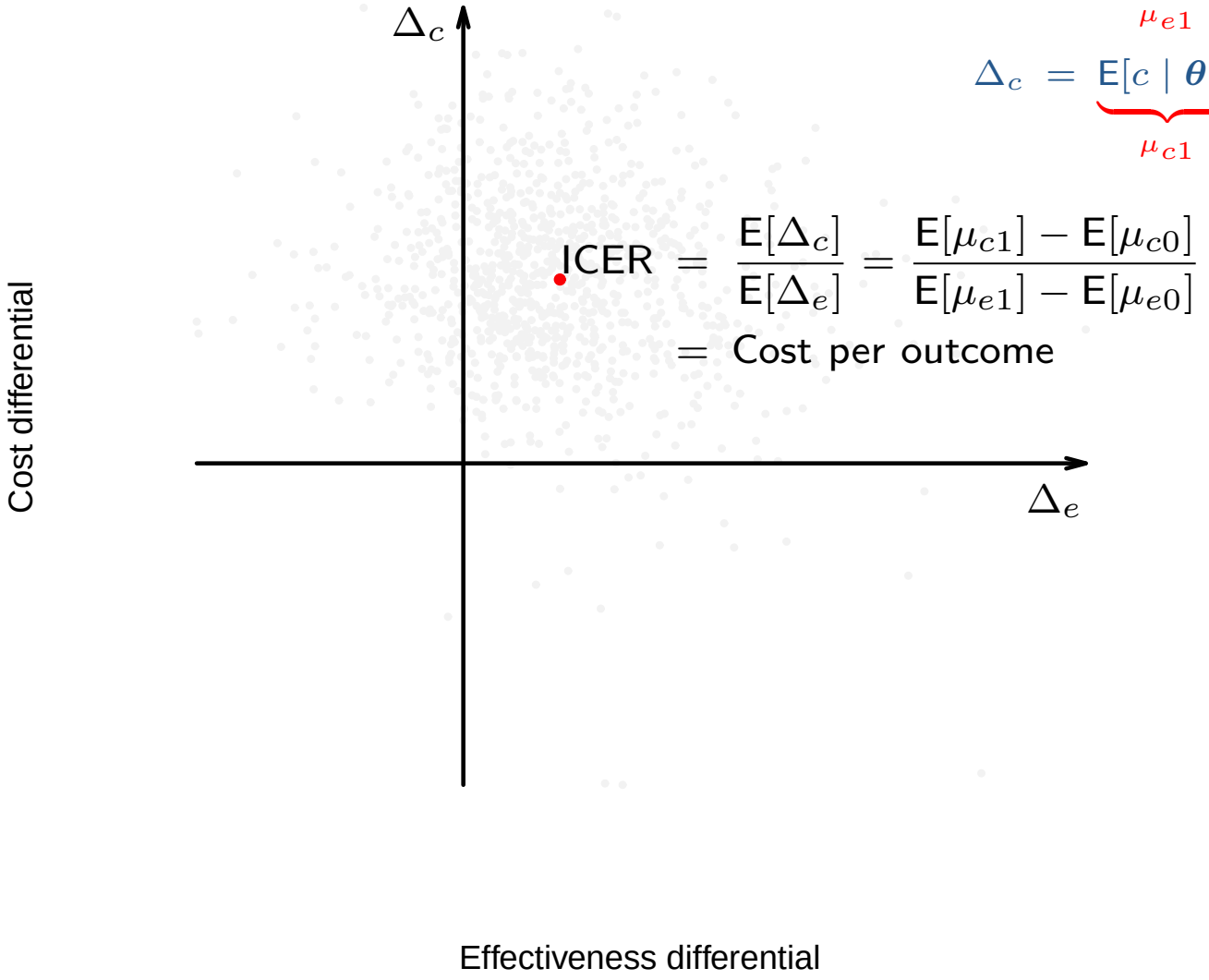
$$\Delta_c = \underbrace{\mathsf{E}[c \mid \boldsymbol{\theta}_1]}_{\mu_{c1}} - \underbrace{\mathsf{E}[c \mid \boldsymbol{\theta}_0]}_{\mu_{c0}}$$

*Induced by $p(\boldsymbol{\theta} \mid \text{data})$

Cost-effectiveness plane

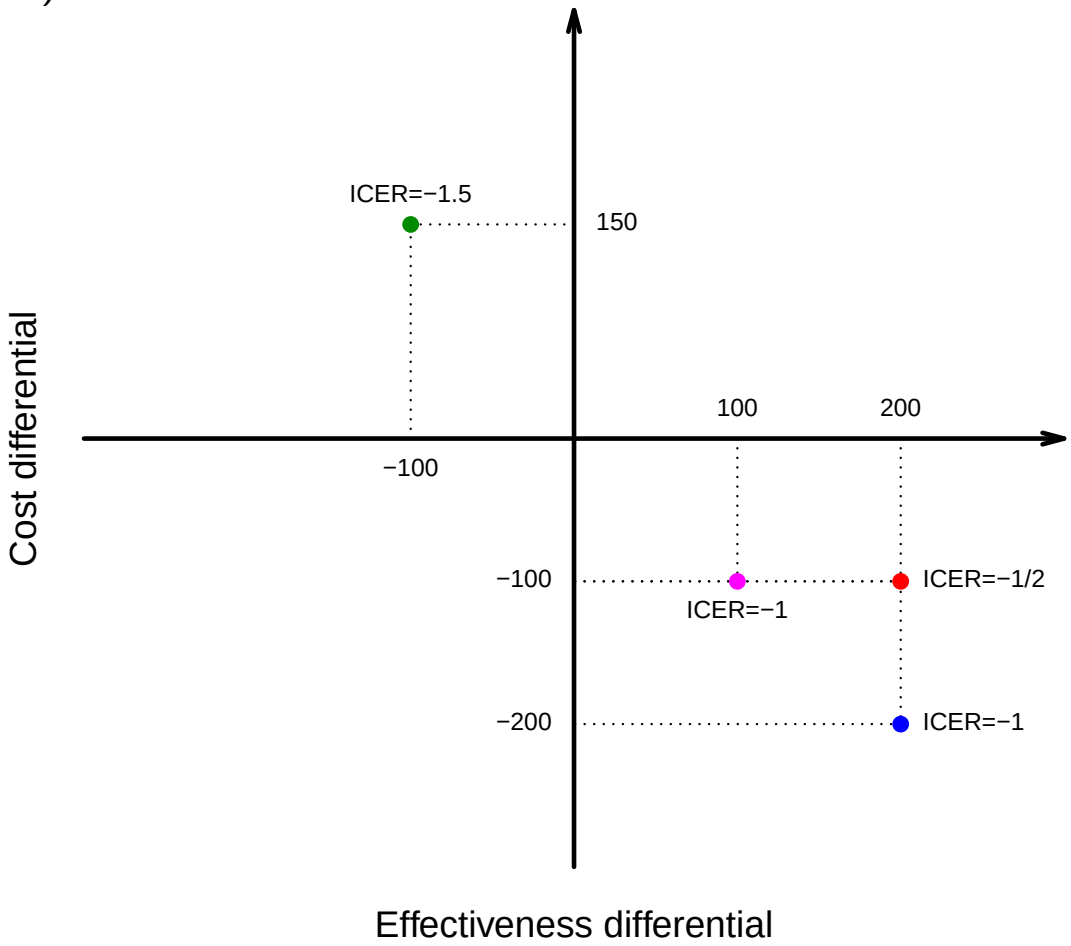
$$\Delta_e = \underbrace{\mathsf{E}[e \mid \boldsymbol{\theta}_1]}_{\mu_{e1}} - \underbrace{\mathsf{E}[e \mid \boldsymbol{\theta}_0]}_{\mu_{e0}}$$

$$\Delta_c = \underbrace{\mathsf{E}[c \mid \boldsymbol{\theta}_1]}_{\mu_{c1}} - \underbrace{\mathsf{E}[c \mid \boldsymbol{\theta}_0]}_{\mu_{c0}}$$



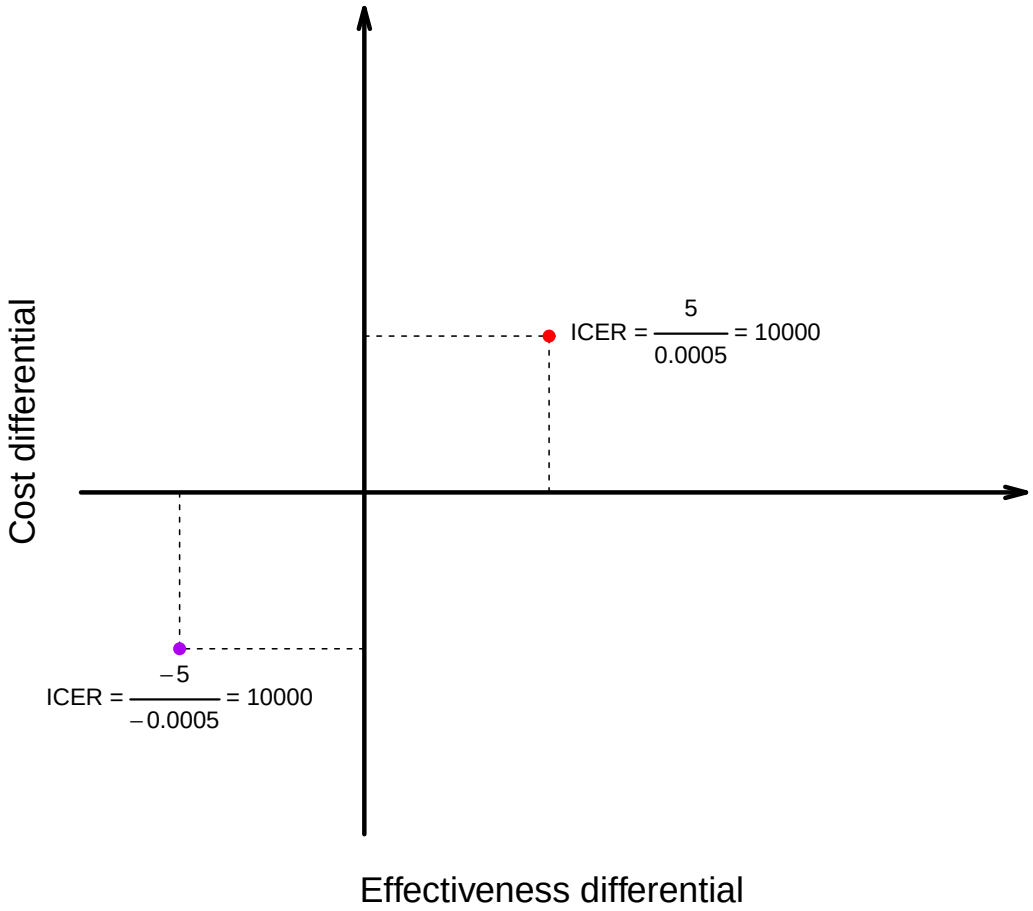
Decision-making based on the ICER

- ICER is not an ordered statistic
  - -200/200 better than -100/200 better than -100/100 in terms of decision, but ratios are -1, -1/2, -1
  - ICERs in the NW quadrant indicate an intervention that is **dominated** (+ costs/- effectiveness)



Decision-making based on the ICER

- Equivalent ICERs can mean very different things!
  - $(\mathrm{E}[\Delta_e], \mathrm{E}[\Delta_c]) = (0.0005, 5)$, indicates that the new treatment produces on average an increase in effectiveness of 0.0005 units at the cost of extra £10 000
  - $(\mathrm{E}[\Delta_e], \mathrm{E}[\Delta_c]) = (-0.0005, -5)$: new intervention less effective, but cheaper
  - In both cases, ICER $=$ £10 000



Decision-theoretic approach to HTA

- Analytic framework for decision-making in the face of uncertainty
- Considers a set of *prescriptive* axioms to ensure **rationality** in decision-making
- Identifies the best course of action given:
  - *Model specification*
  - *Current evidence*

Decision-theoretic approach to HTA

- Analytic framework for decision-making in the face of uncertainty
- Considers a set of *prescriptive* axioms to ensure **rationality** in decision-making
- Identifies the best course of action given:
  - *Model specification*
  - *Current evidence*

Process of rational decision-making

1. Describe uncertainty on *all* unknown quantities by means of a (possibly subjective) *probability distribution* $p(\boldsymbol{\omega}) = p(e, c \mid \boldsymbol{\theta})p(\boldsymbol{\theta})$
2. For each intervention t, outcomes $o = (e, c)$ are valued by means of a pre-specified *measure of utility* $u(e, c; t)$
3. Select as the most "cost-effective" the intervention that is associated with the *maximum expected utility* $\mathcal{U}^t = \mathsf{E}_{\boldsymbol{\omega}}[u(e, c; t)]$

Decision-theoretic approach to HTA

- Analytic framework for decision-making in the face of uncertainty
- Considers a set of *prescriptive* axioms to ensure **rationality** in decision-making
- Identifies the best course of action given:
  - *Model specification*
  - *Current evidence*

Process of rational decision-making

1. Describe uncertainty on *all* unknown quantities by means of a (possibly subjective) *probability distribution* $p(\boldsymbol{\omega}) = p(e, c \mid \boldsymbol{\theta})p(\boldsymbol{\theta})$
2. For each intervention t, outcomes $o = (e, c)$ are valued by means of a pre-specified *measure of utility* $u(e, c; t)$
3. Select as the most "cost-effective" the intervention that is associated with the *maximum expected utility* $\mathcal{U}^t = \mathsf{E}_{\boldsymbol{\omega}}[u(e, c; t)]$

- Typical utility function in HTA: *Monetary Net Benefit* $u(e, c; t) = nb_t = ke_t - c_t$
  - k is the "willingness to pay", i.e. the *cost per extra unit of effectiveness gained*
  - Fixed, *linear* form, which simplifies computations
  - Assumes decision-maker is *risk neutral*. Not necessarily true!

ICER vs EIB

- Under the MNB, the expected utility is

$$\begin{aligned}\mathcal{U}^t = \mathcal{NB}_t &= \mathsf{E}_{\boldsymbol{\omega}}[u(e,c;t)] \\ &= k\mathsf{E}_{\boldsymbol{\omega}}[e_t] - \mathsf{E}_{\boldsymbol{\omega}}[c_t] \\ &= k\mathsf{E}_{\boldsymbol{\theta}}[e \mid \boldsymbol{\theta}_t] - \mathsf{E}_{\boldsymbol{\theta}}[c \mid \boldsymbol{\theta}_t] = k\mathsf{E}[\mu_{et}] - \mathsf{E}[\mu_{ct}]\end{aligned}$$

NB: The expectation is taken with respect to $p(\boldsymbol{\omega})$ so $\mathcal{NB}_t$ is a pure number!

ICER vs EIB

- Under the MNB, the expected utility is

$$\begin{aligned}\mathcal{U}^t = \mathcal{NB}_t &= \mathsf{E}_{\boldsymbol{\omega}}[u(e,c;t)] \\ &= k\mathsf{E}_{\boldsymbol{\omega}}[e_t] - \mathsf{E}_{\boldsymbol{\omega}}[c_t] \\ &= k\mathsf{E}_{\boldsymbol{\theta}}[e \mid \boldsymbol{\theta}_t] - \mathsf{E}_{\boldsymbol{\theta}}[c \mid \boldsymbol{\theta}_t] = k\mathsf{E}[\mu_{et}] - \mathsf{E}[\mu_{ct}]\end{aligned}$$

  NB: The expectation is taken with respect to $p(\boldsymbol{\omega})$ so $\mathcal{NB}_t$ is a pure number!

- Assuming we are considering only two interventions $t = (0, 1)$, decision-making can be effected by looking at the *Expected Incremental Benefit*

$$\begin{aligned}\mathsf{EIB} &= \mathcal{NB}_1 - \mathcal{NB}_0 \\ &= (k\mathsf{E}[\mu_{e1}] - \mathsf{E}[\mu_{c1}]) - (k\mathsf{E}[\mu_{e0}] - \mathsf{E}[\mu_{c0}]) \\ &= k\mathsf{E}[\Delta_e] - \mathsf{E}[\Delta_c]\end{aligned}$$

ICER vs EIB

- Under the MNB, the expected utility is

$$\begin{aligned}\mathcal{U}^t = \mathcal{NB}_t &= \mathsf{E}_{\boldsymbol{\omega}}[u(e,c;t)] \\ &= k\mathsf{E}_{\boldsymbol{\omega}}[e_t] - \mathsf{E}_{\boldsymbol{\omega}}[c_t] \\ &= k\mathsf{E}_{\boldsymbol{\theta}}[e \mid \boldsymbol{\theta}_t] - \mathsf{E}_{\boldsymbol{\theta}}[c \mid \boldsymbol{\theta}_t] = k\mathsf{E}[\mu_{et}] - \mathsf{E}[\mu_{ct}]\end{aligned}$$

  NB: The expectation is taken with respect to $p(\boldsymbol{\omega})$ so $\mathcal{NB}_t$ is a pure number!

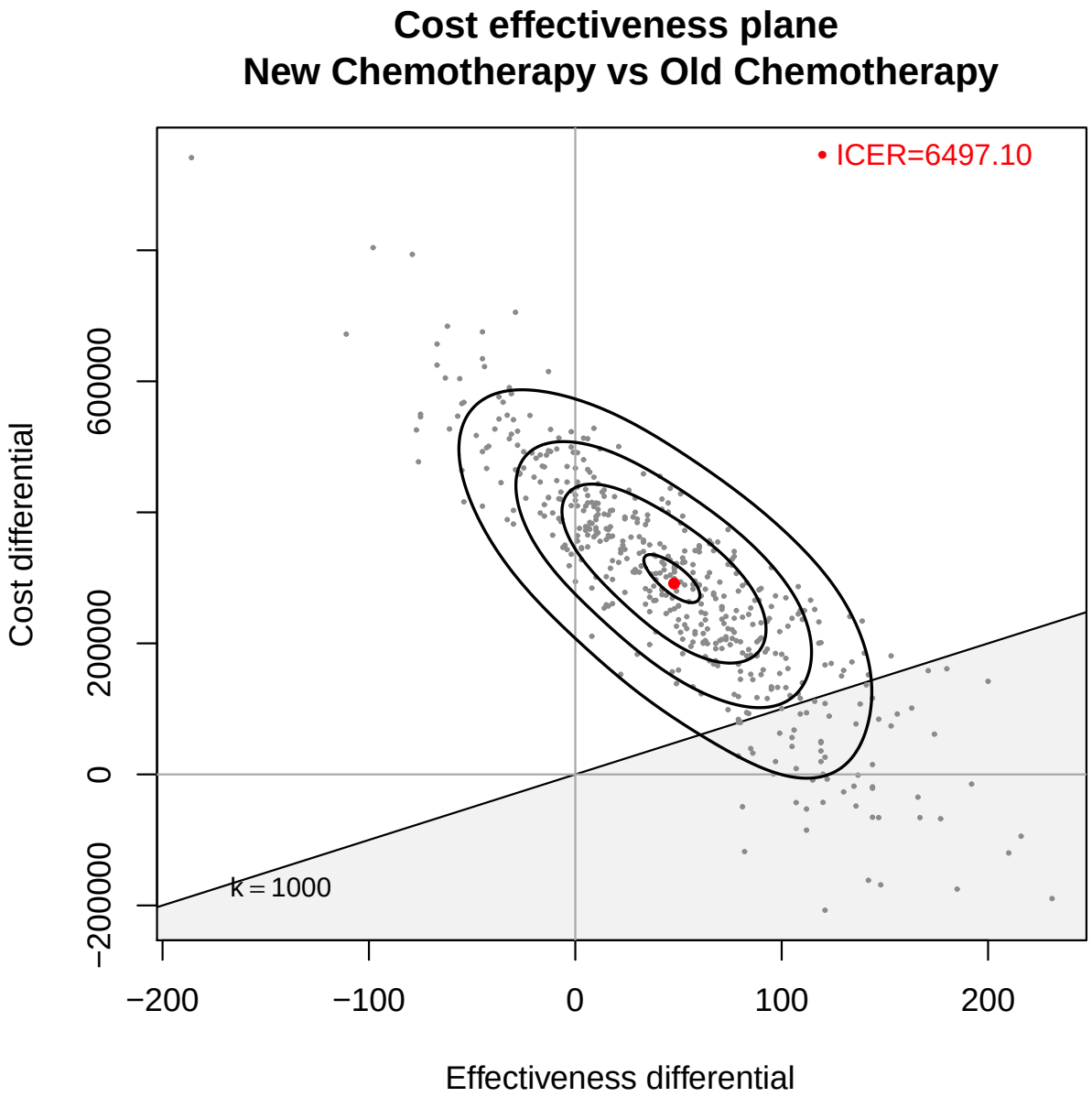
- Assuming we are considering only two interventions $t = (0, 1)$, decision-making can be effected by looking at the *Expected Incremental Benefit*

$$\begin{aligned}\mathsf{EIB} &= \mathcal{NB}_1 - \mathcal{NB}_0 \\ &= (k\mathsf{E}[\mu_{e1}] - \mathsf{E}[\mu_{c1}]) - (k\mathsf{E}[\mu_{e0}] - \mathsf{E}[\mu_{c0}]) \\ &= k\mathsf{E}[\Delta_e] - \mathsf{E}[\Delta_c]\end{aligned}$$

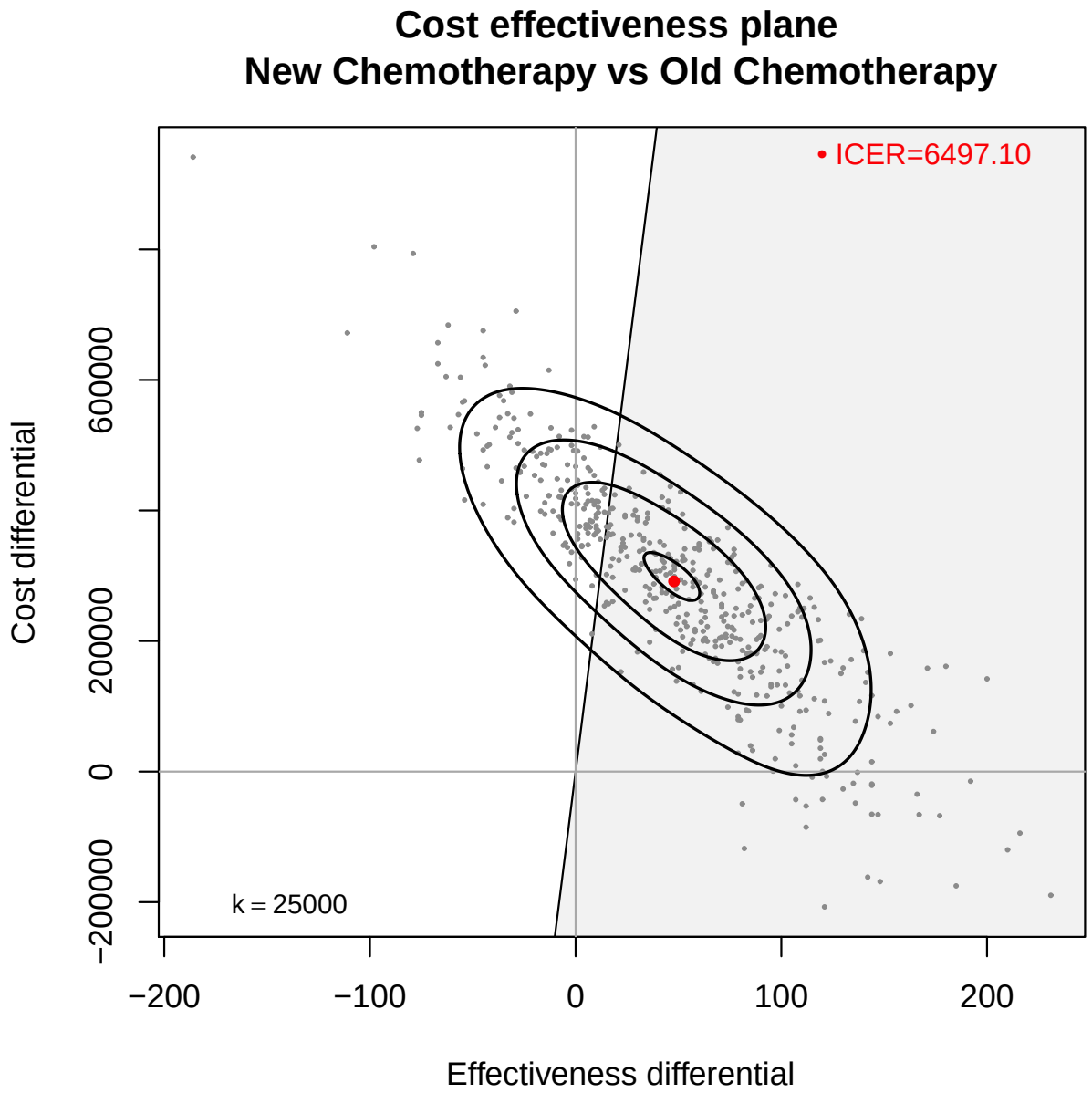
- The reference treatment $t = 1$ is more cost-effective than the comparator $t = 0$ if

$$\mathsf{EIB} > 0 \Rightarrow k > (<)\frac{\mathsf{E}[\Delta_c]}{\mathsf{E}[\Delta_e]} = \mathsf{ICER} \quad \text{if } \mathsf{E}[\Delta_e] > (<)0$$

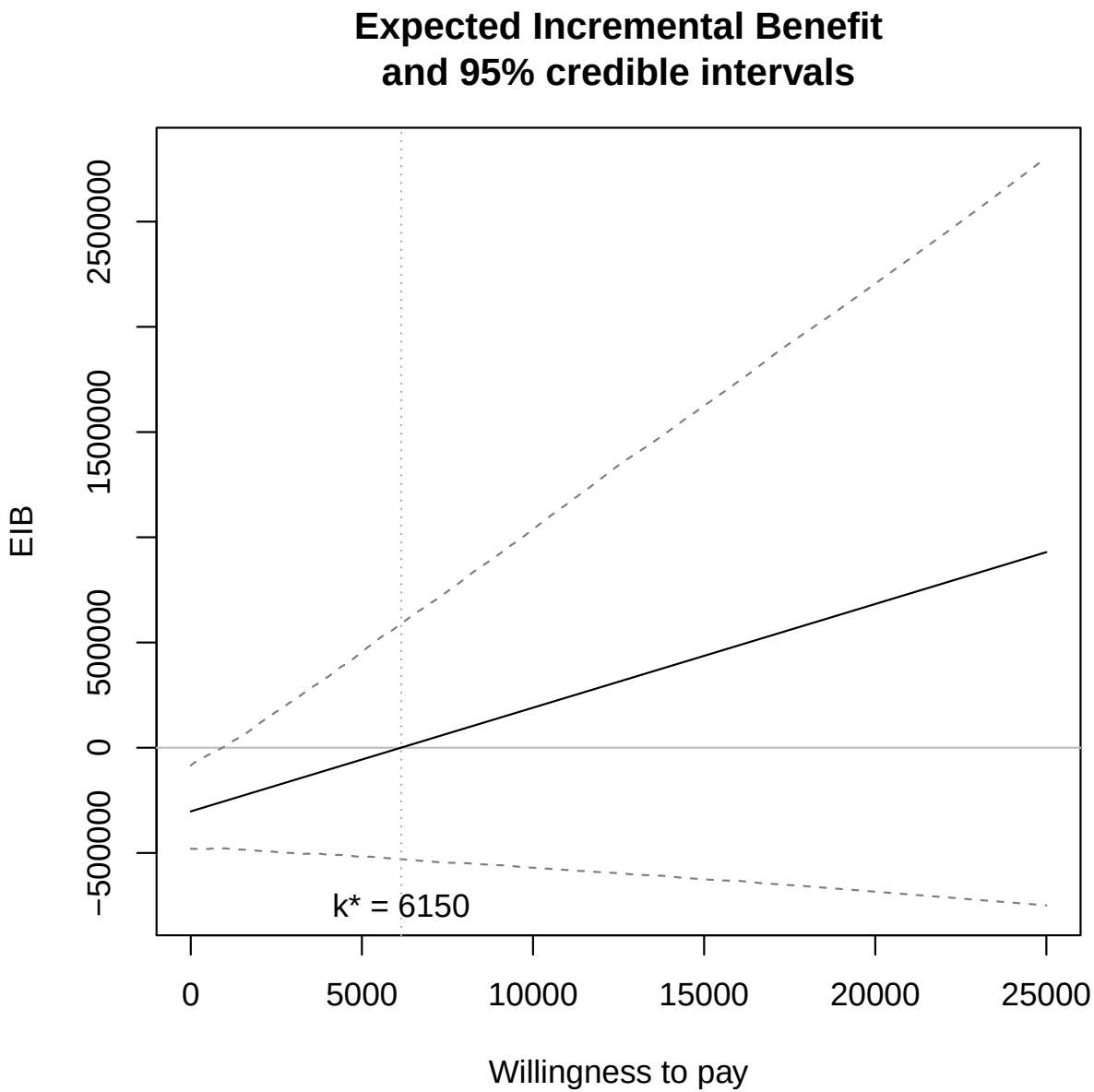
Cost-effectiveness plane vs EIB vs ICER



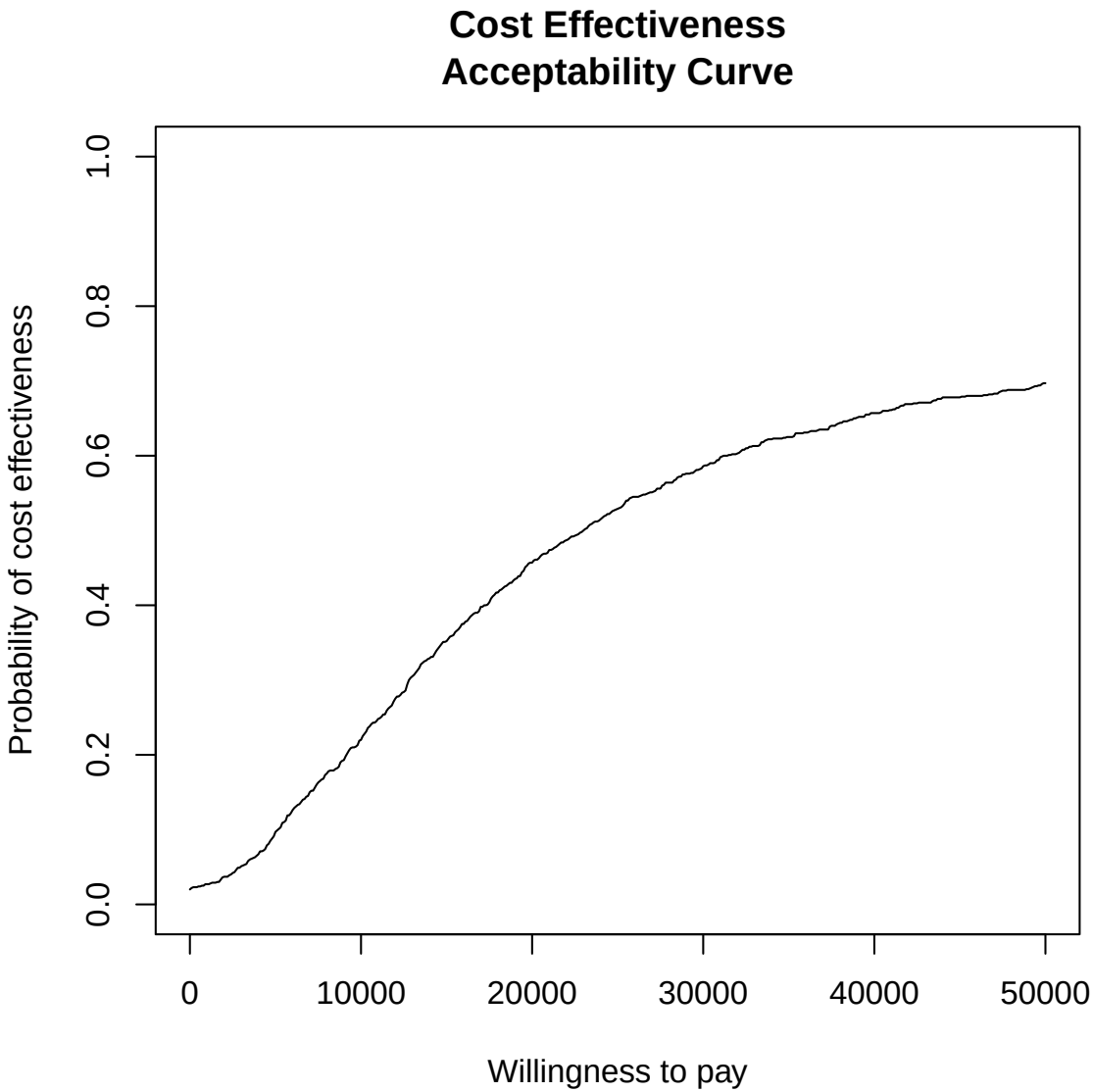
Cost-effectiveness plane vs EIB vs ICER



EIB vs Willingness to pay k



Cost effectiveness acceptability curve



Lecture 3

Decision trees

Model-based economic evaluation

- Trial-based economic evaluation
  - assess cost-effectiveness based on individual-level data on costs and effects
  - shorter-term, selected population
- Model-based economic evaluation
  - construct a model-based representation of disease/clinical history
  - State-transition (usually *Markov*) models for clinical histories very common
  - "populate" model with available + relevant data.
- Model-based evaluation typically used to estimate
  - *long-term* cost-effectiveness
  - in wider population

  by combining data from trials with other sources of evidence.

References

Bayesian Methods in Health Economics, chapter 5.4.
Welton et al (2012). *Evidence Synthesis for Decision Making in Healthcare*

Summary

- Decision tree model
  - Strengths and limitations
  - When to use and when not use
- How to calculate on a decision tree
  - 'Forward' method
  - 'Backward' method
- Examples

References

Decision Making in Health and Medicine, Weinstein et al, Cambridge University Press

Steps in constructing and analysing Decision Trees

A decision tree is a visual representation of a decision analysis:

- *Structure* the tree

Steps in constructing and analysing Decision Trees

A decision tree is a visual representation of a decision analysis:

- *Structure* the tree
- Estimate *probabilities*

Steps in constructing and analysing Decision Trees

A decision tree is a visual representation of a decision analysis:

- *Structure* the tree
- Estimate *probabilities*
- Estimate *payoffs* (assign values to costs and outcomes)

Steps in constructing and analysing Decision Trees

A decision tree is a visual representation of a decision analysis:

- *Structure* the tree
- Estimate *probabilities*
- Estimate *payoffs* (assign values to costs and outcomes)
- Analyse the tree

Steps in constructing and analysing Decision Trees

A decision tree is a visual representation of a decision analysis:

- *Structure* the tree
- Estimate *probabilities*
- Estimate *payoffs* (assign values to costs and outcomes)
- Analyse the tree
  - *Evaluate* the tree
  - Explore *uncertainty*

Structuring the Decision Tree

A decision tree is made up of nodes, branches and outcomes

- Nodes:

Structuring the Decision Tree

A decision tree is made up of nodes, branches and outcomes

- Nodes:
  - Decision node (square): Describes the problem. Deterministic choice.

Structuring the Decision Tree

A decision tree is made up of nodes, branches and outcomes

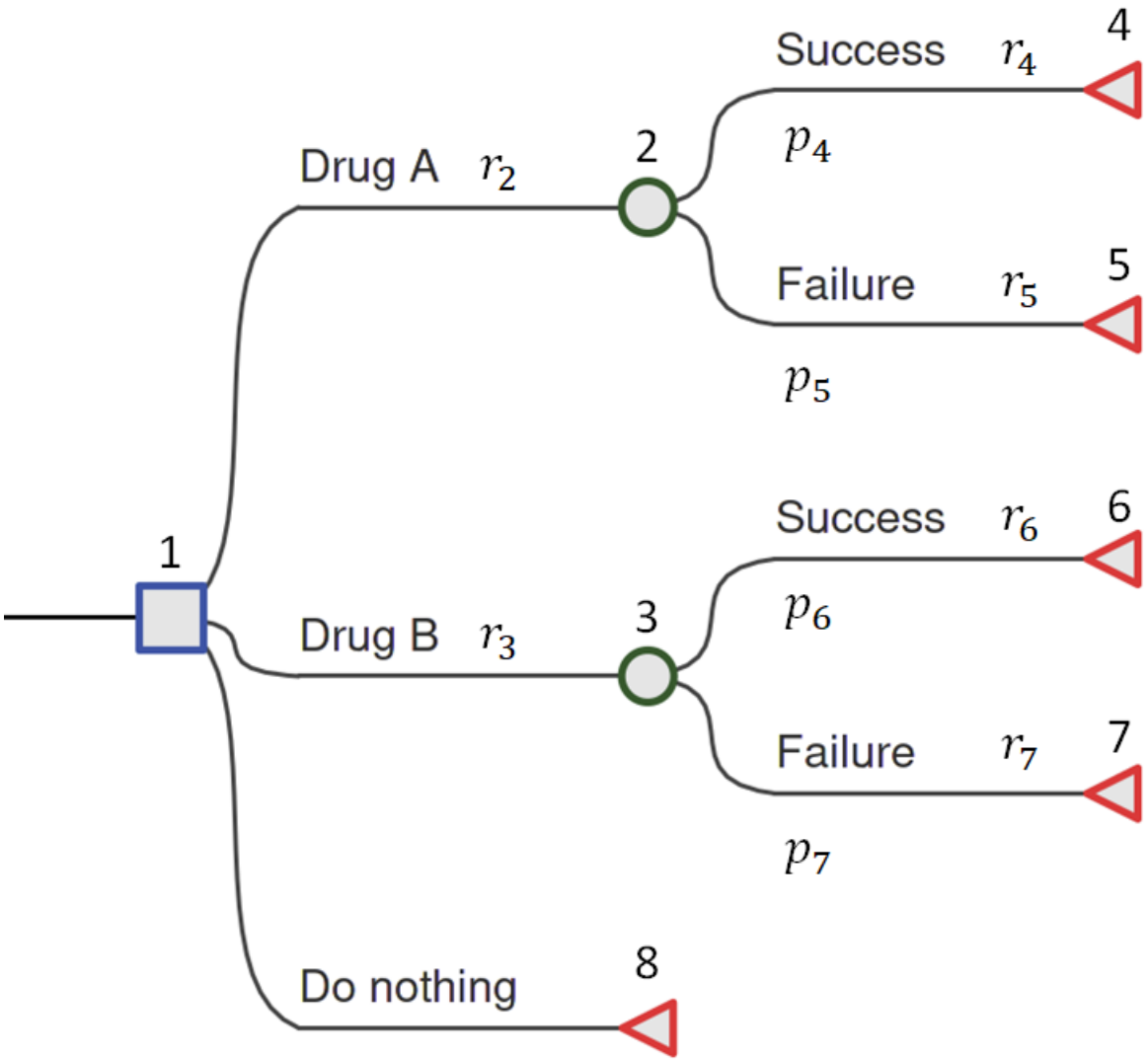
- Nodes:
  - Decision node (square): Describes the problem. Deterministic choice.
  - Chance node (circle): Represents the point at which several possible events can occur.

Structuring the Decision Tree

A decision tree is made up of nodes, branches and outcomes

- Nodes:
  - Decision node (square): Describes the problem. Deterministic choice.
  - Chance node (circle): Represents the point at which several possible events can occur.
  - Terminal node (triangle): Represents the end of a tree with a payoff attached.

A hypothetical example is of a comparison between two drugs, 'Drug A' and 'Drug B'. Each drug has different costs associated with them and different performance in terms of the chance that the drug is successful at treating the patient.



Structuring the Decision Tree

- Branches issuing from a chance node represent possible events patients may experience at that point in the tree.
- Branch probabilities represent the likelihood of each event.
- The sequence of chance nodes from left to right usually follows the sequence of events.
- The events stemming from a chance node must be *mutually exclusive* and probabilities should sum to 1.

Advantages of Decision Trees

- They enable the economic question to be structured in a *meaningful* and *visual* manner.

Advantages of Decision Trees

- They enable the economic question to be structured in a *meaningful* and *visual* manner.
- They allow data informing the model parameters to be assimilated and, where appropriate, synthesised.

Advantages of Decision Trees

- They enable the economic question to be structured in a *meaningful* and *visual* manner.
- They allow data informing the model parameters to be assimilated and, where appropriate, synthesised.
- They are relatively simple to undertake and suitable for:
  - Diseases that occur only once.
  - Decisions about acute care.
  - Decisions with short time frames.

Limitations of Decision Trees

- They do not explicitly account for passage of time:
  - Passage of time accounted for by outcome measure.
  - Limited ability to account for long term outcomes.

Limitations of Decision Trees

- They do not explicitly account for passage of time:
  - Passage of time accounted for by outcome measure.
  - Limited ability to account for long term outcomes.
- Possible to add branches but results in a complex model.
- Other modelling techniques can handle repeated events better.
- Structure of tree only allows for one-way progression of patient through model: Not movement back and forth between states.
- Decision trees can still be useful as a sub-model.

Tuberculosis—diagnosis, management, prevention, and control: summary of updated NICE guidance

BMJ 2016 ; 352 doi: https://doi.org/10.1136/bmj.h6747 (Published 13 January 2016)
Cite this as: *BMJ* 2016;352:h6747

Lucy Elizabeth Hoppe, technical analyst (clinical) [1], *Rachel Kettle, technical advisor (public health)* [2], *Michael Eisenhut, consultant paediatrician and member of the Guideline Development Group* [3], *Ibrahim Abubakar, professor of infectious disease epidemiology and co-chair of the Guideline Development Group* [4] *on behalf of the Guideline Development Group*

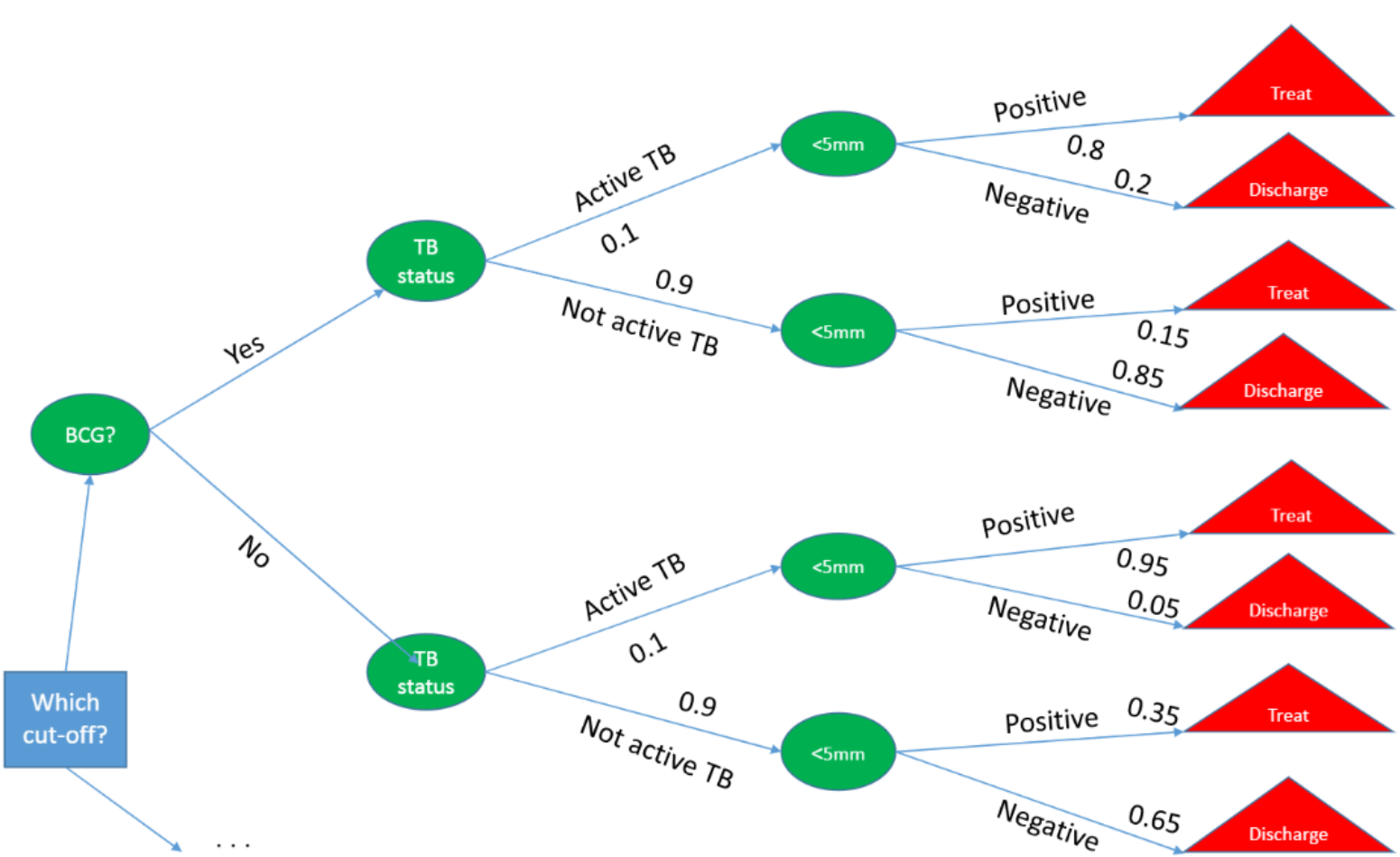
Author affiliations

Correspondence to: L E Hoppe lucy.hoppe@nice.org.uk

What you need to know

- Undertake tuberculosis (TB) testing in close contacts of people with pulmonary or laryngeal TB, people who are immunocompromised and at high risk of TB, and new entrants from high incidence countries who present to healthcare services

Simple decision tree example: New TST guidelines



cost	utility
£60	50
£10	-100
£60	-10
£10	0
£70	50
£20	-100
£70	-10
£20	0

Calculations on a decision tree

- Conditional probabilities two or more events
- Probability of both A and B occurring, divided by the probability of B occurring
- The joint probability of A and B measures the probability that A and B occur together at the same moment.
- The marginal probability of A is the individual probability of A, ignoring any value of event B.

Probabilities of moving along branches

- Complete set of all possible outcomes is set of terminal nodes on the RHS

Probabilities of moving along branches

- Complete set of all possible outcomes is set of terminal nodes on the RHS
- By chaining the conditional probabilities for a given pathway we obtain the total or *joint* probability of reaching a terminal node. Each pathway through the tree is a mutually exclusive sequence of events. Consider one such pathway through the tree with the following n nodes.

$$x_0 \rightarrow x_{[1]} \rightarrow x_{[2]} \rightarrow \cdots \rightarrow x_{[n]}$$

where

  - x_0 is the root node (often the decision node)
  - $x_{[n]}$ is the terminal node and the square brackets indicate that this is the ith ordered nodes in the sequence.

Probabilities of moving along branches

- Complete set of all possible outcomes is set of terminal nodes on the RHS
- By chaining the conditional probabilities for a given pathway we obtain the total or *joint* probability of reaching a terminal node. Each pathway through the tree is a mutually exclusive sequence of events. Consider one such pathway through the tree with the following n nodes.

$$x_0 \rightarrow x_{[1]} \rightarrow x_{[2]} \rightarrow \cdots \rightarrow x_{[n]}$$

where

  - x_0 is the root node (often the decision node)
  - $x_{[n]}$ is the terminal node and the square brackets indicate that this is the ith ordered nodes in the sequence.

- By the product rule joint probability of this path is

$$p(x_{[1]}, x_{[2]}, \ldots, x_{[n]}) = p(x_{[2]} \mid x_{[1]})p(x_{[3]} \mid x_{[2]}) \cdots p(x_{[n]} \mid x_{[n-1]}) = \prod_{i}^{[n-1]} p(x_{[i+1]} \mid x_{[i]})$$

Probabilities of moving along branches

- Complete set of all possible outcomes is set of terminal nodes on the RHS
- By chaining the conditional probabilities for a given pathway we obtain the total or *joint* probability of reaching a terminal node. Each pathway through the tree is a mutually exclusive sequence of events. Consider one such pathway through the tree with the following n nodes.

$$x_0 \rightarrow x_{[1]} \rightarrow x_{[2]} \rightarrow \cdots \rightarrow x_{[n]}$$

where

  - x_0 is the root node (often the decision node)
  - $x_{[n]}$ is the terminal node and the square brackets indicate that this is the ith ordered nodes in the sequence.

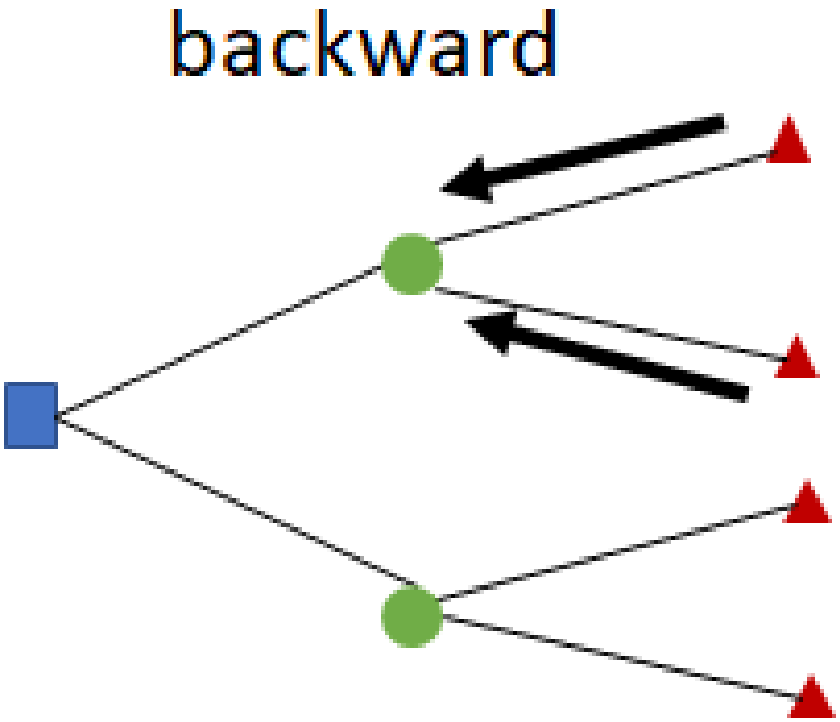
- By the product rule joint probability of this path is

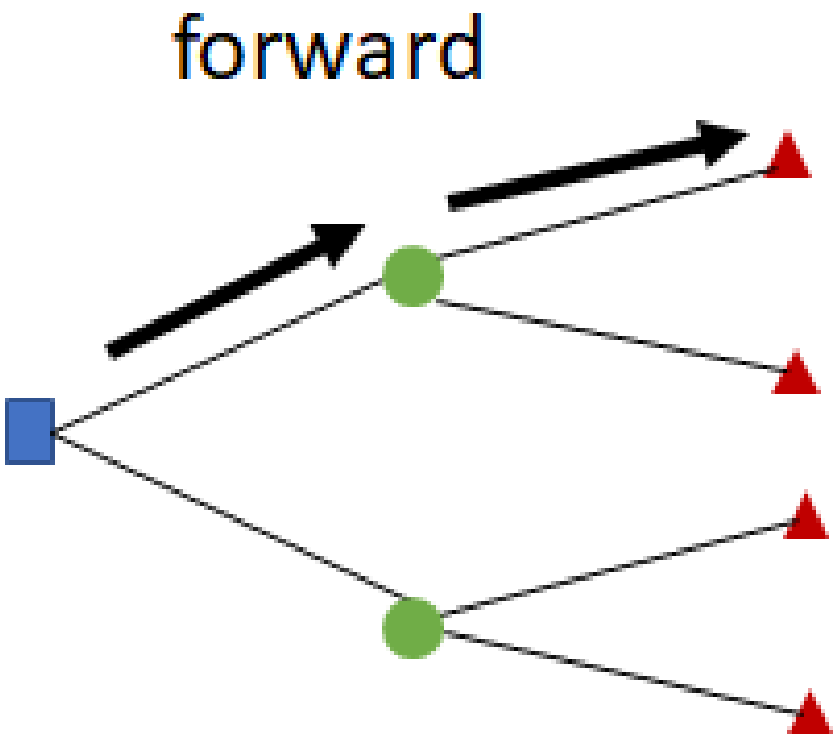
$$p(x_{[1]}, x_{[2]}, \ldots, x_{[n]}) = p(x_{[2]} \mid x_{[1]})p(x_{[3]} \mid x_{[2]}) \cdots p(x_{[n]} \mid x_{[n-1]}) = \prod_{i}^{[n-1]} p(x_{[i+1]} \mid x_{[i]})$$

- For the above pathway, the corresponding cost and effects are $c_{[1]}, c_{[2]}, \ldots, c_{[n]}$ and $e_{[1]}, e_{[2]}, \ldots, e_{[n]}$.

How to calculate expected values

There are two alternative approaches used for calculating the expected cost and effectiveness on a decision tree.





"Backward" computation

- Weighted average of the total values of the child nodes of a parent chance node
  - Weights are the probability of traversing each branch to the child nodes.

"Backward" computation

- Weighted average of the total values of the child nodes of a parent chance node
  - Weights are the probability of traversing each branch to the child nodes.
- Act of stepping backwards in this way through the tree is called *folding back* (or *rolling back*).

"Backward" computation

- Weighted average of the total values of the child nodes of a parent chance node
  - Weights are the probability of traversing each branch to the child nodes.
- Act of stepping backwards in this way through the tree is called *folding back* (or *rolling back*).
- So called because the children are folded up or collapsed, so that the chance node is now represented by its expected value.

"Backward" computation

- Weighted average of the total values of the child nodes of a parent chance node
  - Weights are the probability of traversing each branch to the child nodes.
- Act of stepping backwards in this way through the tree is called *folding back* (or *rolling back*).
- So called because the children are folded up or collapsed, so that the chance node is now represented by its expected value.
- Starting at the right-most terminal nodes the expected values at each chance node are calculate in turn and the tree can thus be folded back all the way to the decision node

"Backward" computation

- Weighted average of the total values of the child nodes of a parent chance node
  - Weights are the probability of traversing each branch to the child nodes.
- Act of stepping backwards in this way through the tree is called *folding back* (or *rolling back*).
- So called because the children are folded up or collapsed, so that the chance node is now represented by its expected value.
- Starting at the right-most terminal nodes the expected values at each chance node are calculate in turn and the tree can thus be folded back all the way to the decision node
- Example of something called a *recursive* function which is a function that calls itself during its execution.
  - A well-known example is when calculating the Fibonacci series.
- This approach is part of a whole field of stochastic optimisation in applied probability called Markov decision process (MDP).

"Backward" computation

- Recall the conditional probabilities $p_{ij} = p(x_j \mid x_i)$, then the expected value is

$$\mathbb{E}[V_i] = \begin{cases} r_i & \text{if} \quad i \in \mathcal{S}_{term} \\ r_i + \sum p_{ij}\mathbb{E}[V_j] & \text{otherwise} \end{cases}$$

where

  - $\mathbb{E}$ is the weighted average of the values, V is the random variable total node value, e.g. cost or QALYs
  - r_i is the (unit) value at node i
  - $\mathcal{S}_{term}$ is the set of all terminal nodes. Note that at a terminal node the expected value is simply the value at that node.

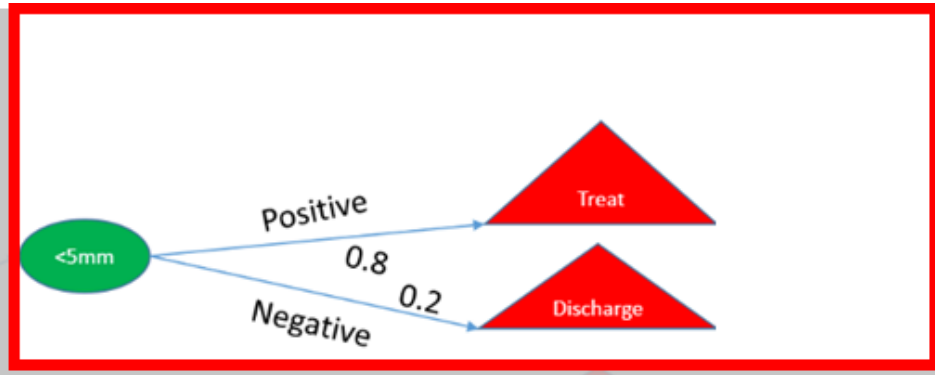
"Backward" computation

- Recall the conditional probabilities $p_{ij} = p(x_j \mid x_i)$, then the expected value is

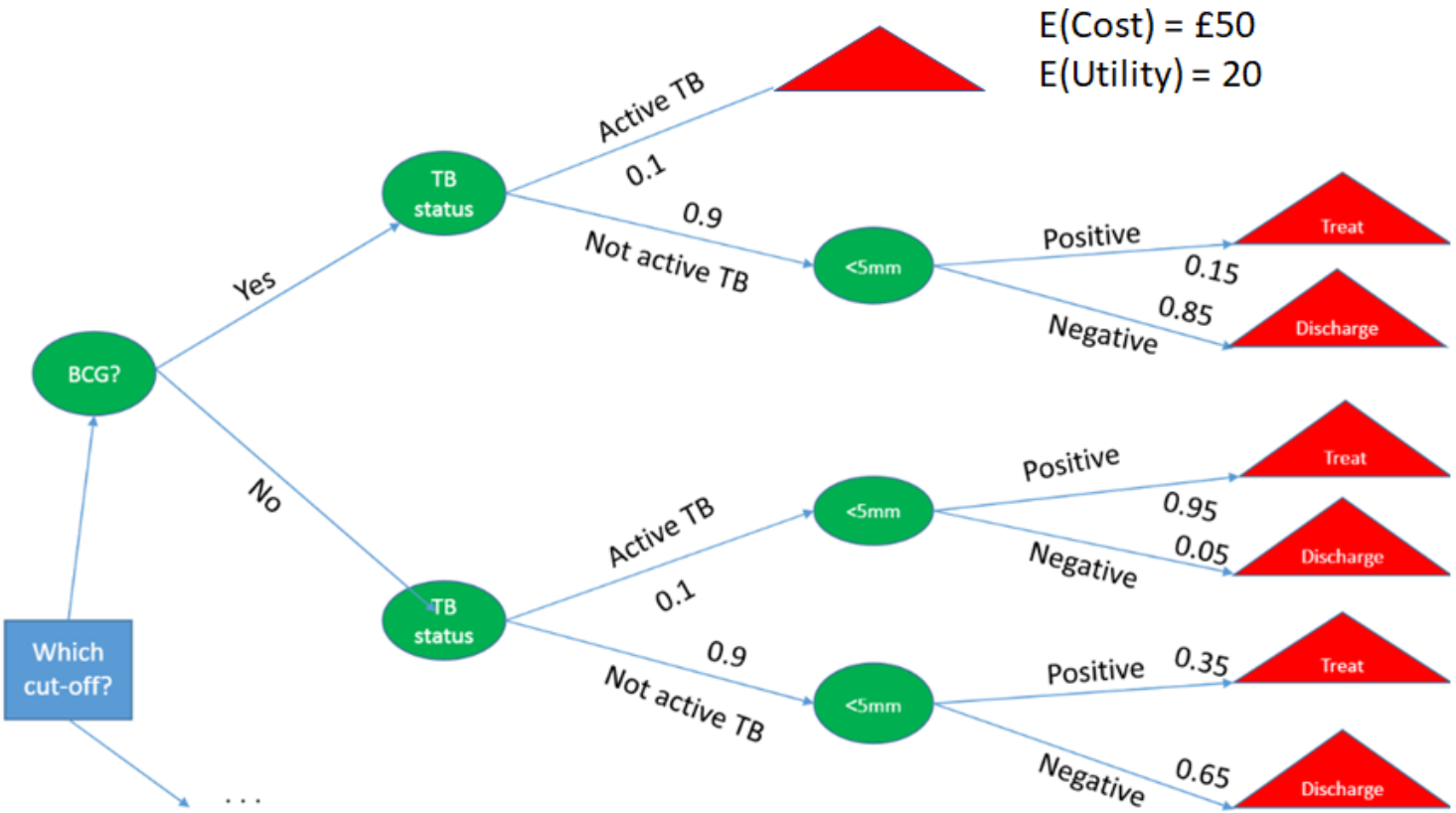
$$\mathbb{E}[V_i] = \begin{cases} r_i & \text{if} \quad i \in \mathcal{S}_{term} \\ r_i + \sum p_{ij}\mathbb{E}[V_j] & \text{otherwise} \end{cases}$$

  where
  - $\mathbb{E}$ is the weighted average of the values, V is the random variable total node value, e.g. cost or QALYs
  - r_i is the (unit) value at node i
  - $\mathcal{S}_{term}$ is the set of all terminal nodes. Note that at a terminal node the expected value is simply the value at that node.
- An advantage of using this method is that total expected values can be obtained at each node and so if there are multiple decision node not at the root of the tree the recursive approach can fold-back sub-trees.

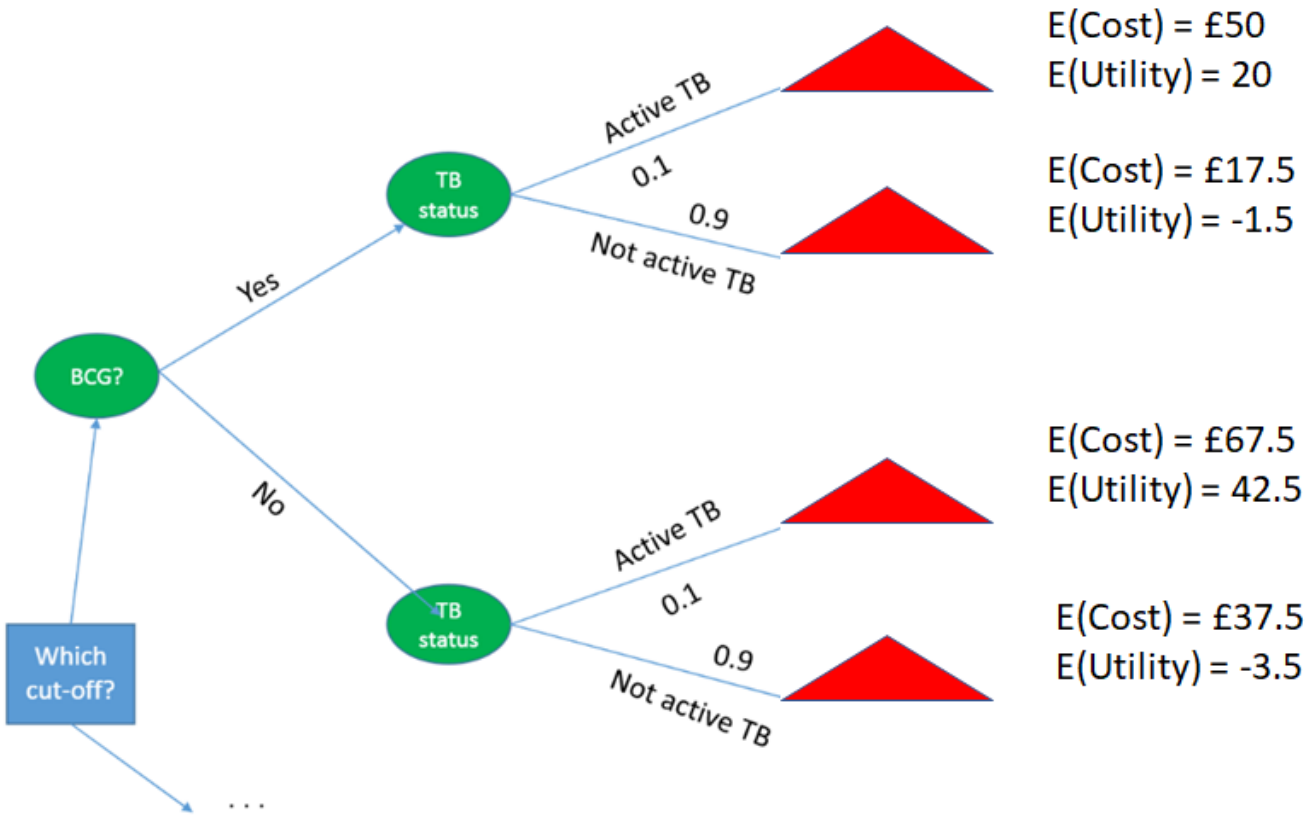
Simple decision tree example: New TST guidelines



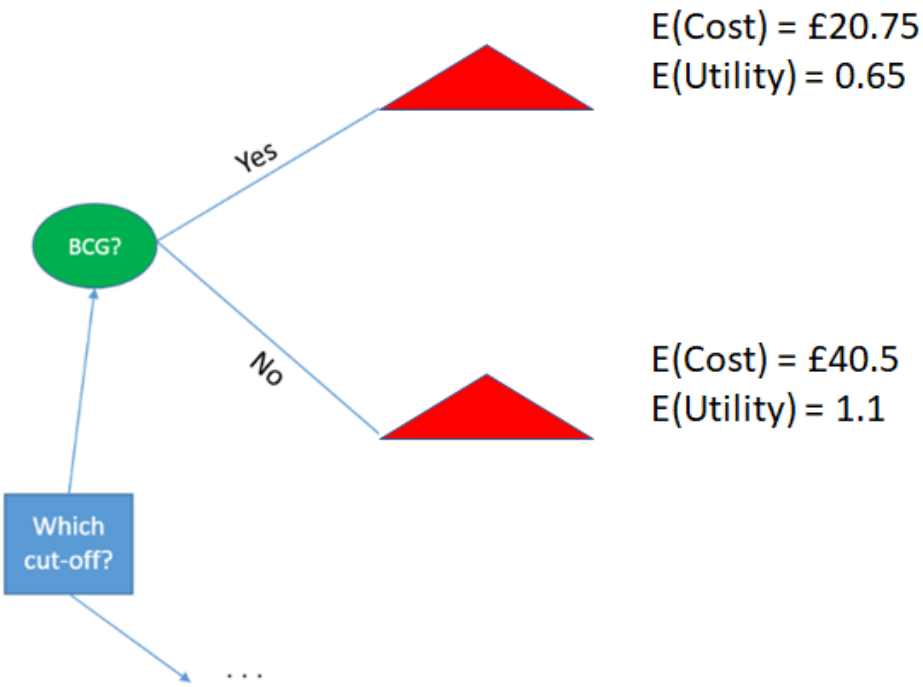
Simple decision tree example: New TST guidelines



Simple decision tree example: New TST guidelines



Simple decision tree example: New TST guidelines



"Forward" computation

- May be more intuitive

"Forward" computation

- May be more intuitive
- Calculate the total health and costs, and joint probability along all of the distinct pathways of the tree corresponding to a decision.

"Forward" computation

- May be more intuitive
- Calculate the total health and costs, and joint probability along all of the distinct pathways of the tree corresponding to a decision.
- The weighted average of the costs or health values give the expected value at the decision node.

"Forward" computation

- May be more intuitive
- Calculate the total health and costs, and joint probability along all of the distinct pathways of the tree corresponding to a decision.
- The weighted average of the costs or health values give the expected value at the decision node.
- Formally

$$\mathbb{E}[V] = \sum_{j \in S_{term}} r_j^* p_j^*$$

where

- $r_j^* = r_{[1]} + r_{[2]} + \cdots + r_{[n]}$ is the total of the values along each pathway with terminal node j
- $p_j^* = p(x_{[1]}, x_{[2]}, \ldots, x_{[n]})$ is the joint probabilities of traversing the unique path with terminal node j.
- The set of possible terminal nodes if the decision-maker takes a given decision is denoted S_{term}.

TB Combined probabilities for each branch

If probability of BCG is 0.1 (c.f. NHS Immunisation Statistics, England 2012-13) then

$$
\begin{aligned}
p_1^* &= 0.1 \times 0.1 \times 0.8 = 0.008 \\
p_2^* &= 0.1 \times 0.1 \times 0.2 = 0.002 \\
p_3^* &= 0.1 \times 0.9 \times 0.15 = 0.0135 \\
p_4^* &= 0.1 \times 0.9 \times 0.85 = 0.0765 \\
p_5^* &= 0.9 \times 0.1 \times 0.95 = 0.0855 \\
p_6^* &= 0.9 \times 0.1 \times 0.05 = 0.0045 \\
p_7^* &= 0.9 \times 0.9 \times 0.35 = 0.2835 \\
p_8^* &= 0.9 \times 0.9 \times 0.65 = 0.5265
\end{aligned}
$$

TB expected cost and health impact

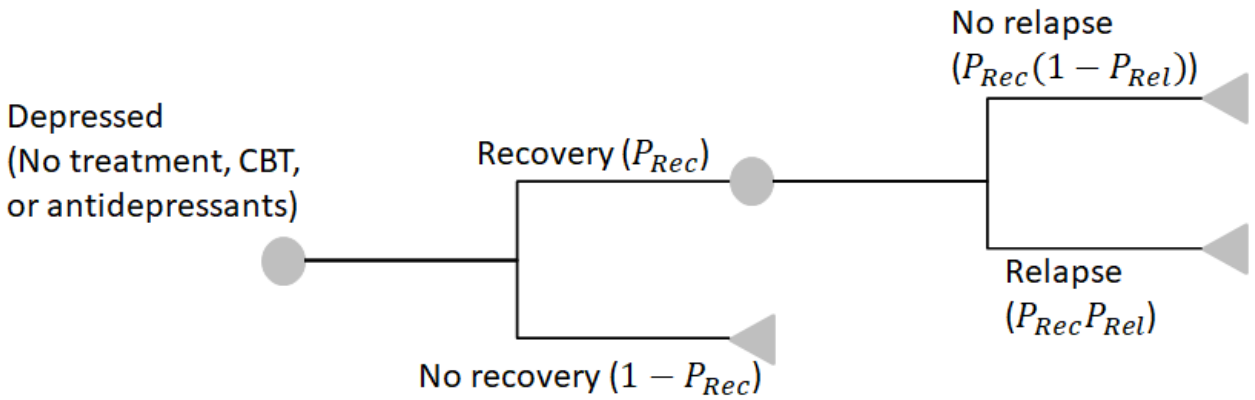
Cost:

$$
\begin{aligned}
&0.008 \times 60 + 0.003 \times 10 + 0.0135 \times 60 + 0.0765 \times 10 \\
&+0.0855 \times 70 + 0.0045 \times 20 + 0.2835 \times 70 + 0.5265 \times 20 \\
&= 38.535
\end{aligned}
$$

Health:

$$
\begin{aligned}
&0.008 \times 50 + 0.003 \times (-100) + 0.0135 \times (-10) + 0.0765 \times 0 \\
&+0.0855 \times 50 + 0.0045 \times (-100) + 0.2835 \times (-10) + 0.5265 \times 0 \\
&= 0.955
\end{aligned}
$$

Total costs and effects for treatment k



Costs	Effects
c_{Rec}	q_{Rec}
c_{Rel}	q_{Rel}
$c_{No\ rec}$	$q_{No\ rec}$

Costs if recover, no relapse

Costs if recover, then relapse

Costs if no recovery

$$Total\ costs_k = TC_k + P_{Rec,k} \times (1 - P_{Rel,k}) \times c_{Rec} + P_{Rec,k} \times P_{Rel,k} \times c_{Rel} + (1 - P_{Rec,k}) \times c_{No\ rec}$$

$$Total\ effects_k = P_{Rec,k} \times (1 - P_{Rel,k}) \times q_{Rec} + P_{Rec,k} \times P_{Rel,k} \times q_{Rel} + (1 - P_{Rec,k}) \times q_{No\ rec}$$

QALYs if recover, no relapse

QALYs if recover, then relapse

QALYs if no recovery

Lecture 4

Markov models

Outline of this lecture

- Theory of Markov modelling
- Example: Estimating Markov model transition probabilities from individual-level counts of transitions
- Bayesian inference ideas
  - Dirichlet and Multinomial distributions
  - Advantages of Bayes: including prior information, natural quantification of uncertainty (probabilistic sensitivity analysis).
- Implementing Markov models in R
- Brief discussion of decision modelling more generally, from Bayesian perspective

Markov models in discrete time

Assume

- a set of S "clinically relevant" states: exhaustive and mutually exclusive
- a discrete time axis, indexed by "cycles" or time units j
- a set of allowed transitions in a diagram of the states

Arrows connecting two states denote that a person can move

- from the state at the start of the arrow at time j
- to the state at the end of the arrow at time $j + 1$
- Absence of an arrow implies transition is not allowed by model

Markov models in discrete time

Assume

- a set of S "clinically relevant" states: exhaustive and mutually exclusive
- a discrete time axis, indexed by "cycles" or time units j
- a set of allowed transitions in a diagram of the states

Arrows connecting two states denote that a person can move

- from the state at the start of the arrow at time j
- to the state at the end of the arrow at time $j + 1$
- Absence of an arrow implies transition is not allowed by model

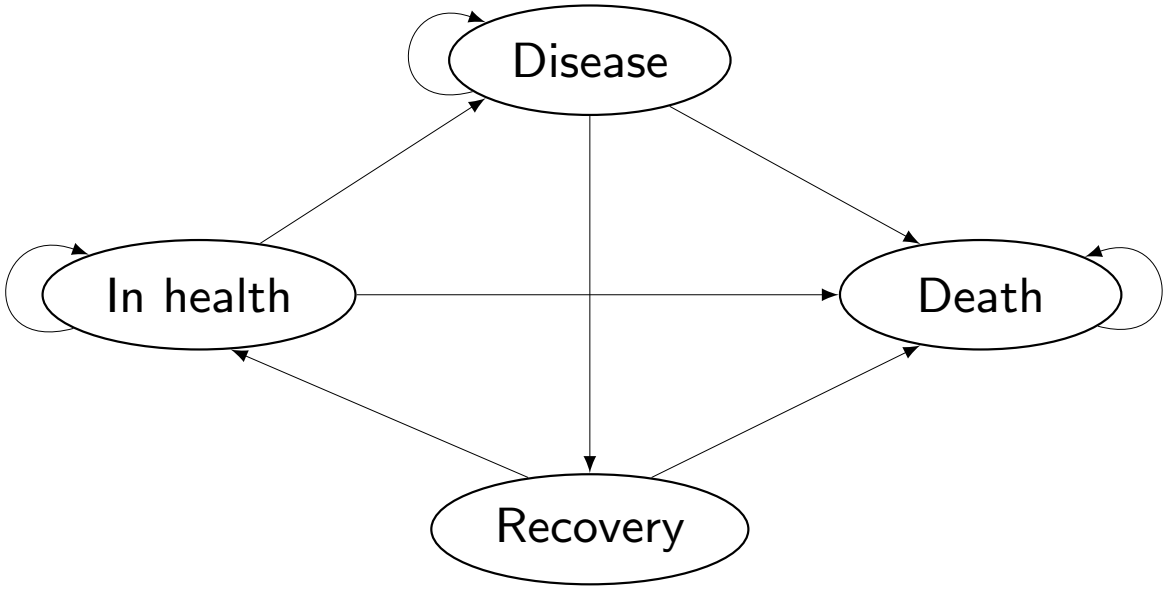
Movements occur according to suitable *transition probabilities*

$$\boldsymbol{\pi}_j = \boldsymbol{\pi}_{j-1}\boldsymbol{\Lambda}_j$$

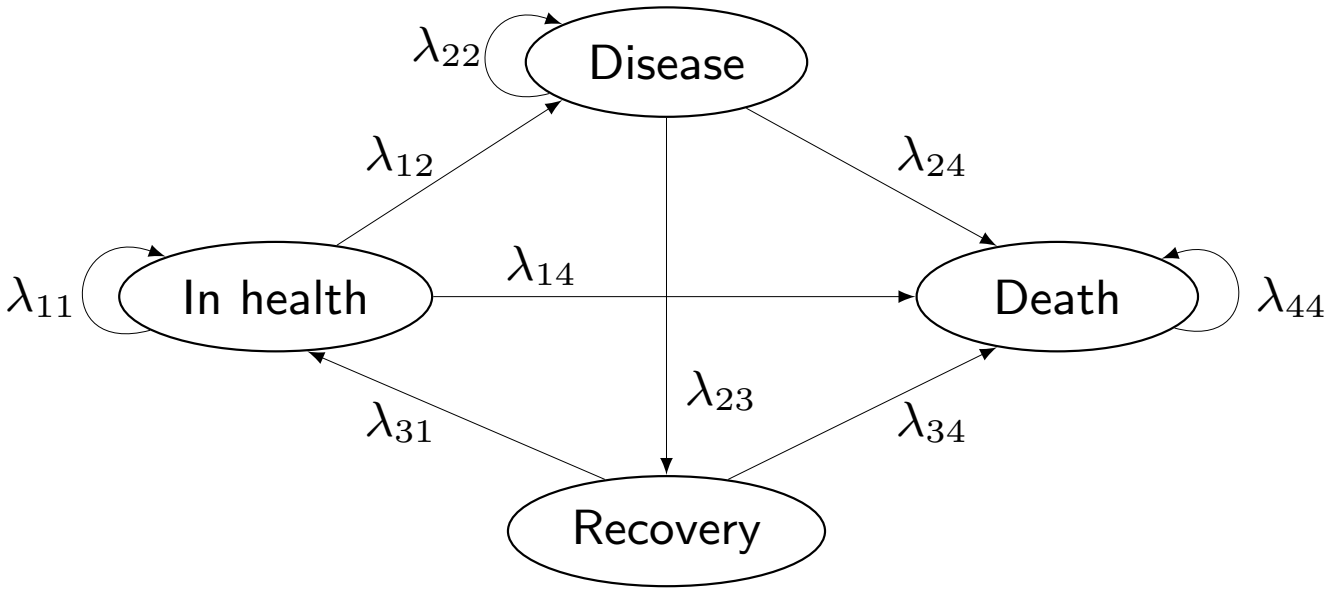
- $\boldsymbol{\pi}_j$ is the vector of probabilities of occupying each state at time j
- $\boldsymbol{\Lambda}_j = [\Lambda_{j;s,s'}]$ is the *transition probability matrix* at time j: s, s' entry is the probability of moving from state s to state s' at time j

Markov models

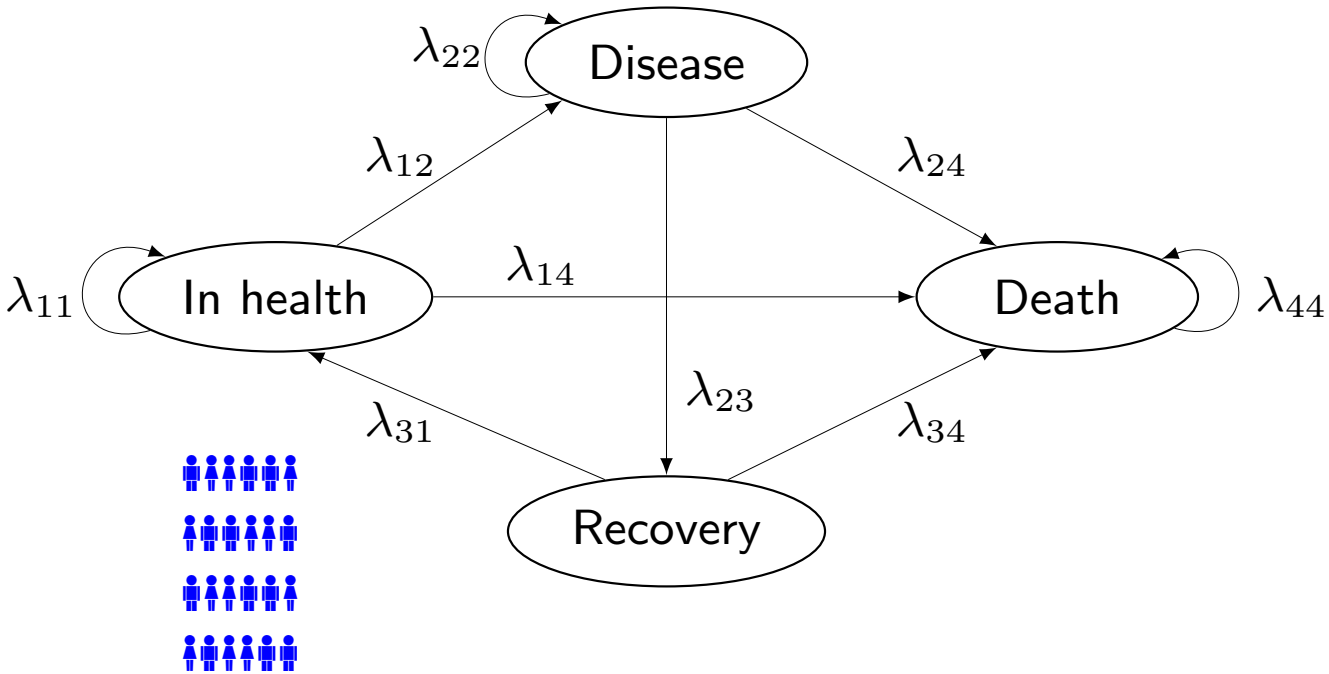
1. Define a structure



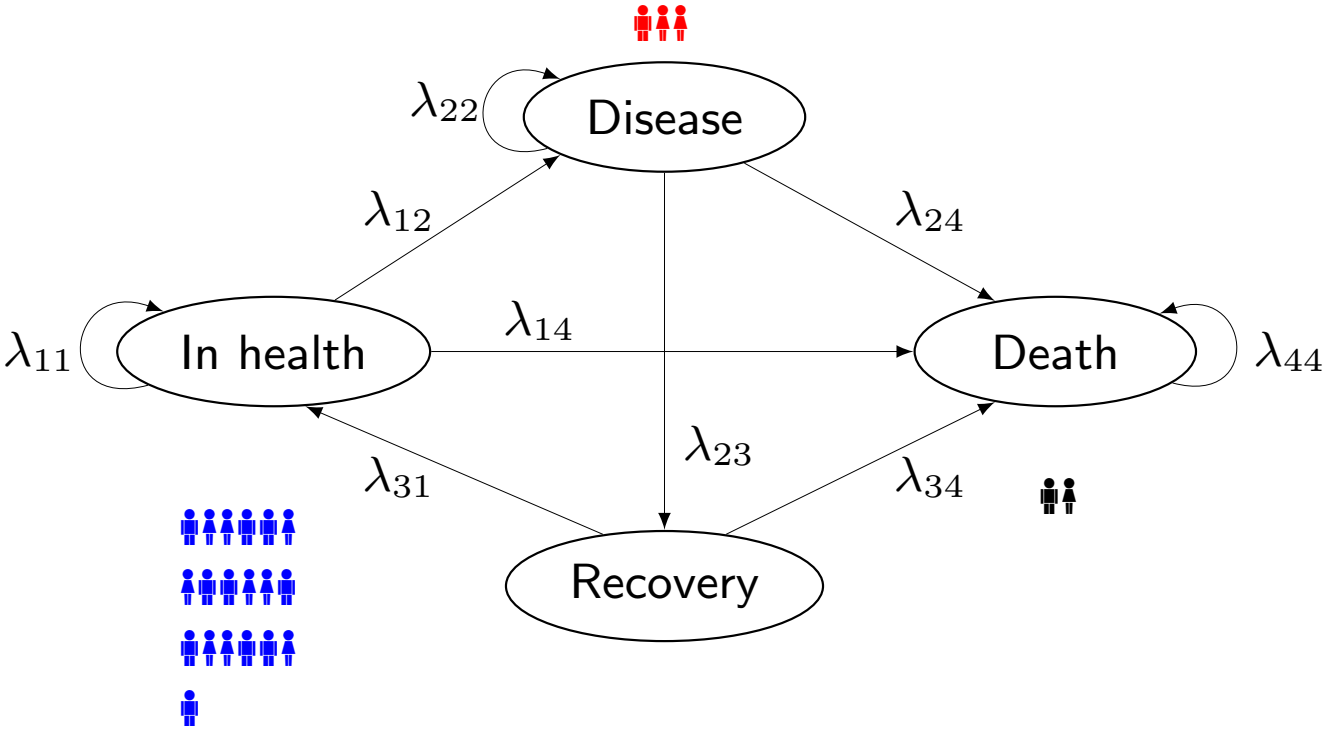
2. Estimate the transition probabilities from available, relevant data
Define costs and utilities associated with occupying each state s at each time j



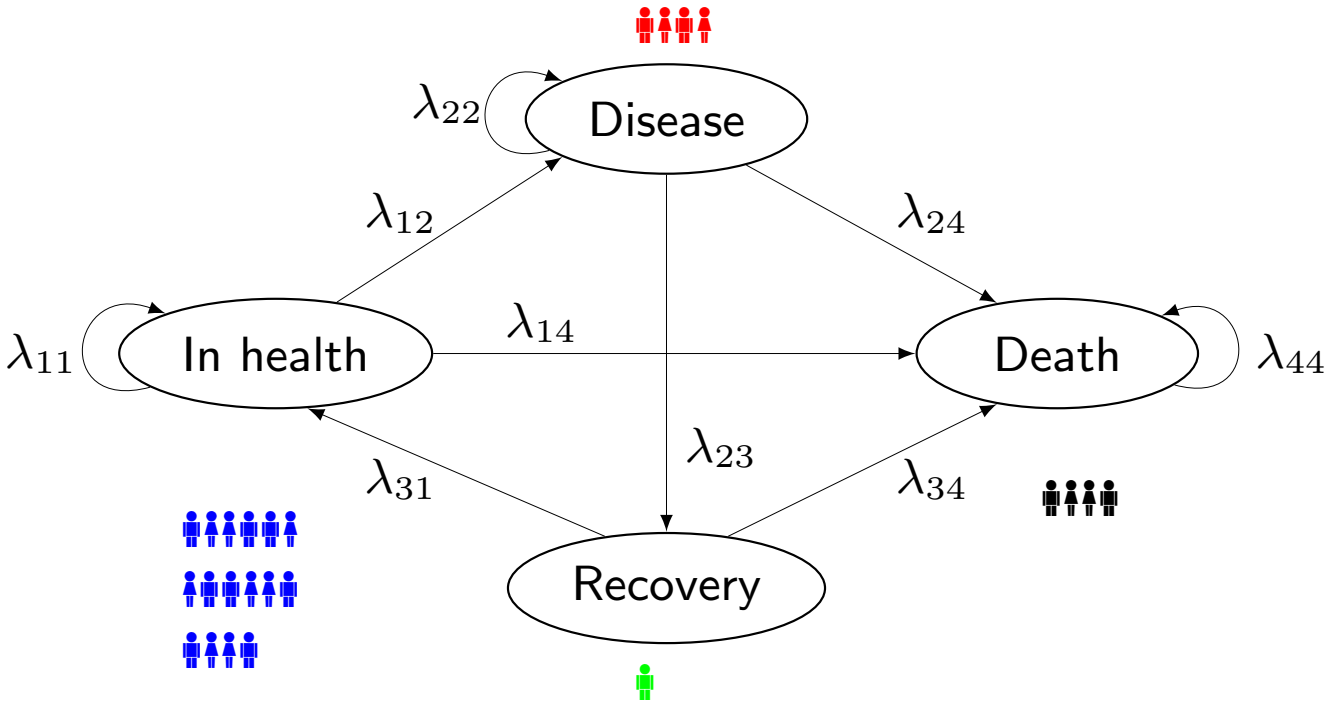
3. Run the simulation, recording:
proportion of people in each state $\rightarrow$ expected costs and effects,
at each time:
$j = 0$



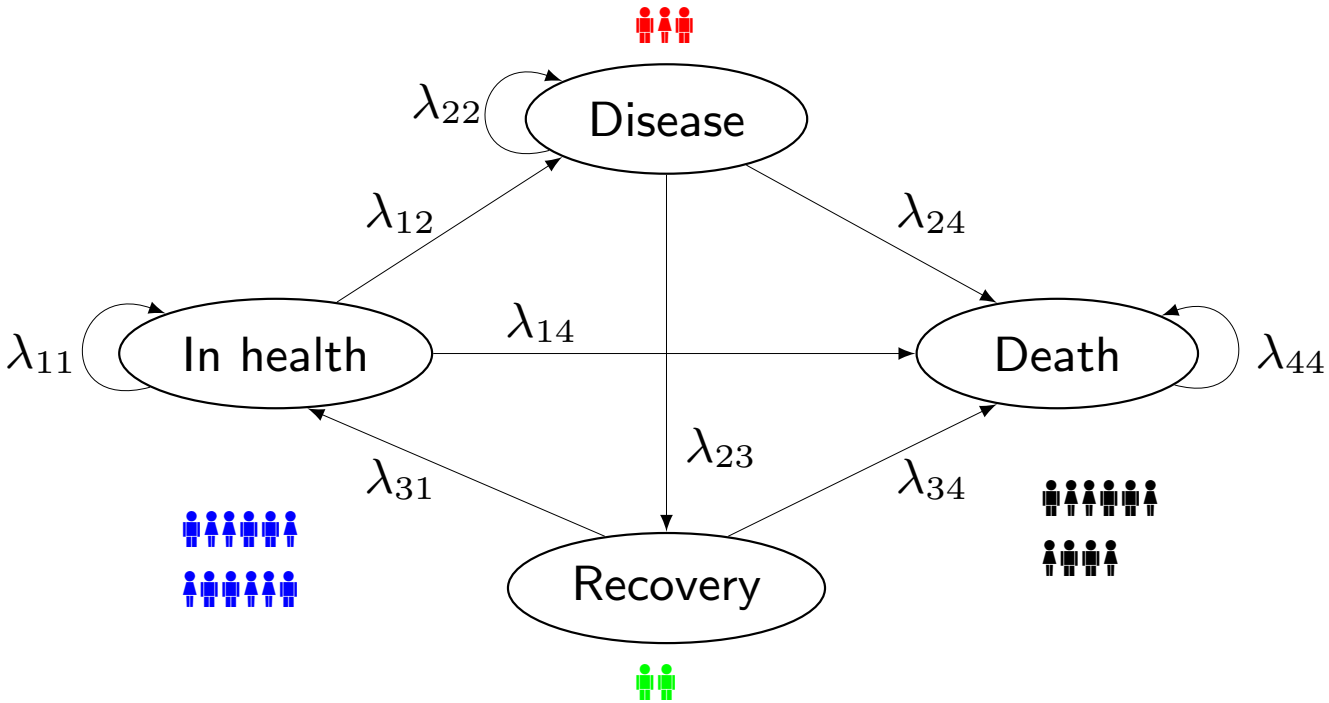
3. Run the simulation: $j = 1$



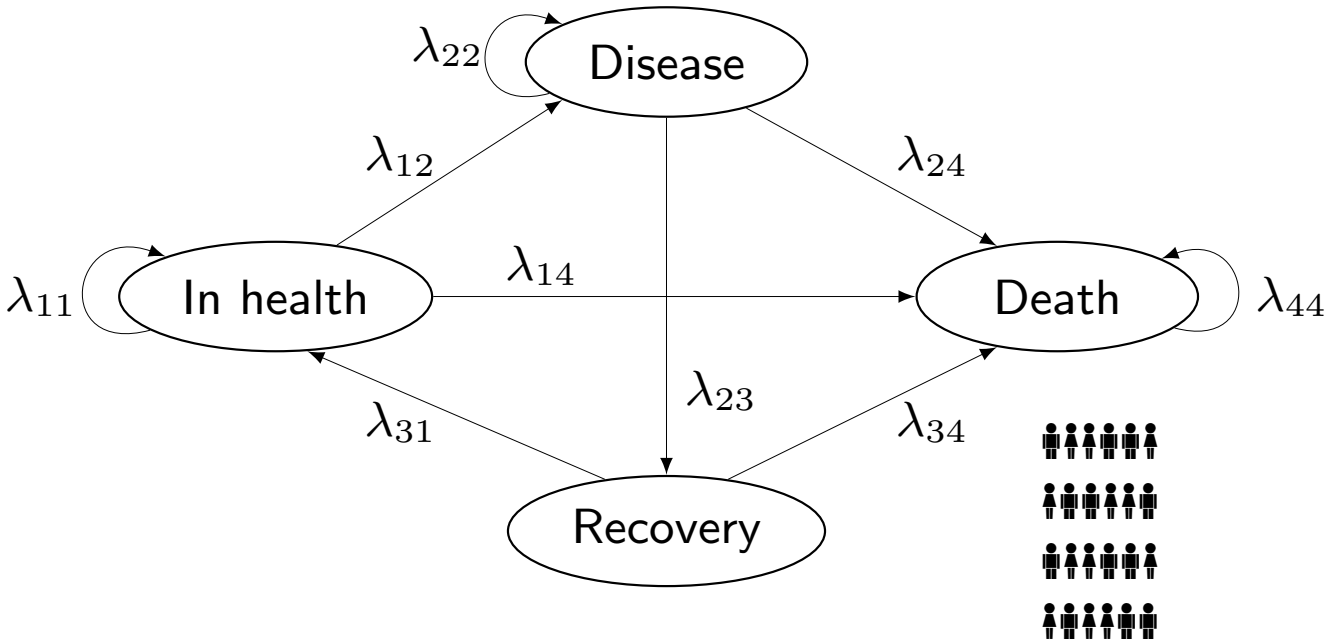
3. Run the simulation: $j = 2$



3. Run the simulation: $j = 3$



3. Run the simulation: $j = J$



Markov assumption (discrete time)

Markov models are multi-state models in which next transition depends only on the current state

$$\Pr(\text{state } s' \text{ at time } j+1 \mid \text{history of the process}) =$$
$$\Pr(\text{state } s' \text{ at time } j+1 \mid \text{state } s \text{ at time } j)$$

Markov assumption (discrete time)

Markov models are multi-state models in which next transition depends only on the current state

$$\Pr(\text{state } s' \text{ at time } j+1 \mid \text{history of the process}) =$$
$$\Pr(\text{state } s' \text{ at time } j+1 \mid \text{state } s \text{ at time } j)$$

A Markov model is *time-homogeneous* if this transition probability is the same for all times j.

- Counterexample: risk of death depends on age.

Markov assumption (discrete time)

Markov models are multi-state models in which next transition depends only on the current state

$$\Pr(\text{state } s' \text{ at time } j+1 \mid \text{history of the process}) =$$
$$\Pr(\text{state } s' \text{ at time } j+1 \mid \text{state } s \text{ at time } j)$$

A Markov model is *time-homogeneous* if this transition probability is the same for all times j.

- Counterexample: risk of death depends on age.

Non-Markov models can often be made Markov by adding extra states

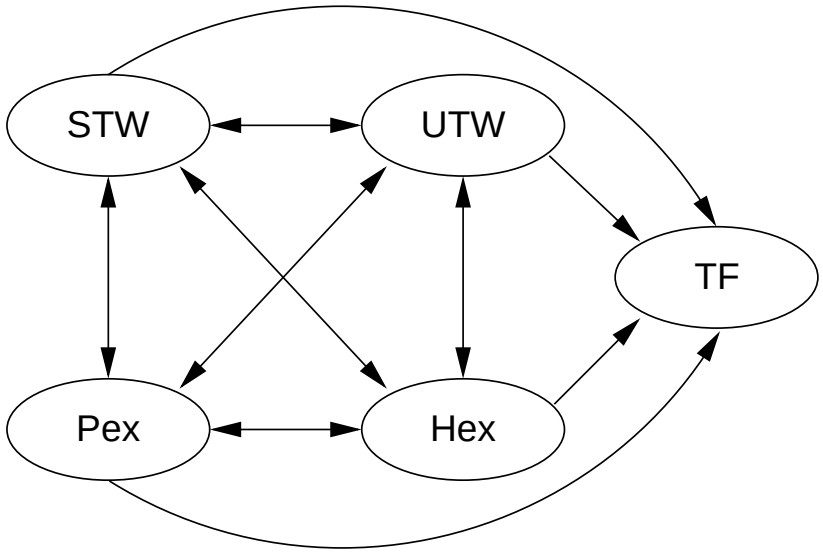
- e.g. states health→disease→ death
- suppose risk of death changes with time spent with disease
- split disease into "early disease"→"late disease", with differing risks of death

Remember these models are only approximations to a true process that is usually continuous-time, continuous-state

Example: asthma[1]

Five-state model for management of asthma:

- STW: Successfully treated week
- UTW: Unsuccessfully treated week
- Hex: Hospital-managed exacerbation
- Pex: Primary care-managed exacerbation
- TF: Treatment failure - enters a "usual care" pattern (absorbing)



[1] Briggs A, Ades AE and Price MJ. Medical Decision Making (2003); 23:341-350

Data

Estimate the Markov model transition probabilities using individual-level transition data

- From a RCT with 2 treatments
  - SFC: Salmeterol (50μg) / fluticasone propionate (100μg) in combination
    - ★ new treatment, $t = 2$
  - FP: Fluticasone propionate alone (100μg)
    - ★ existing treatment, $t = 1$
- 12 week trial
- For each arm we count the number of transitions between states from week j to $j + 1$
- From the Markov assumption, these can be considered independent

Data

SFC						
	Number in state at week $j+1$					Total in state
$r_{ss'}$	STW	UTW	Hex	Pex	TF	at week j (n_s)
STW	210	60	0	1	1	272
UTW	88	641	0	4	13	746
Hex	0	0	0	0	0	0
Pex	1	0	0	0	1	2
TF	0	0	0	0	81	81
FP						
STW	66	32	0	0	2	100
UTW	42	752	0	5	20	819
Hex	0	0	0	0	0	0
Pex	0	4	0	1	0	5
TF	0	0	0	0	156	156

These are summed over all weeks

Estimating transition probabilities

Could estimate (s, s') transition probability by dividing total

- number of transitions observed from s to s', by
- number of weeks observed in state s

But can't estimate transition rates out of Hex since nobody went into Hex!
Also very few went into Pex

- estimates / SEs for rates *out of* Pex will be unstable.

Economic model needs to

- include possibility of Hex / Pex – expensive states!
- account for *uncertainty* in the transition rates.

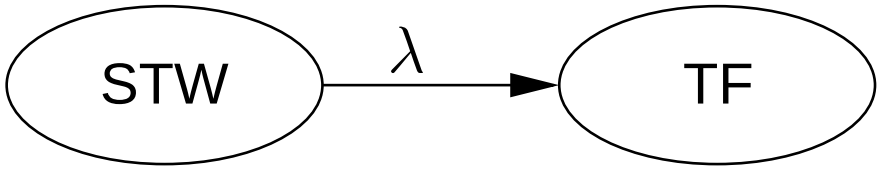
$\rightarrow$ use Bayesian inference

- – combine *priors* on Hex/Pex/other states with data
- $\rightarrow$ *posterior* distribution of transition probabilities.

Binomial distribution for an event

Suppose there are two states in the model (*successfully treated week* / *treatment failure*).

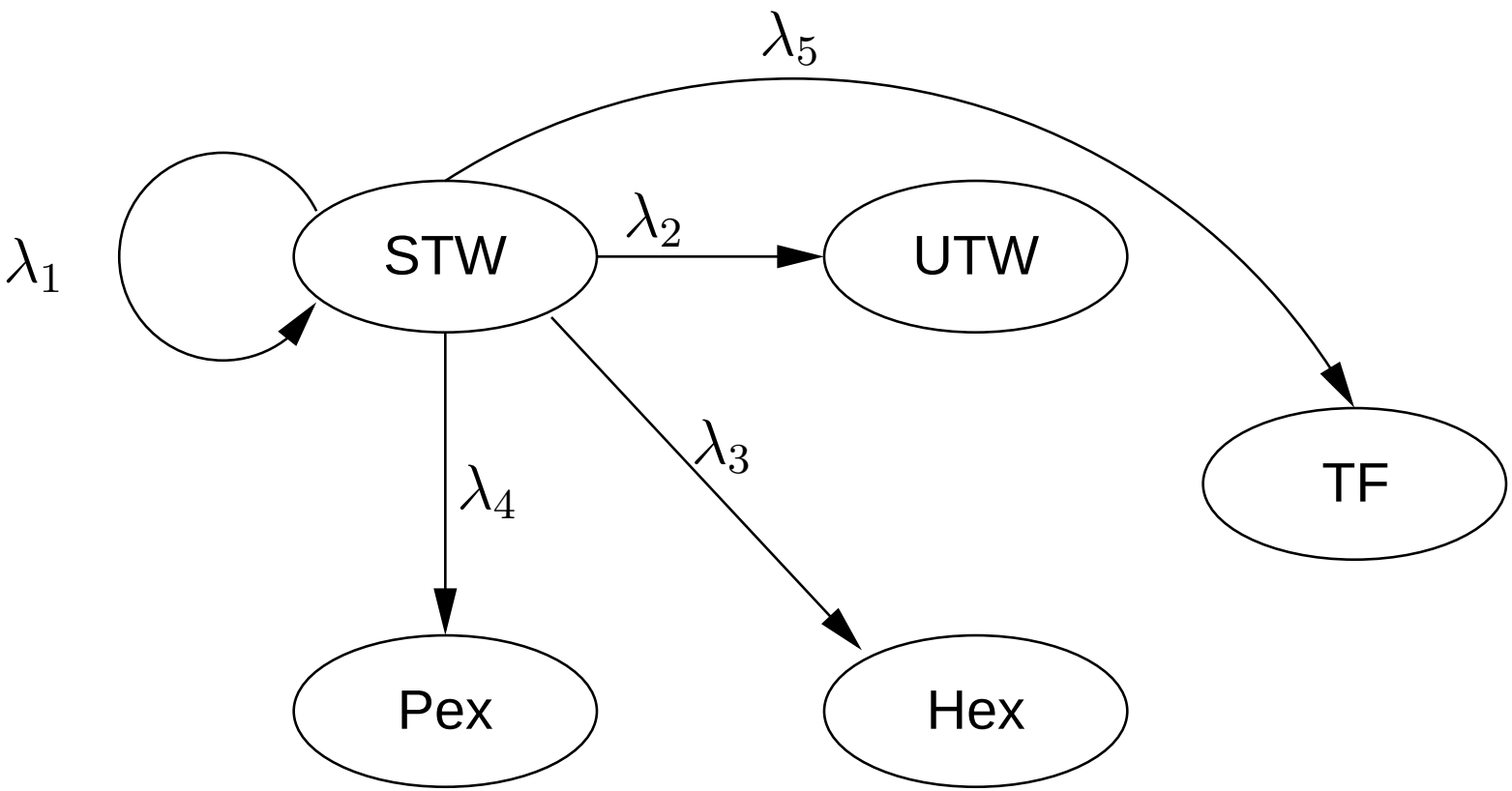
Out of n people currently under treatment, the following week r people have had treatment failure.



Total in STW at j	Number in state at $j+1$	
at week j	STW	TF
n	$n-r$	r

Beta distribution is a convenient prior for λ (conjugacy)

Multinomial model for several events



Number in state at $j+1$					Total in STW at j
STW	UTW	Hex	Pex	TF	at week j
r_1	r_2	r_3	r_4	r_5	n

Bayesian estimation of multinomial model

Likelihood

$$\begin{array}{lcl} r_1, \ldots, r_5 & \sim & \text{Multinomial}(\lambda_1, \ldots, \lambda_5, n) \\ p(\boldsymbol{r} \mid \boldsymbol{\lambda}) & = & \frac{n!}{r_1! \cdots r_5!} \lambda_1^{r_1} \ldots \lambda_5^{r_5} \quad \propto \quad \lambda_1^{r_1} \ldots \lambda_5^{r_5} \\ \sum \lambda_s & = & 1 \end{array}$$

Prior distribution for five transition probabilities

$$\begin{array}{lcl} (\lambda_1, \ldots, \lambda_5) & \sim & \text{Dirichlet}(a_1, \ldots, a_5) \\ p(\lambda_1, \ldots, \lambda_5) & = & \frac{\Gamma(a_1 + \cdots + a_5)}{\Gamma(a_1) \cdots \Gamma(a_5)} \lambda_1^{(a_1 - 1)} \cdots \lambda_5^{(a_5 - 1)} \quad \propto \quad \lambda_1^{(a_1 - 1)} \cdots \lambda_5^{(a_5 - 1)} \end{array}$$

Posterior distribution

$$\begin{array}{rcl} p(\boldsymbol{\lambda} \mid \boldsymbol{r}) & \propto & \lambda_1^{(a_1 + r_1 - 1)} \cdots \lambda_5^{(a_5 + r_5 - 1)} \\ \boldsymbol{\lambda} & \sim & \text{Dirichlet}(a_1 + r_1, \ldots, a_5 + r_5) \end{array}$$

Dirichlet prior distribution

- General prior for a set of numbers that add up to 1.
  $(p_1, \ldots, p_S) \sim \text{Dirichlet}(a_1, \ldots, a_S)$
  - a_s proportional to *expected* probability p_s of outcome s
  - scale of the a_s indicates prior *precision* of $(p_1, \ldots, p_S)$

- **NB**: Dirichlet$(1, \ldots, 1)$ is a flat prior.
- Given a sample size $\sum a_s$, a_s is your prior expectation for the number of patients you would expect in each state s

Multinomial / Dirichlet model in BUGS (cfr BMHE, §5.4.1)

```
model {
# Multinomial distribution for r, s=1,...,4: non-absorbing states
   for(s in 1:(S-1)){
      r.sfc[s,1:S] ~ dmulti(lambda.sfc[s,1:S], n.sfc[s])
      r.fp[s,1:S] ~ dmulti(lambda.fp[s,1:S], n.fp[s])
   }

# Dirichlet prior distributions for the transition probabilities lambda
   for(s in 1:(S-1)){
      lambda.sfc[s,1:S] ~ ddirch(prior.sfc[s,1:S])
      lambda.fp[s,1:S] ~ ddirch(prior.fp[s,1:S])
   }

# Fixed transition probabilities for the absorbing state - prob 1 of staying there/prob 0 of moving
   for(s in 1:S){
      lambda.sfc[S,s] <- p.fixed[s]
      lambda.fp[S,s] <- p.fixed[s]
   }
}
```

Simulating from a Dirichlet in R

However the Dirichlet posterior is known here.

We can simulate directly from a known Dirichlet distribution in R by exploiting that a Dirichlet random vector is a function of Gamma random variables:

- Simulate $Z_s \sim \text{Gamma}(a_s, 1)$ for $s = 1, \ldots, S$. Then
- $(Z_1, \ldots, Z_S) / \sum_{i=1}^{S} Z_i \sim \text{Dirichlet}(a_1, \ldots, a_S)$

Simulating from a Dirichlet in R

However the Dirichlet posterior is known here.

We can simulate directly from a known Dirichlet distribution in R by exploiting that a Dirichlet random vector is a function of Gamma random variables:

- Simulate $Z_s \sim \text{Gamma}(a_s, 1)$ for $s = 1, \ldots, S$. Then
- $(Z_1, \ldots, Z_S) / \sum_{i=1}^{S} Z_i \sim \text{Dirichlet}(a_1, \ldots, a_S)$

For example, to simulate 1000 draws from a Dirichlet(1,1,1,1):

```
a <- c(1, 1, 1, 1)
dir_rand <- matrix(nrow=1000, ncol=4)
for (i in 1:4)
    dir_rand[,i] <- rgamma(1000, a[i], 1)
dir_rand <- dir_rand / rowSums(dir_rand)
```

Simulating from a Dirichlet in R

However the Dirichlet posterior is known here.

We can simulate directly from a known Dirichlet distribution in R by exploiting that a Dirichlet random vector is a function of Gamma random variables:

- Simulate $Z_s \sim \text{Gamma}(a_s, 1)$ for $s = 1, \ldots, S$. Then
- $(Z_1, \ldots, Z_S) / \sum_{i=1}^{S} Z_i \sim \text{Dirichlet}(a_1, \ldots, a_S)$

For example, to simulate 1000 draws from a Dirichlet(1,1,1,1):

```
a <- c(1, 1, 1, 1)
dir_rand <- matrix(nrow=1000, ncol=4)
for (i in 1:4)
    dir_rand[,i] <- rgamma(1000, a[i], 1)
dir_rand <- dir_rand / rowSums(dir_rand)
```

Or functions to simulate directly from the Dirichlet are provided in several R add-on packages available from CRAN, e.g.

```
install.packages("VGAM") # if not already installed
library(VGAM)
dir_rand <- rdiric(1000, a)
```

Estimated transition probabilities versus observed data

After obtaining the posterior Dirichet distributions

	Raw transition counts					*Bayesian posterior mean*				
To:	STW	UTW	Hex	Pex	TF	STW	UTW	Hex	Pex	TF
From:	SFC									
STW	210	60	0	1	1	0.76	0.22	0.004	0.01	0.01
UTW	88	641	0	4	13	0.12	0.85	0.001	0.01	0.02
Hex	0	0	0	0	0	0.20	0.20	0.20	0.20	0.20
Pex	1	0	0	0	1	0.29	0.14	0.14	0.14	0.28
TF	0	0	0	0	81	0	0	0	0	1
	FP									
STW	66	32	0	0	2	0.64	0.31	0.01	0.01	0.03
UTW	42	752	0	5	20	0.05	0.91	0.001	0.00	0.03
Hex	0	0	0	0	0	0.20	0.20	0.20	0.20	0.20
Pex	0	4	0	1	0	0.10	0.50	0.10	0.20	0.10
TF	0	0	0	0	156	0	0	0	0	1

Information from the prior has been combined with the data

Cost-effectiveness modelling

Calculate *expected costs* and *expected benefits* of each treatment over the long term:

- *Cohort simulation:* Estimate proportion of population in each state at each successive time (from Markov *transition probabilities*)
  - at the start, everyone in "base" state (e.g. just received treatment)
  - usually in the long run, everyone will be in the "absorbing" state (e.g. death)
- Assign cost and benefit to each state in model
- Accumulate expected cost and benefit over times

NB: Check BMHE 5.4 + practical!

Discounting

- Costs and outcomes can occur at different times with respect to when the intervention is implemented
  - Society tends to value benefits that arrive closer to the present time more than those that are achievable later in the future
- *Discounting* accounts for differential timing by reducing the value of costs & effects in the future
  - Particularly relevant for economic evaluation spanning over a time horizon > 1 year (e.g. Markov models)

Discounting

- Costs and outcomes can occur at different times with respect to when the intervention is implemented
  - Society tends to value benefits that arrive closer to the present time more than those that are achievable later in the future
- *Discounting* accounts for differential timing by reducing the value of costs & effects in the future
  - Particularly relevant for economic evaluation spanning over a time horizon > 1 year (e.g. Markov models)
- If cost at some future time j is c_{tj}, then *discounted* cost is $c_{tj}/(1+d)^j$.
- Discount rate d: NICE suggest 3.5% for costs and outcomes
- **Present Value** of intervention (treatment) t over a time horizon J is

$$\text{PV}_t^c = \sum_{j=0}^{J} \frac{c_{tj}}{(1+d)^j} \qquad \text{PV}_t^e = \sum_{j=0}^{J} \frac{e_{tj}}{(1+d)^j}$$

(for costs c and clinical benefits e, respectively)

Simulating prevalences of states over time

- π_{js} is the **expected proportion** of patients in state s at time j
- $\lambda_{ss'}$ is the probability of going from state s to state s' in one time step
- Thus

$$\pi_{j+1,s} = \pi_{j1}\lambda_{1s} + \cdots + \pi_{jS}\lambda_{Ss}$$

which we can write in matrix form as

$$\begin{aligned} (\pi_{j+1,1}, \ldots, \pi_{j+1,S}) &= (\pi_{j,1}, \ldots, \pi_{j,S}) \begin{pmatrix} \lambda_{11} & \cdots & \lambda_{1S} \\ \vdots & \ddots & \vdots \\ \lambda_{S1} & \cdots & \lambda_{SS} \end{pmatrix} \\ \boldsymbol{\pi}_{j+1} &= \boldsymbol{\pi}_j \boldsymbol{\Lambda} \end{aligned}$$

- **NB**: $\boldsymbol{\Lambda}$ may depend on time j

Simulating prevalences of states over time

- π_{js} is the **expected proportion** of patients in state s at time j
- $\lambda_{ss'}$ is the probability of going from state s to state s' in one time step
- Thus

$$\pi_{j+1,s} = \pi_{j1}\lambda_{1s} + \cdots + \pi_{jS}\lambda_{Ss}$$

  which we can write in matrix form as

$$\begin{aligned} (\pi_{j+1,1}, \ldots, \pi_{j+1,S}) &= (\pi_{j,1}, \ldots, \pi_{j,S}) \begin{pmatrix} \lambda_{11} & \cdots & \lambda_{1S} \\ \vdots & \ddots & \vdots \\ \lambda_{S1} & \cdots & \lambda_{SS} \end{pmatrix} \\ \boldsymbol{\pi}_{j+1} &= \boldsymbol{\pi}_j \boldsymbol{\Lambda} \end{aligned}$$

- **NB**: $\boldsymbol{\Lambda}$ may depend on time j
- Instead of calculating **expected** proportions π_{js}, could simulate the **absolute number** m_{js} of patients in a cohort of size M, in state s, at time j.
  - may help to illustrate what model is doing, though decision is generally based on *expected* cost (average cost for an infinite population)

Coding this in R (see BMHE 5.4)

Markov cost-effectiveness models can be implemented straightforwardly in R.

```
pi_sfc <- pi_fp <- array(dim = c(n.sims, # number of PSA samples
                                 S,    # number of states
                                 J+1)) # number of times
for (i in 1:n_sims){                    # Assumes only 1 patient
    pi_sfc[i,,1] <- c(1,0,0,0,0)        #  initially in the state
    pi_fp[i,,1] <- c(1,0,0,0,0)         #  'In health'
}

for (i in 1:n.sims) {
    for (j in 2:(J+1)){
        for (s in 1:S){
            pi_sfc[i,s,j] <- sum(pi_sfc[i,,j-1]*lambda_sfc[i,,s])
            pi_fp[i,s,j] <- sum(pi_fp[i,,j-1]*lambda_fp[i,,s])
        }
    }
}
```

Many different ways to write code like this: see the questions sheet / solutions for another way.

See also heemod, an R package designed for health economic modelling
https://CRAN.R-project.org/package=heemod

Costs and utilities

- Assign a cost to each state for each treatment

	Cost per week (£)				
Treatment	STW	UTW	Hex	Pex	TF
SFC	7.96	7.96	1821.17	100.79	TF cost (j)
FP	2.38	2.38	1815.58	95.21	TF cost (j)

- In this example, assume the expected cost for week j in SFC group is

$$\pi_{j1} \times \text{cost}_{(SFC,1)} + \cdots + \pi_{j4} \times \text{cost}_{(SFC,4)} + \pi_{j5} \times \text{TF cost}(j)$$

  - $\text{TF cost}(j) = \pi_{j1}\text{cost}_{(FP,1)} + \cdots + \pi_{j4}\text{cost}_{(FP,4)}$ = average cost over states, weighted by probability of occupying state, under old treatment

- Utility measure is the number of weeks in the successfully controlled state

Expected costs and effects from a Markov model in R

Combining probabilities of state occupancy (pi_sfc) with state-specific costs and utilities, for SFC treatment group (with similar code for FP group)

```
state_costs_sfc <- c(STW=7.96, UTW=7.96, Hex=1821.17, Pex=100.79, TF=NA)
state_util_sfc <- c(1, 0, 0, 0, 0)
cost_sfc <- eff_sfc <- numeric(nsim)

## Loop over PSA samples
for (i in 1:nsim){
    cts <- ets <- numeric(J)

    ## Loop over times j
    for (j in 1:J){
        state_costs_sfc["TF"] <- state_costs_sfc[1:4] %*% pi_sfc[i,1:3,j]
        # costs for time j, averaged over states
        cts[j] <- sum(pi_sfc[i,,j] * state_costs_sfc)
        # effects for time j, averaged over states
        ets[j] <- sum(pi_sfc[i,,j] * state_util_sfc)
    }
    # expected costs and effects summed over all times, and averaged over states
    # for PSA sample i
    cost_sfc[i] <- sum(cts)
    eff_sfc[i] <- sum(ets)
}
```

Summary: Bayesian health economic model

At each iteration of "probabilistic sensitivity analysis", for each arm

- **Step 1**:
  - draw a realization of the transition probabilities $\mathbf{\Lambda}$ from their posterior distribution
  - define starting state and set time $j = 1$

Summary: Bayesian health economic model

At each iteration of "probabilistic sensitivity analysis", for each arm

- **Step 1**:
  - draw a realization of the transition probabilities $\mathbf{\Lambda}$ from their posterior distribution
  - define starting state and set time $j = 1$
- **Step 2**: Markov model
  - Set time $j = j + 1$
  - find prevalence of states π_j at this time
  - combine π_j with state-specific cost and utility to calculate expected cost and utility at this time
  - repeat step 2 up to time horizon $j = J (= 13$ in this case)

Summary: Bayesian health economic model

At each iteration of "probabilistic sensitivity analysis", for each arm

- **Step 1**:
  - draw a realization of the transition probabilities $\mathbf{\Lambda}$ from their posterior distribution
  - define starting state and set time $j = 1$
- **Step 2**: Markov model
  - Set time $j = j + 1$
  - find prevalence of states π_j at this time
  - combine π_j with state-specific cost and utility to calculate expected cost and utility at this time
  - repeat step 2 up to time horizon $j = J (= 13$ in this case)
- **Step 3**: Cost-effectiveness
  - calculate expected *costs*
  - calculate expected *effects* (here, proportion of time spent in STW)

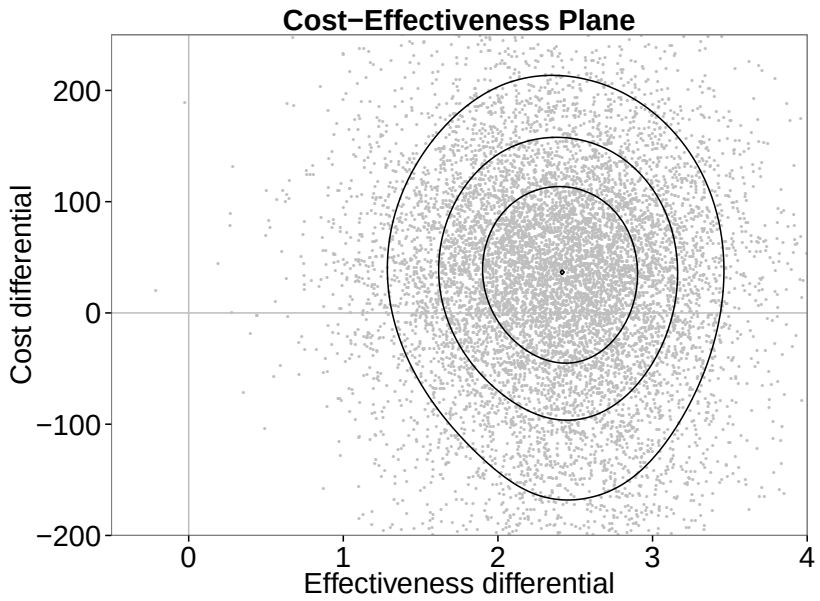
  summed over times

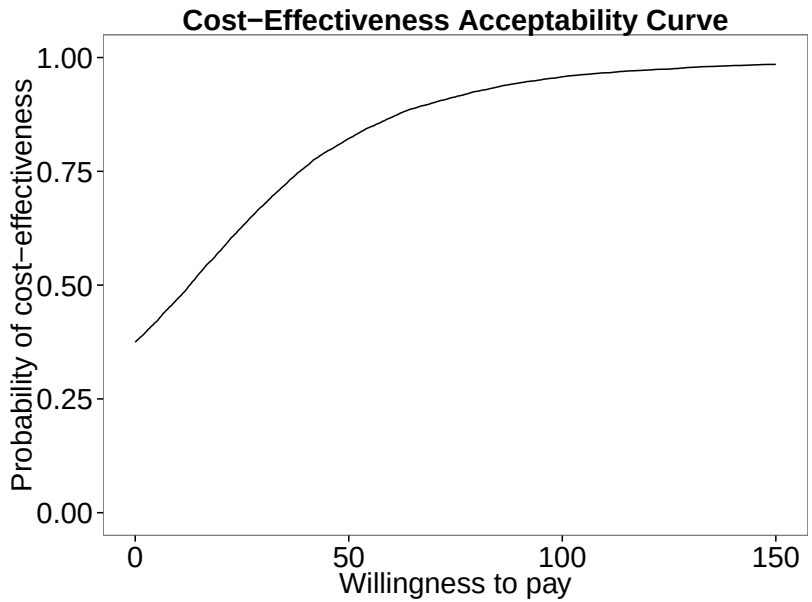
Gives a sample from the **posterior distribution** of expected costs and effects: vectors `cost_sfc`, `eff_sfc`, `cost_fp`, `eff_fp` in R code

Results

Sample from *joint posterior distribution* of costs and effects for each treatment gives, e.g.

- *incremental cost-effectiveness ratio* = mean cost difference / mean benefit difference
- "cost-effectiveness plane"
- incremental net benefit (with CI)
- cost-effectiveness acceptability curve





Bayesian decision modelling more generally

Example given here was simple

- Markov model with all transition probabilities estimated from individual-level transition count data.

Many more complicated examples, e.g.

- Transition probabilities for baseline disease progression estimated from published data
- Treatment effects estimated from RCTs / (network-)meta-analysis $\rightarrow$ transition probabilities for treated groups
- Combinations of individual-level and published aggregate data...

Examples in:

- *Bayesian Methods in Health Economics*,
- *Evidence Synthesis for Decision Making in Healthcare*,
- http://nicedsu.org.uk/technical-support-documents/

Lecture 5

Uncertainty analysis

Handling uncertainty in economic evaluations

- Population uncertainty: Sub-group analysis
- Parameter uncertainty: Sensitivity analysis
- Structural uncertainty: Sensitivity analysis
- Collect more data
- Sensitivity analysis (SA)
- Deterministic sensitivity analysis
  - One-way SA
  - Two-way SA
  - Scenario analysis (best or worst case, what if..)
- Probabilistic sensitivity analysis (PSA)
  - Monte Carlo simulation
- Cost-effectiveness acceptability curves (CEAC)

How robust are health economic evaluations

- Do limitations in either the quality or availability of evidence affect the recommended decision?
- If the decision is not altered despite 'reasonable' variations in key assumptions/parameters, then the analysis can be considered to be 'robust'
- Two types of uncertainty:
  - Structural (is the model design correct?)
  - Parameter (are the values correct?)
- In economic evaluation, some form of sensitivity analysis is frequently carried out in order to allow for uncertainty
- This uncertainty may be present in the evaluation for several reasons:
  - Data are unavailable and assumptions are necessary
  - Available but inaccurate
- In this type of analysis the values recorded for important parameters are varied, usually one at a time, in order to determine whether the results are sensitive to the assumptions made

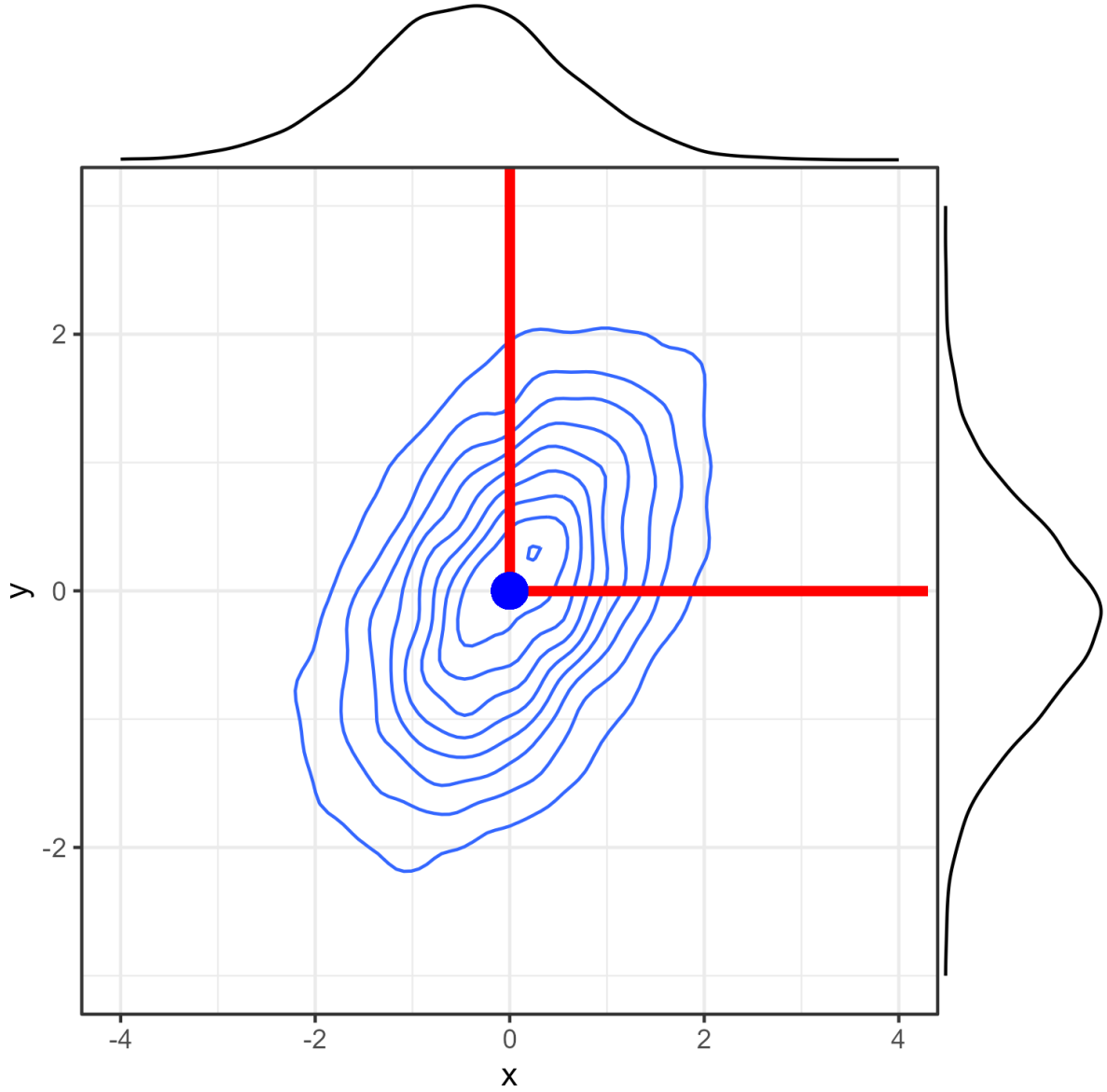
Uncertainty in economic evaluations

- Structural: scenario analysis
- Re-run the analysis with alternate assumptions and model structures
- Parameter: sensitivity analysis (SA)
- Re-run the analysis with different parameter values
- Type of sensitivity analysis:
  - One-way SA
  - Multi-way SA
  - Extreme values SA
  - Probabilistic SA

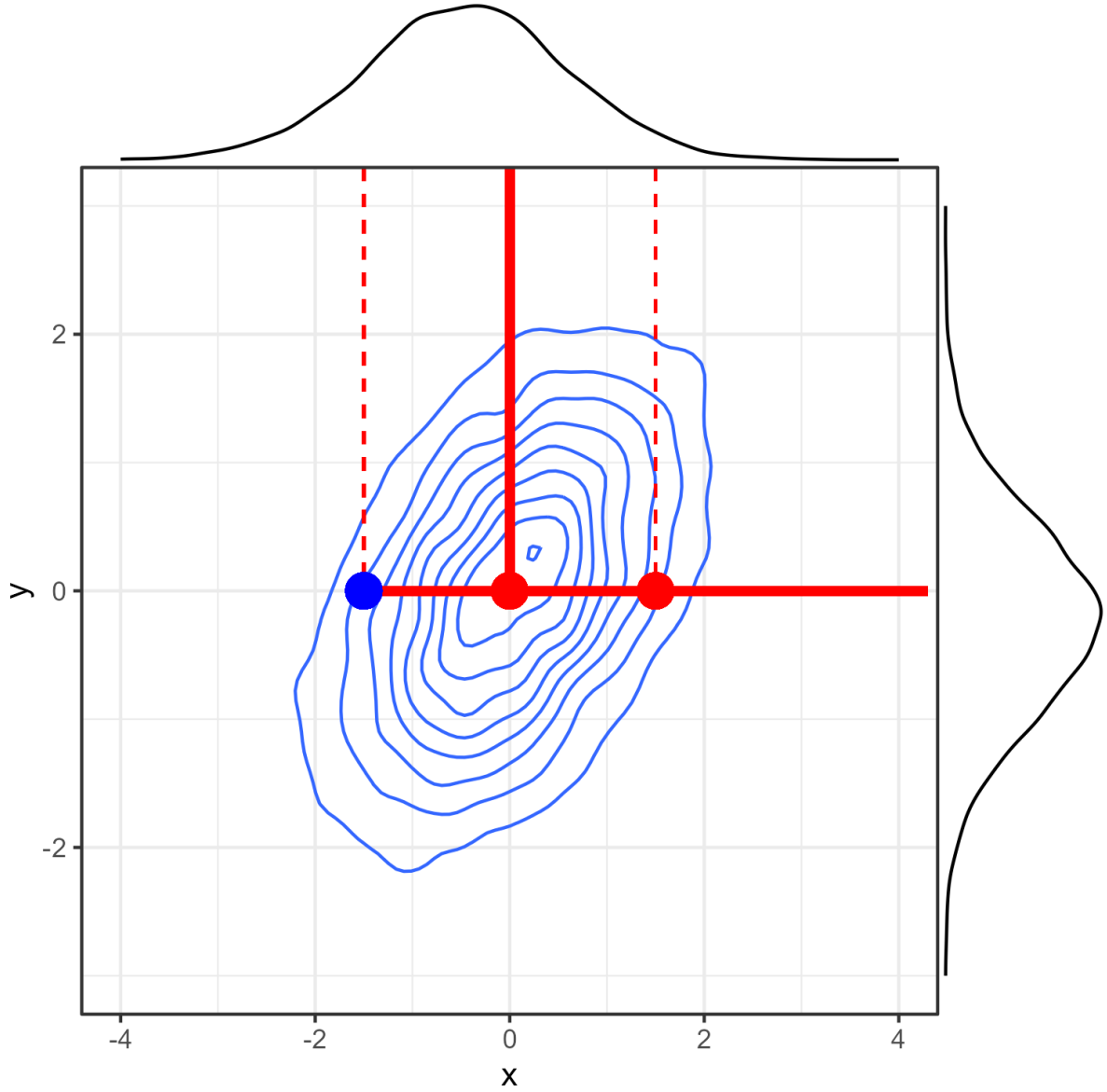
Types of sensitivity analysis

- Simple sensitivity analysis entails varying one or more of the components of an evaluation to see how it affects the results
- Probabilistic sensitivity analysis assigns ranges and distribution to variables and computer programs are used to select values at random from each range and to record the results
- By using these different methods of sensitivity analysis it is possible to show whether the results of a particular study over a range of assumptions or hinge on the accuracy of particular assumptions

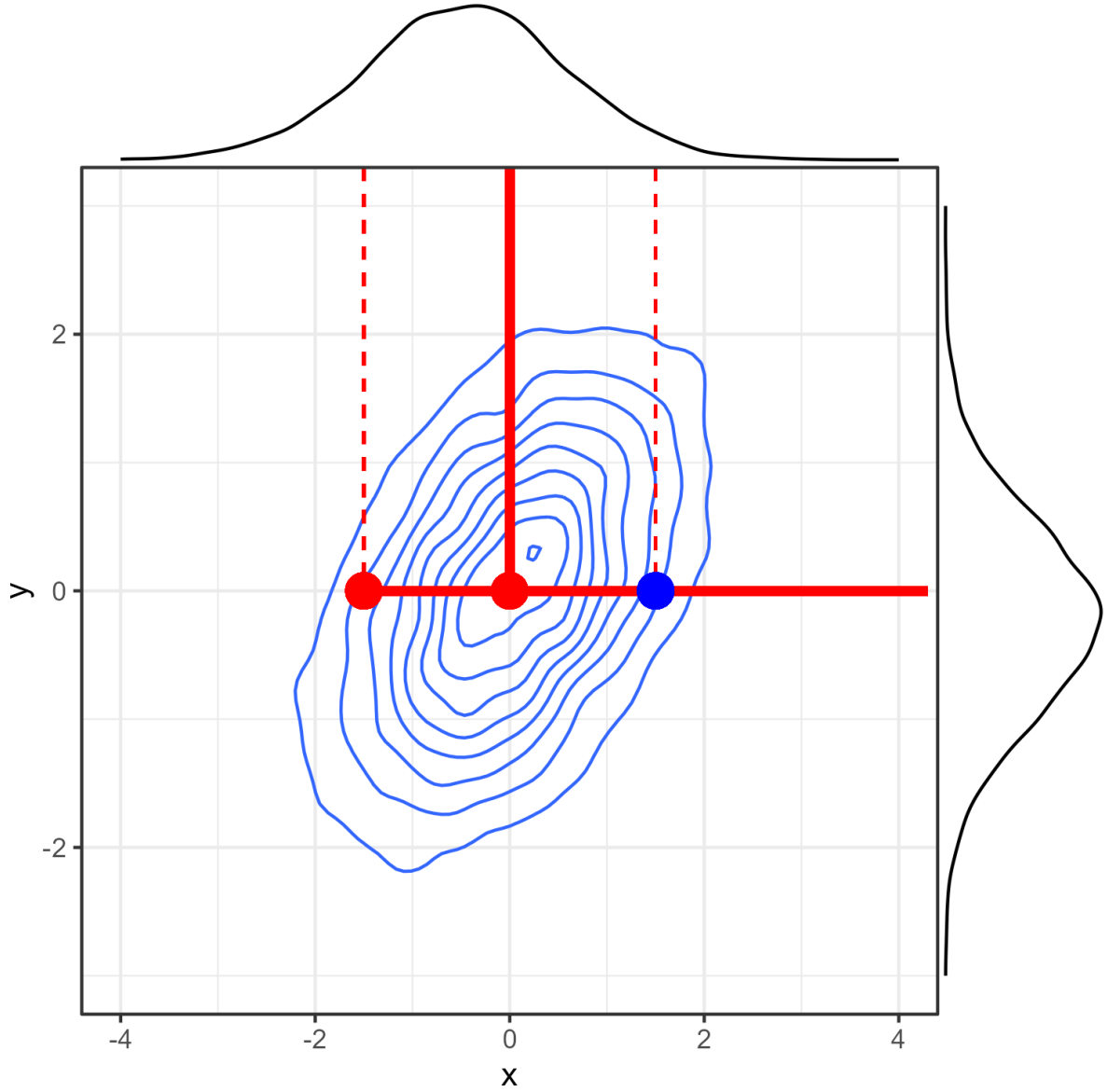
A two input example



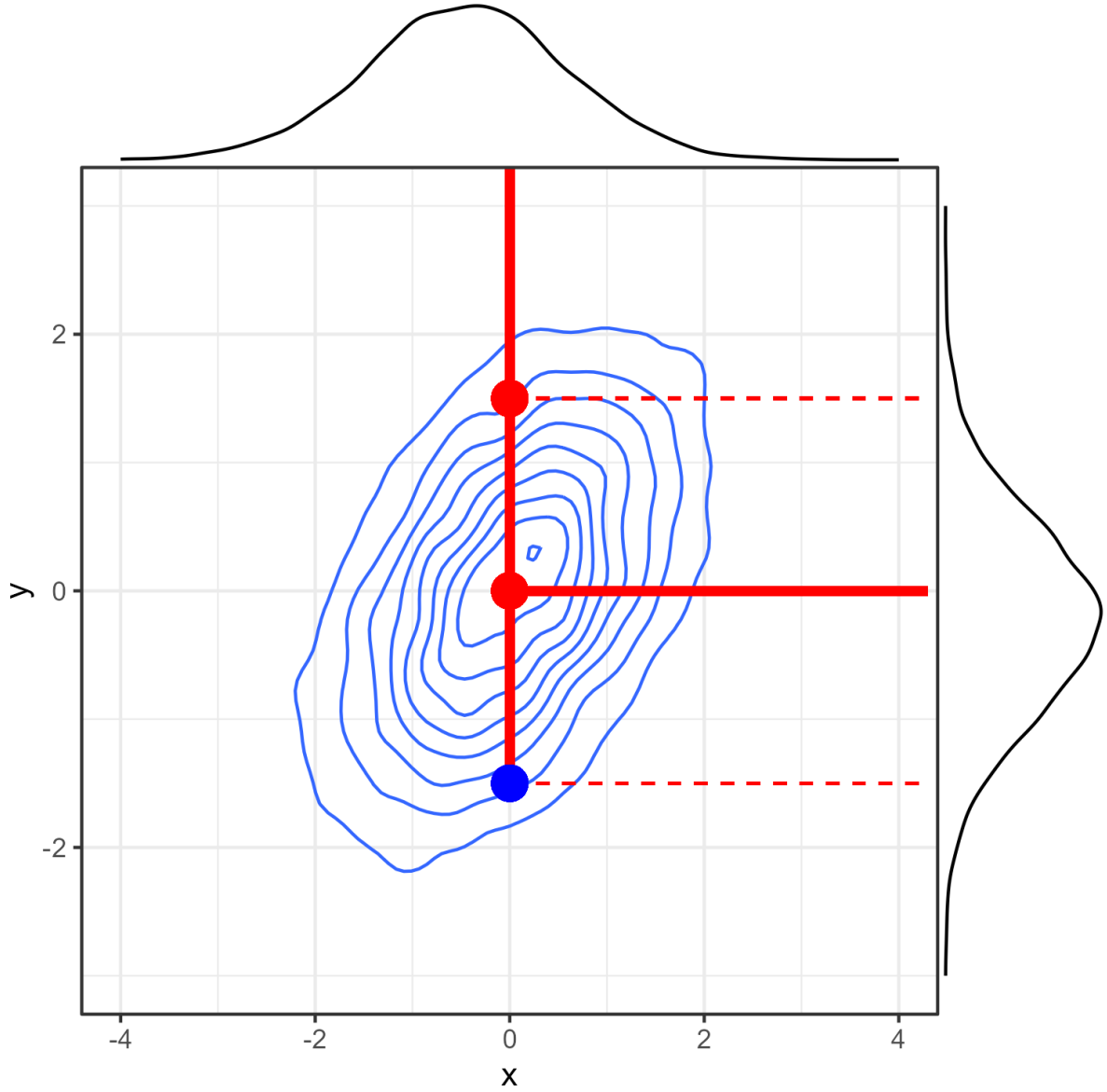
A two input example



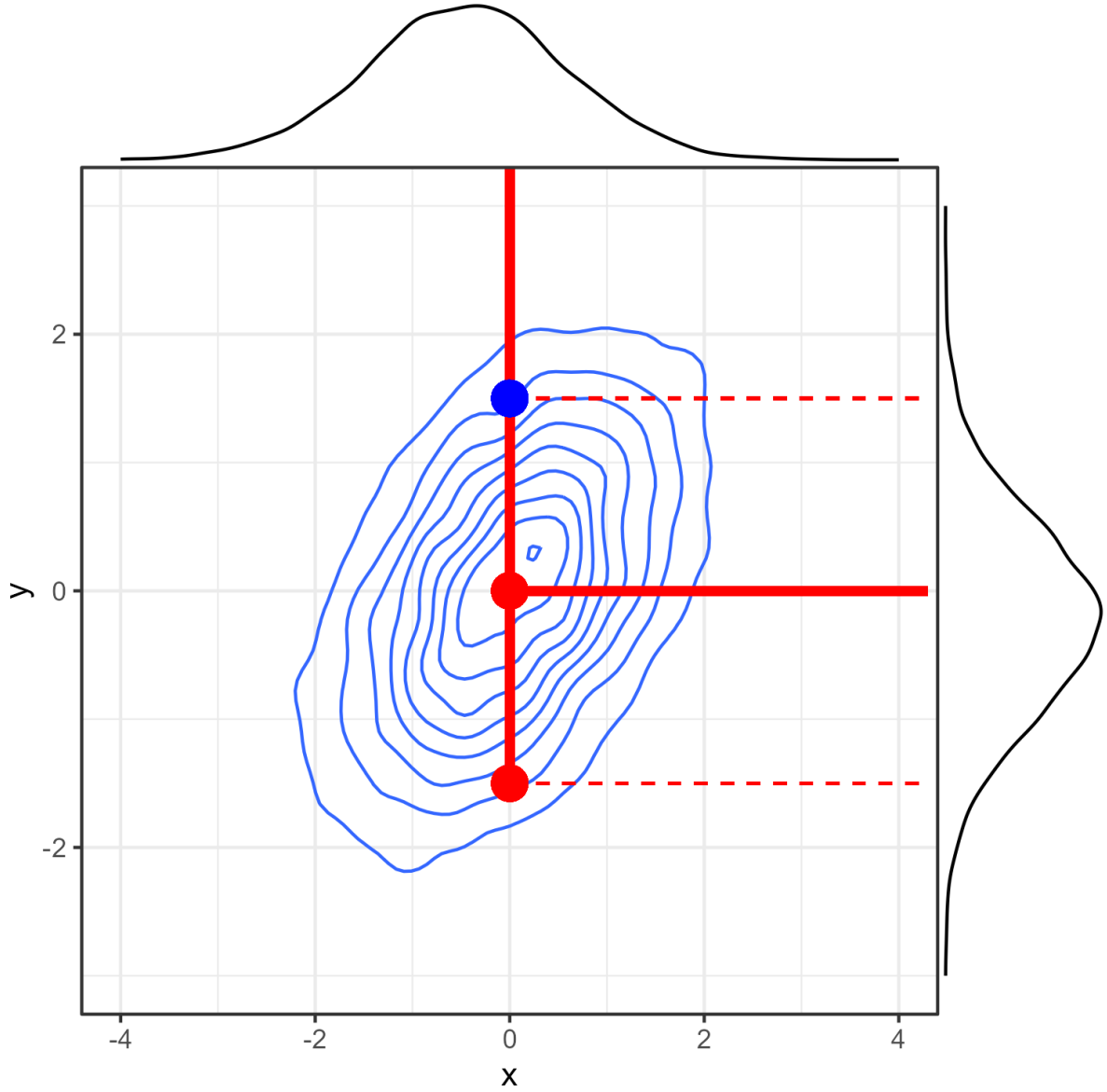
A two input example



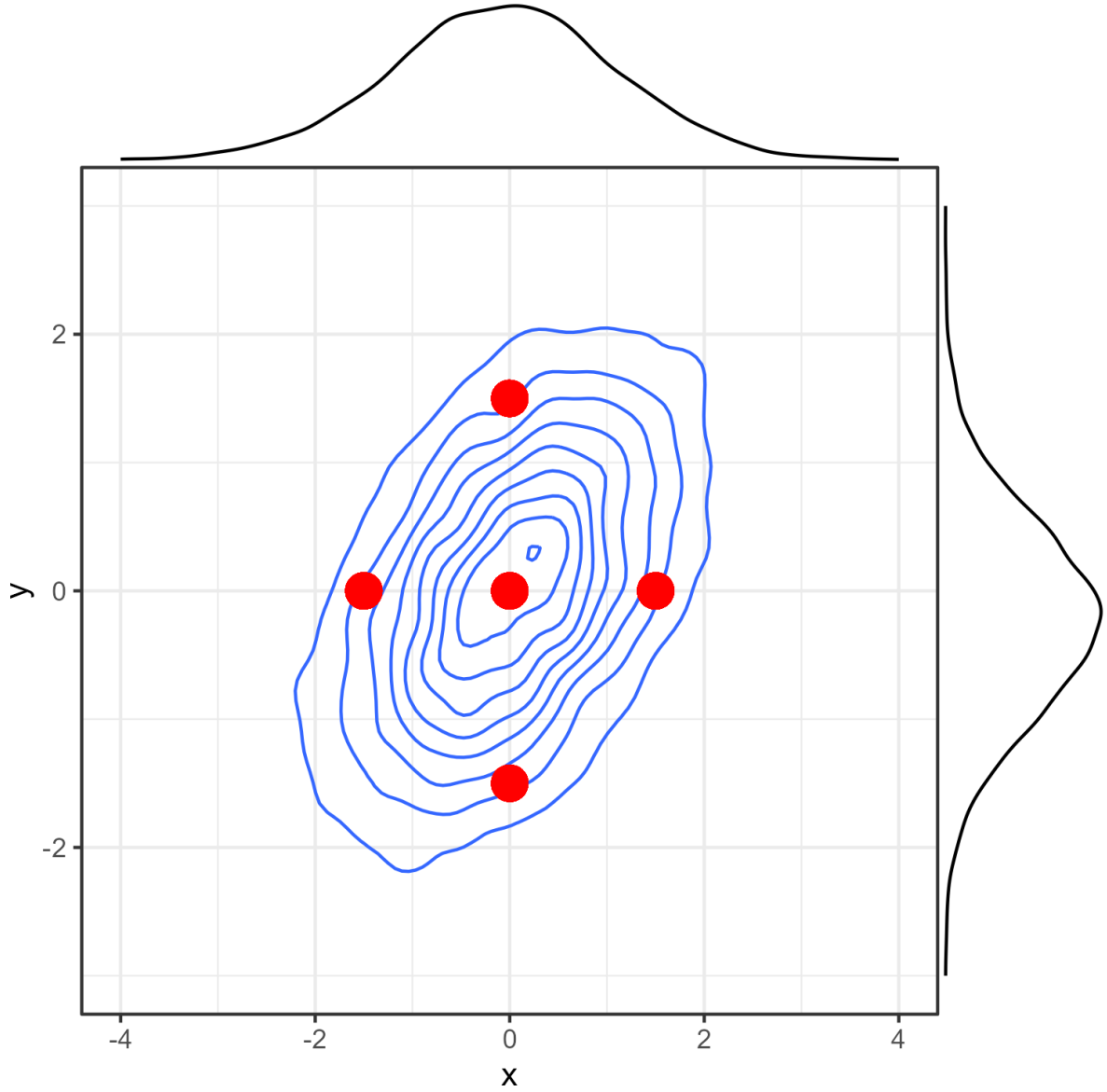
A two input example



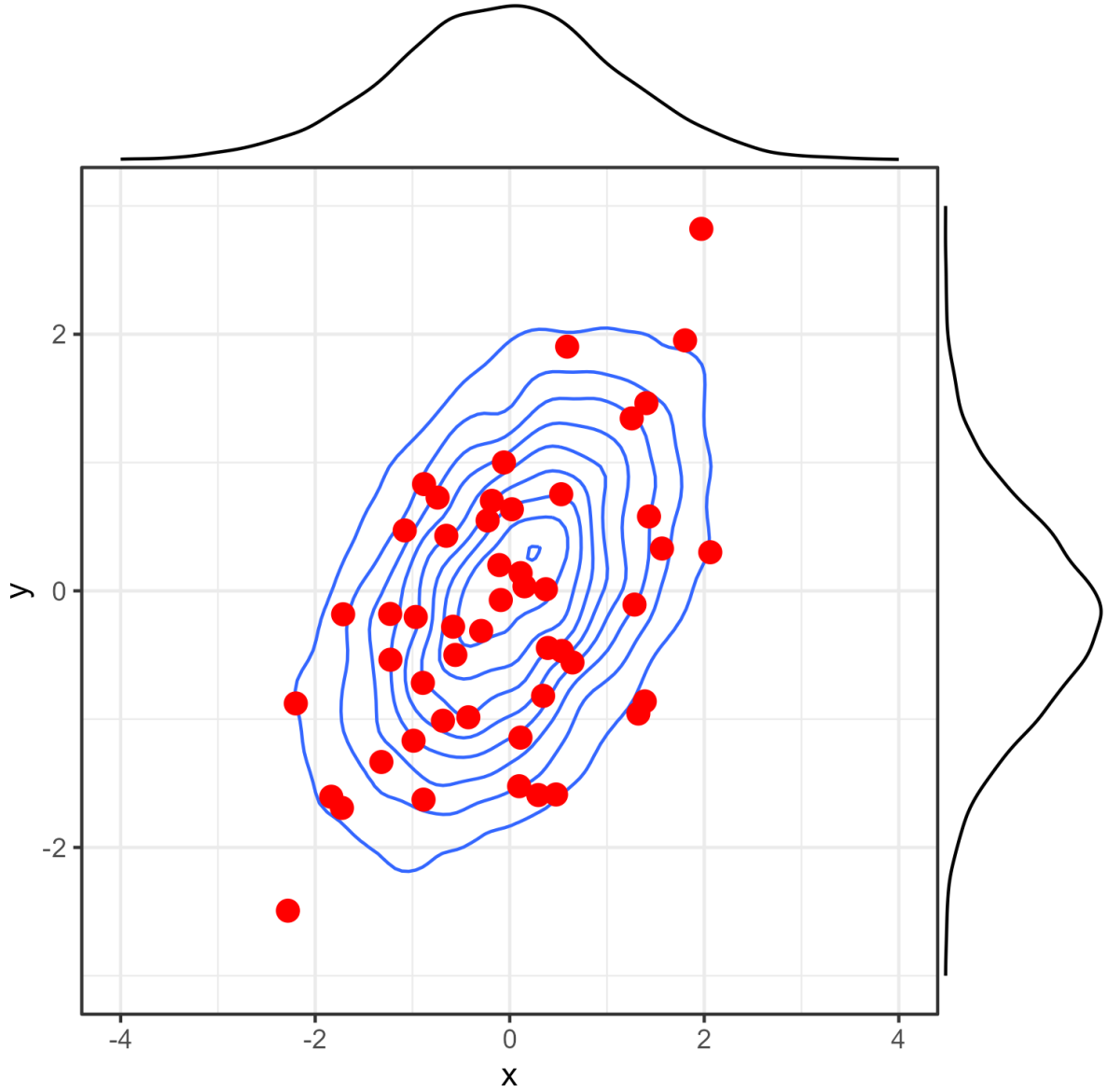
A two input example



A two input example

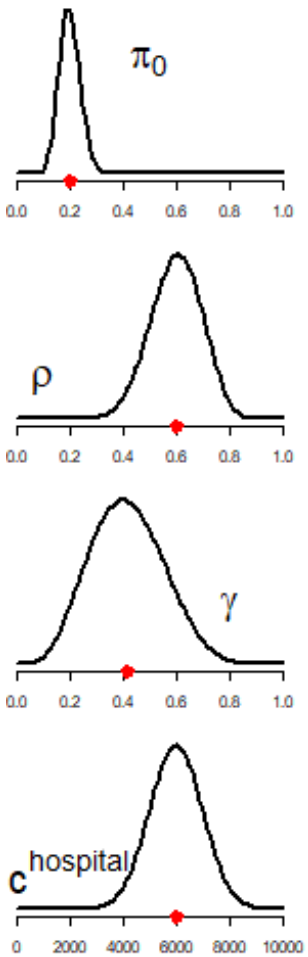


A two input example



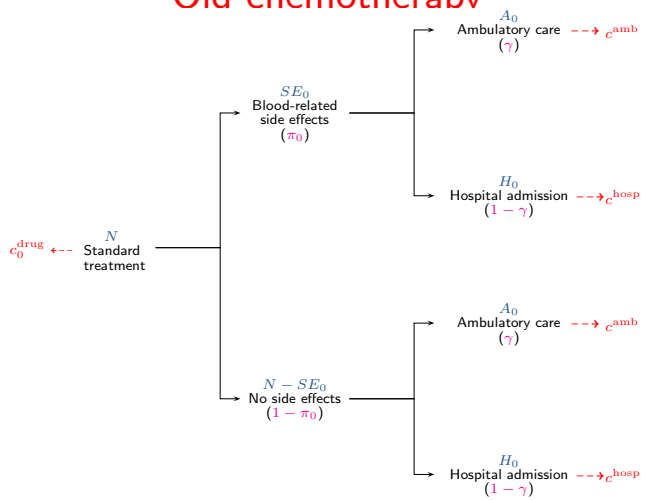
4. Uncertainty analysis

Parameters

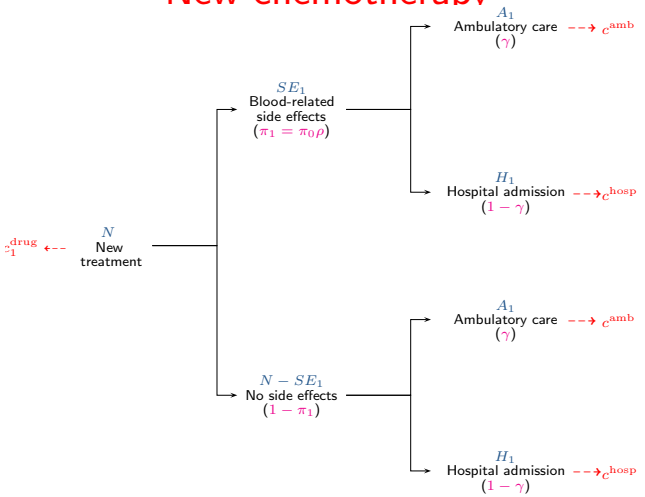


Model structure

Old chemotherapy



New chemotherapy



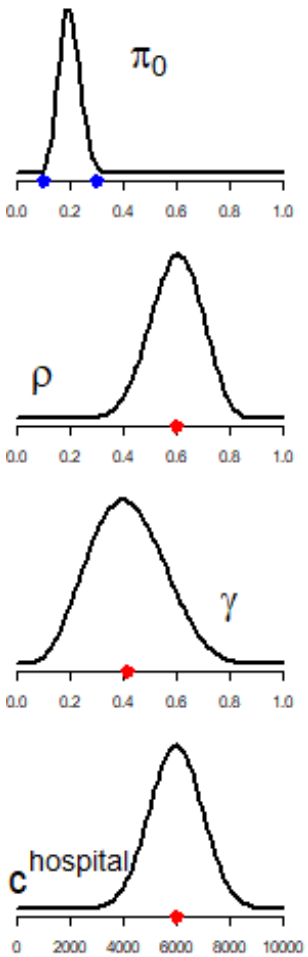
Decision analysis

Old chemotherapy	
Benefits	Costs

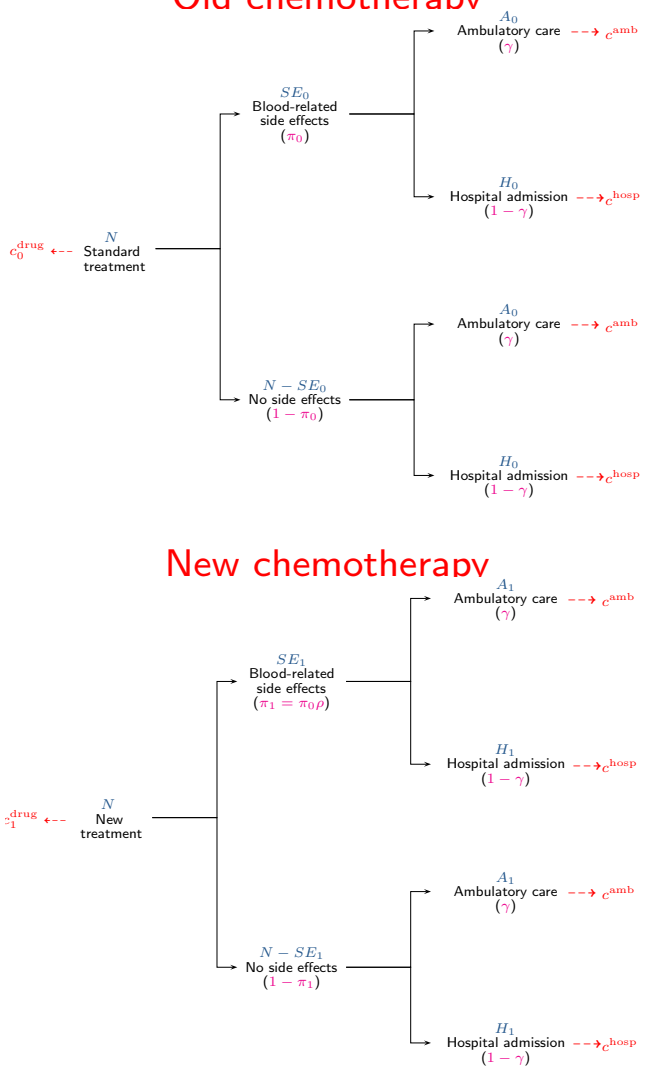
New chemotherapy	
Benefits	Costs

4. Uncertainty analysis

Parameters



Model structure



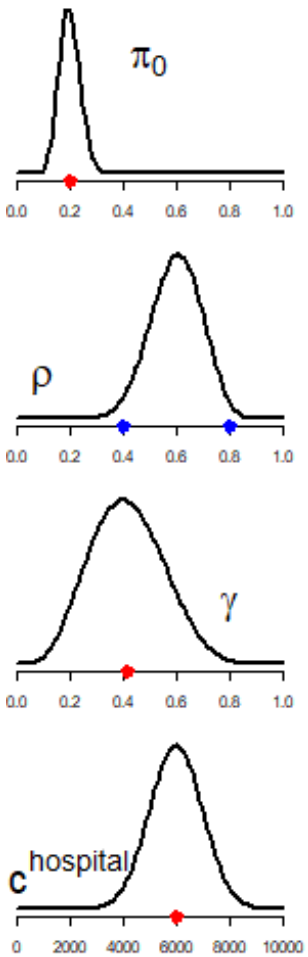
Decision analysis

Old chemotherapy	
Benefits	Costs

New chemotherapy	
Benefits	Costs

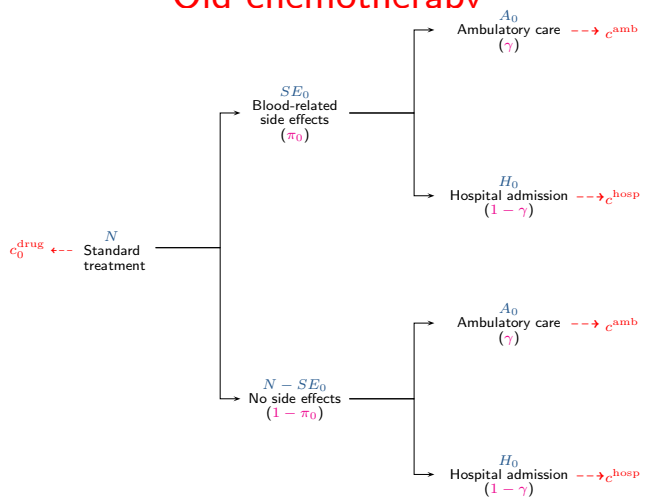
4. Uncertainty analysis

Parameters

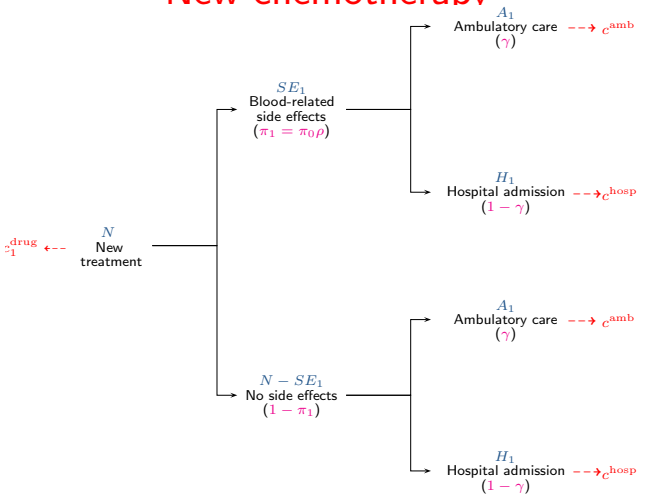


Model structure

Old chemotherapy



New chemotherapy



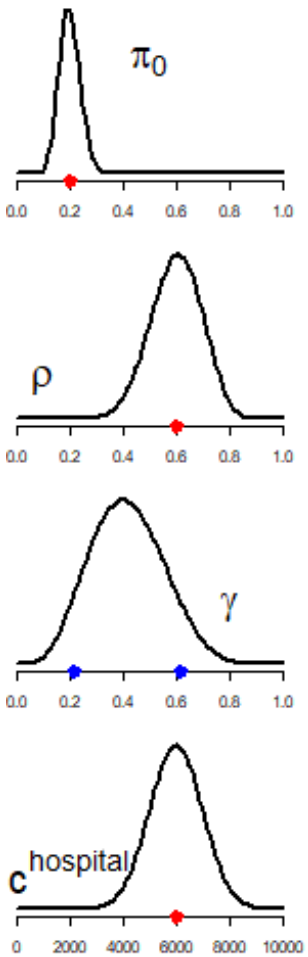
Decision analysis

Old chemotherapy	
Benefits	Costs

New chemotherapy	
Benefits	Costs

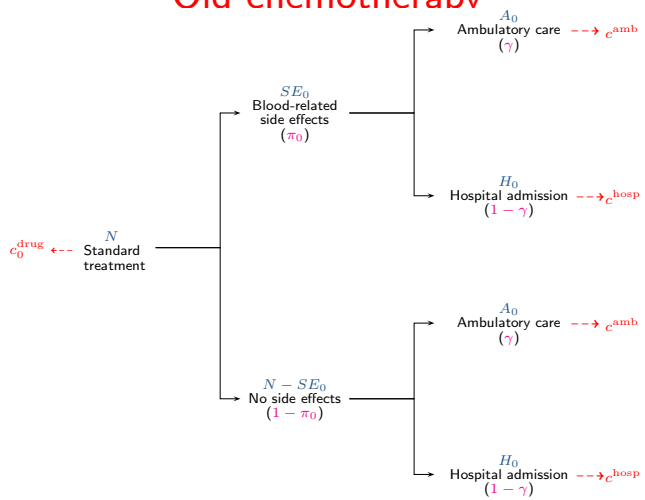
4. Uncertainty analysis

Parameters

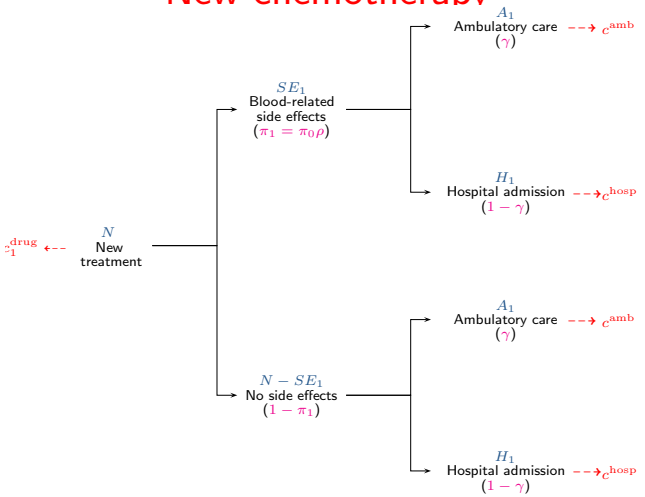


Model structure

Old chemotherapy



New chemotherapy



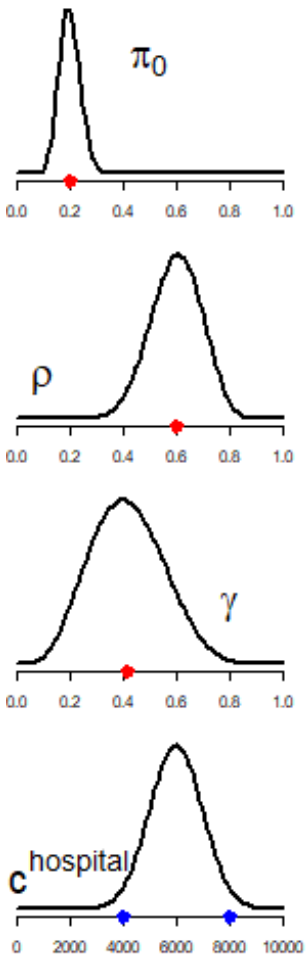
Decision analysis

Old chemotherapy	
Benefits	Costs

New chemotherapy	
Benefits	Costs

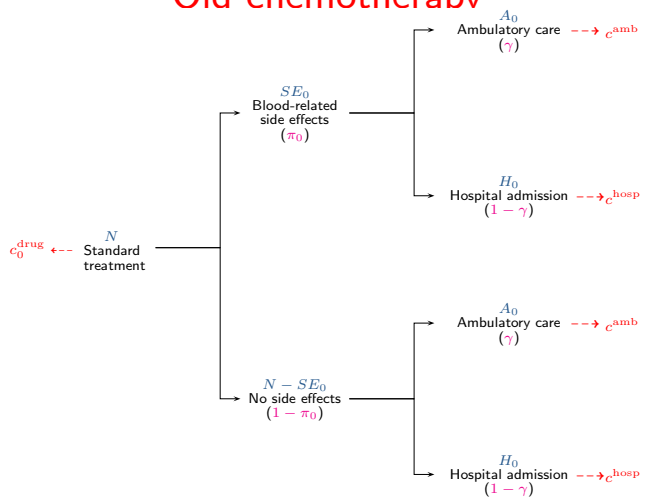
4. Uncertainty analysis

Parameters

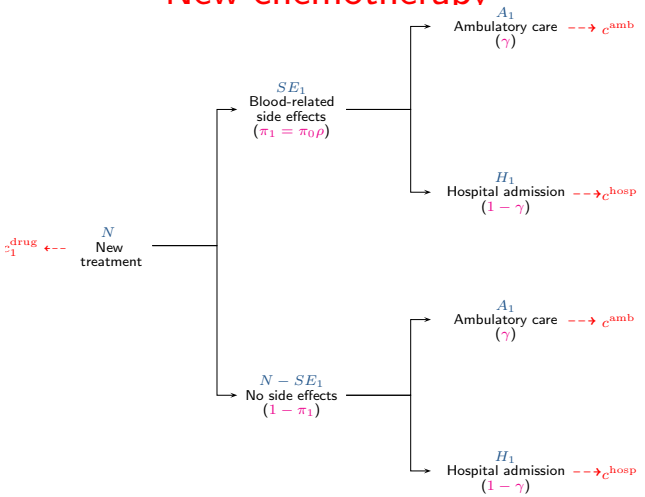


Model structure

Old chemotherapy



New chemotherapy



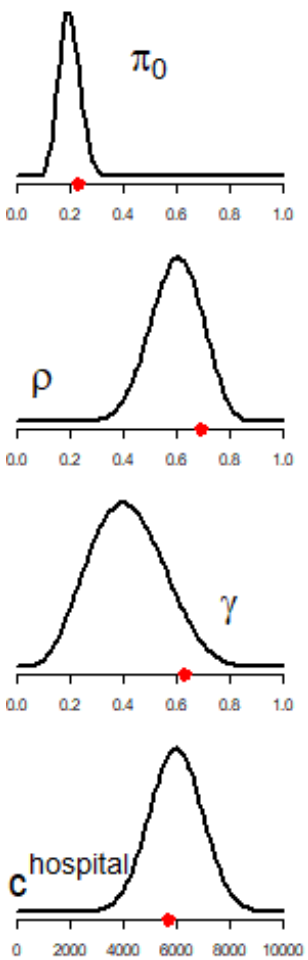
Decision analysis

Old chemotherapy	
Benefits	Costs

New chemotherapy	
Benefits	Costs

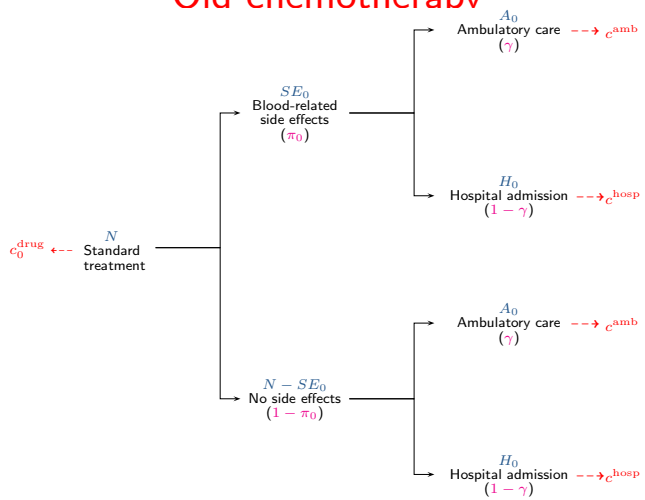
4. Uncertainty analysis

Parameters

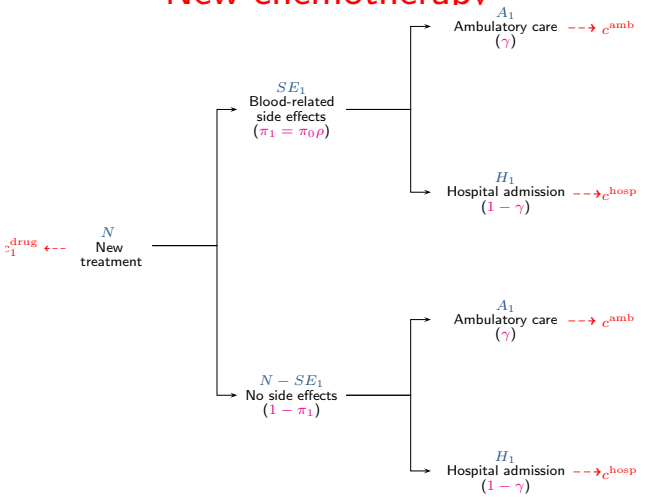


Model structure

Old chemotherapy



New chemotherapy



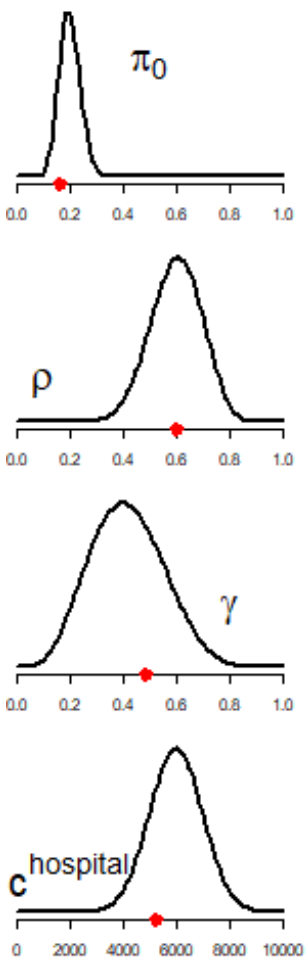
Decision analysis

Old chemotherapy	
Benefits	Costs
741	670 382.1

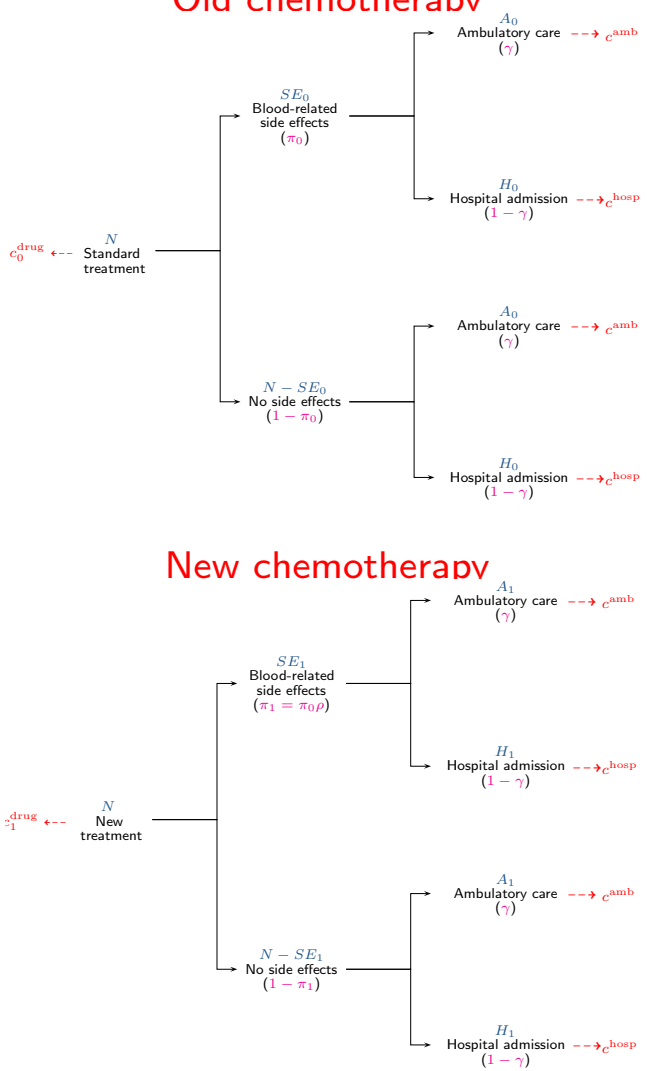
New chemotherapy	
Benefits	Costs
732	1 131 978

4. Uncertainty analysis

Parameters



Model structure



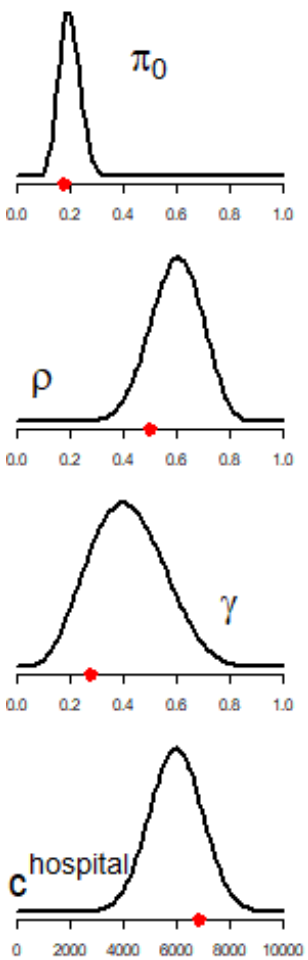
Decision analysis

Old chemotherapy	
Benefits	Costs
741	670 382.1
699	871 273.3

New chemotherapy	
Benefits	Costs
732	1 131 978
664	1 325 654

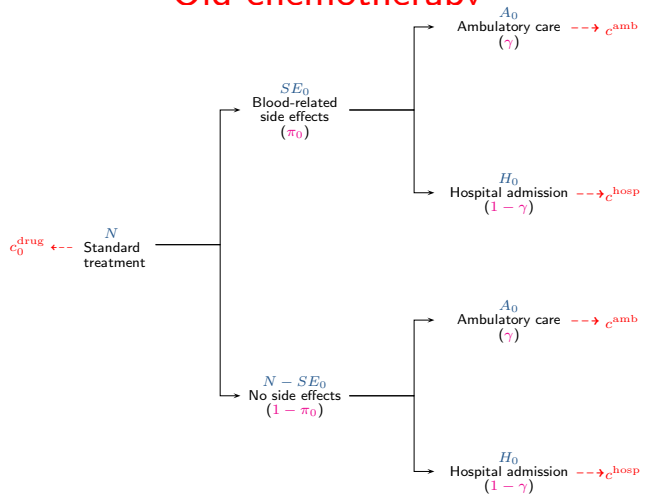
4. Uncertainty analysis

Parameters

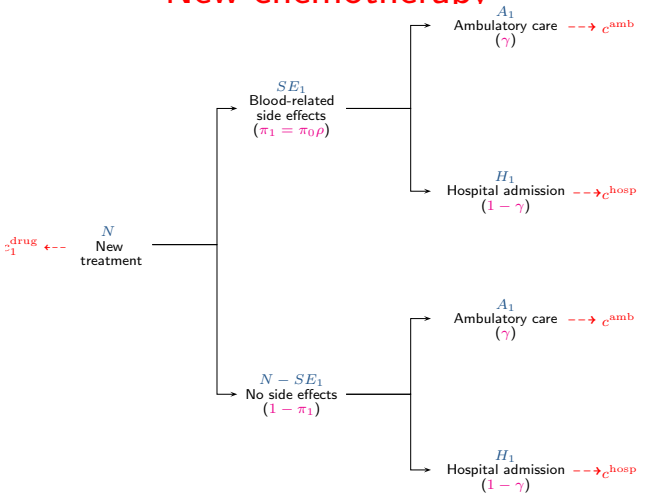


Model structure

Old chemotherapy



New chemotherapy



Decision analysis

Old chemotherapy	
Benefits	Costs
741	670 382.1
699	871 273.3
...	...
726	425 822.2
716.2	**790 381.2**

New chemotherapy	
Benefits	Costs
732	1 131 978
664	1 325 654
...	...
811	766 411.4
774.5	**1 066 849.8**

$$\text{ICER} = \frac{276\,468.6}{58.3} = 6\,497.1$$

Is this *all* we need?

(see VoI)

- The CEAC only deals with the **probability** of making the "right decision"
- But it does not account for the **payoff/penalty** associated with making the "wrong" one!

Is this *all* we need?

(see Vol)

- The CEAC only deals with the **probability** of making the "right decision"
- But it does not account for the **payoff/penalty** associated with making the "wrong" one!

- **Example 1**: Intervention $t = 1$ is the most cost-effective, given current evidence
  - $\Pr(t = 1 \text{ is cost-effective}) = 0.51$
  - If we get it wrong: Increase in costs = £3
    Decrease in effectiveness = 0.000001 QALYs
  - Large uncertainty/negligible consequences $\Rightarrow$ **can afford uncertainty**

Is this *all* we need?

(see Vol)

- The CEAC only deals with the **probability** of making the "right decision"
- But it does not account for the **payoff/penalty** associated with making the "wrong" one!

- **Example 1**: Intervention $t = 1$ is the most cost-effective, given current evidence
  - $\Pr(t = 1 \text{ is cost-effective}) = 0.51$
  - If we get it wrong: Increase in costs = £3
    Decrease in effectiveness = 0.000001 QALYs
  - Large uncertainty/negligible consequences ⇒ **can afford uncertainty**

- **Example 2**: Intervention $t = 1$ is the most cost-effective, given current evidence
  - $\Pr(t = 1 \text{ is cost-effective}) = 0.999$
  - If we get it wrong: Increase in costs = £1 000 000 000
    Decrease in effectiveness = 999999 QALYs
  - Tiny uncertainty/dire consequences ⇒ **probably should think about it...**