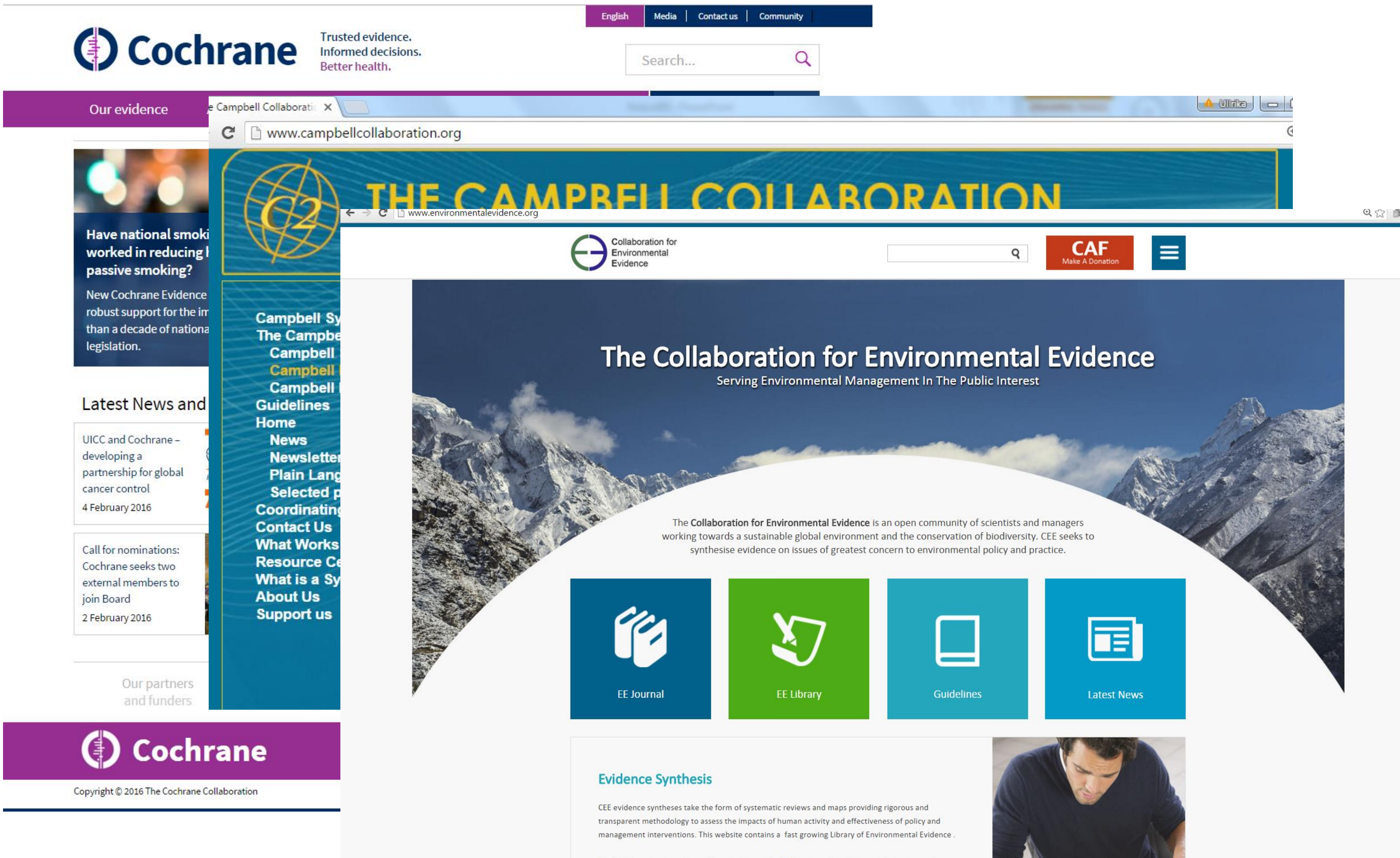
Robust Evidence Synthesis

Ullrika Sahlin

Thursday 9.00-12.30

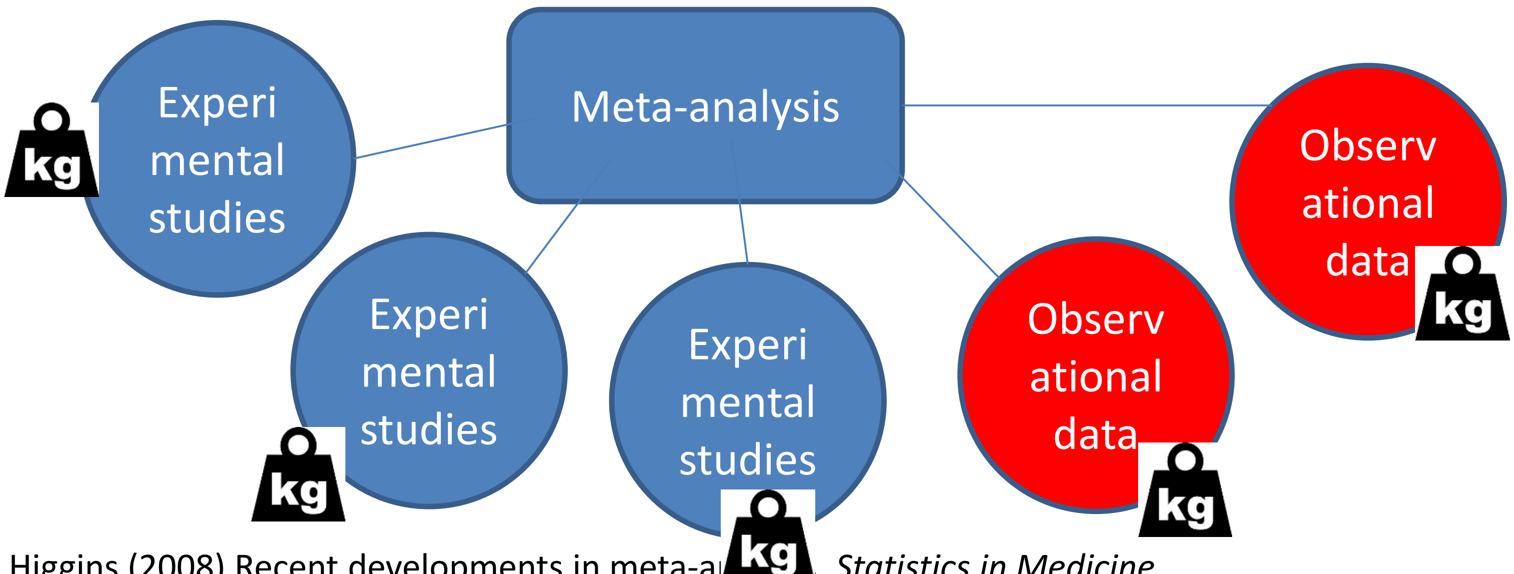
Evidence-based



Meta-analysis

Statistical technique to combine results from multiple independent studies

Consider differences in quality in studies



Sutton & Higgins (2008) Recent developments in meta-analysis. *Statistics in Medicine*

Sutton & Abrams (2001). Bayesian methods in meta-analysis and evidence synthesis. *Statistical Methods in Medical Reserach*

Weed (2005) Weight of Evidence: A review of Concept and Methods. *Risk Analysis*

Meta-analysis

Charnley

Stanmore

Table IV. Summary of evidence on revision hazards for Charnley and Stanmore prostheses: hazard ratios < 1 are in favour of Stanmore.

Source	Charnley: Number of patients	Charnley: Revision rate	Stanmore: Number of patients	Stanmore: Revision rate	Estimated hazard ratio (HR)	(95% int.)
					Fixed-effects model	
Registry	28 525	5.9%	865	3.2%	0.55	(0.37–0.77)
RCT	200	3.5%	213	4.0%	1.34	(0.45–3.46)
Case series	208	16.0%	982	7.0%	0.44	(0.28–0.66)

Spiegelhalter and Best (2003). Bayesian approaches to mulitple sources of evidence and uncertainty in complex cost-effectiveness modelling. *Statist. Med.*

COLLABORATION FOR ENVIRONMENTAL EVIDENCE

Systematic reviews for conservation and environmental management

CONTEXT

In cities, climate change may increase human exposure to high temperatures (including heat waves), **ground-level ozone and ultra-violet even more than in surrounding countryside.**

Could this be mitigated by greening urban areas (increasing the abundance and cover of vegetation)?

This question was addressed by a systematic review of the accessible scientific referenced and grey literature. The systematic review takes into account the quality of the research and possible biases, in order to provide a rigorous, transparent, replicable and updatable review of the scientific evidence .

From the CEE Library

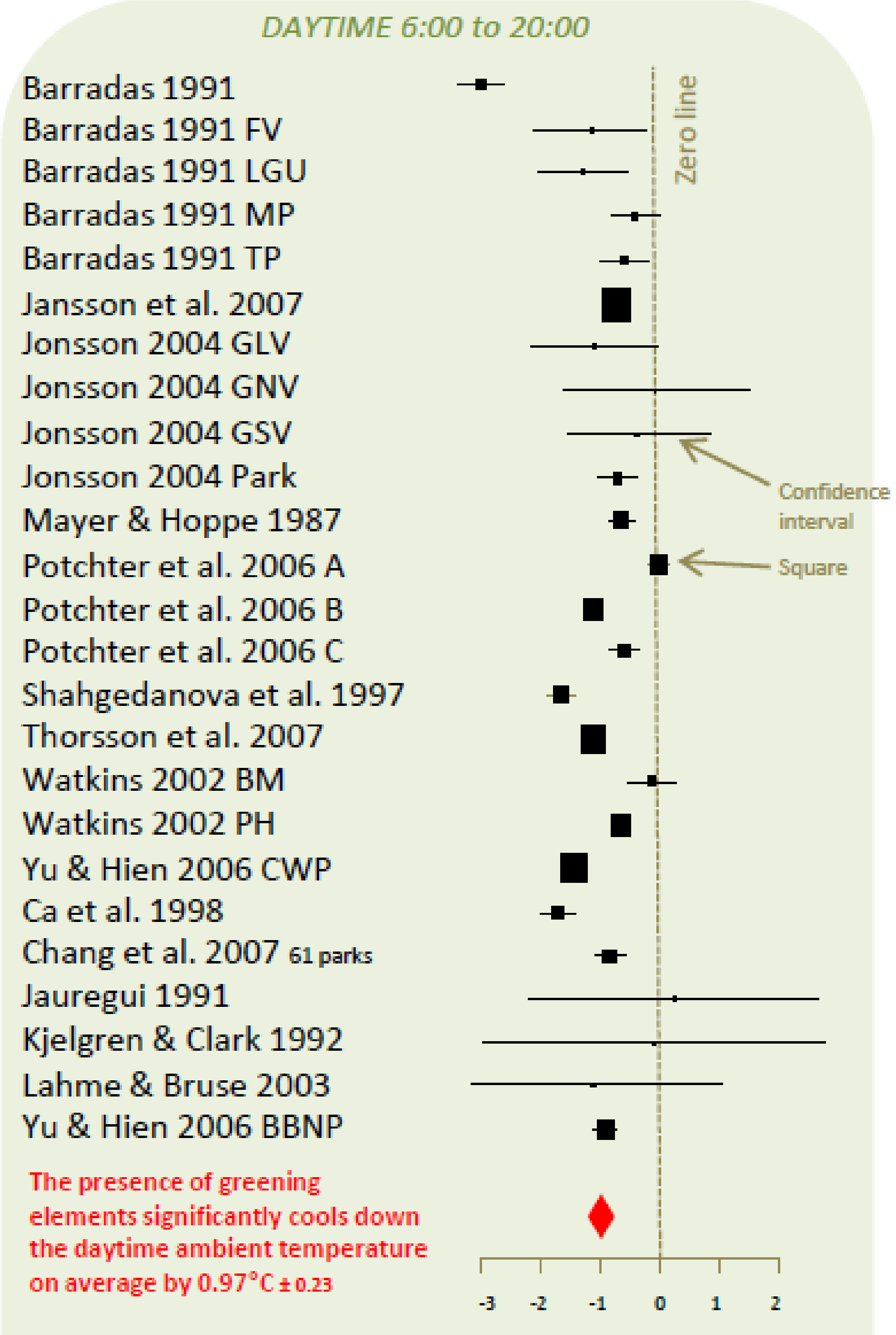
GREENING CITIES TO MITIGATE IMPACTS of CLIMATE CHANGE

DINGS

On average, a park is 1°C cooler than a built-up area

Many factors can moderate this difference, such as the park size, proportion of paved areas, wind, irrigation, season and latitude, weather and surroundings.

3 studies report that the cooling effect extended beyond the boundaries of the park or trees, further studies are needed to confirm this result.



http://www.environmentalevidence.org/completed-reviews/how-effective-is-greening-of-urban-areas-in-reducing-human-exposure-to-ground-level-ozone-concentrations-uv-exposure-and-the-urban-heat-island-effect

GRADE

Underlying methodology	Quality rating
Randomized trials; or double-upgraded observational studies.	High
Downgraded randomized trials; or upgraded observational studies.	Moderate
Double-downgraded randomized trials; or observational studies.	Low
Triple-downgraded randomized trials; or downgraded observational studies; or case series/case reports.	Very low

Other quality dimensions in the GRADE system

- Inconsistency
- Indirectness
- Publication bias
- Imprecision

Imprecision – not what you think – but almost

ELSEVIER

Journal of Clinical Epidemiology 64 (2011) 1283–1293

Journal of Clinical Epidemiology

GRADE guidelines 6. Rating the quality of evidence—imprecision

Gordon H. Guyatt[a,b,*], Andrew D. Oxman[c], Regina Kunz[d,e], Jan Brozek[a], Pablo Alonso-Coello[f], David Rind[g], PJ Devereaux[a], Victor M. Montori[h], Bo Freyschuss[i], Gunn Vist[c], Roman Jaeschke[b], John W. Williams Jr.[j], Mohammad Hassan Murad[h], David Sinclair[k], Yngve Falck-Ytter[l], Joerg Meerpohl[m,n], Craig Whittington[o], Kristian Thorlund[a], Jeff Andrews[p], Holger J. Schünemann[a,b]

[a]*Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario L8N 3Z5, Canada*
[b]*Department of Medicine, McMaster University, Hamilton, Ontario L8N 3Z5, Canada*
[c]*Norwegian Knowledge Centre for the Health Services, PO Box 7004, St Olavs plass, 0130 Oslo, Norway*

From Guyatt et al

Confidence intervals capture the extent of imprecision – mostly

To a large extent, CIs inform the impact of random error on evidence quality. Within the frequentist (in contrast to Bayesian) framework, the CI represents that range of results which, were an experiment repeated numerous times and the CI recalculated for each experiment, a particular proportion of the CIs (typically 95%), would include the true underlying value.

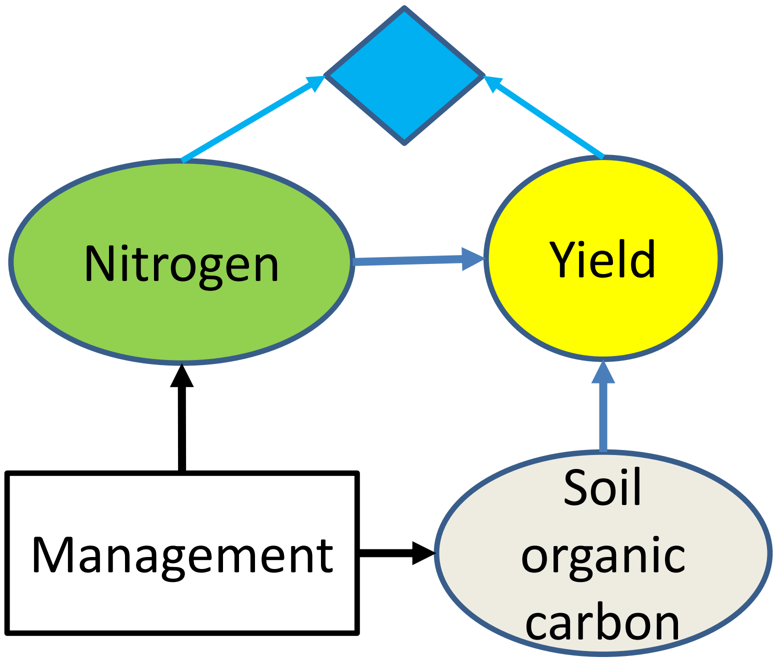
Conceptually easier than this defintion is to think of the CI as the range in which the trugh plausibility lies.

From Guyatt et al

When considering the quality of evidence, the issue is whether the CI around the estimate of treatment effect is sufficiently narrow. If it is not, we rate down the evidence quality by one level.

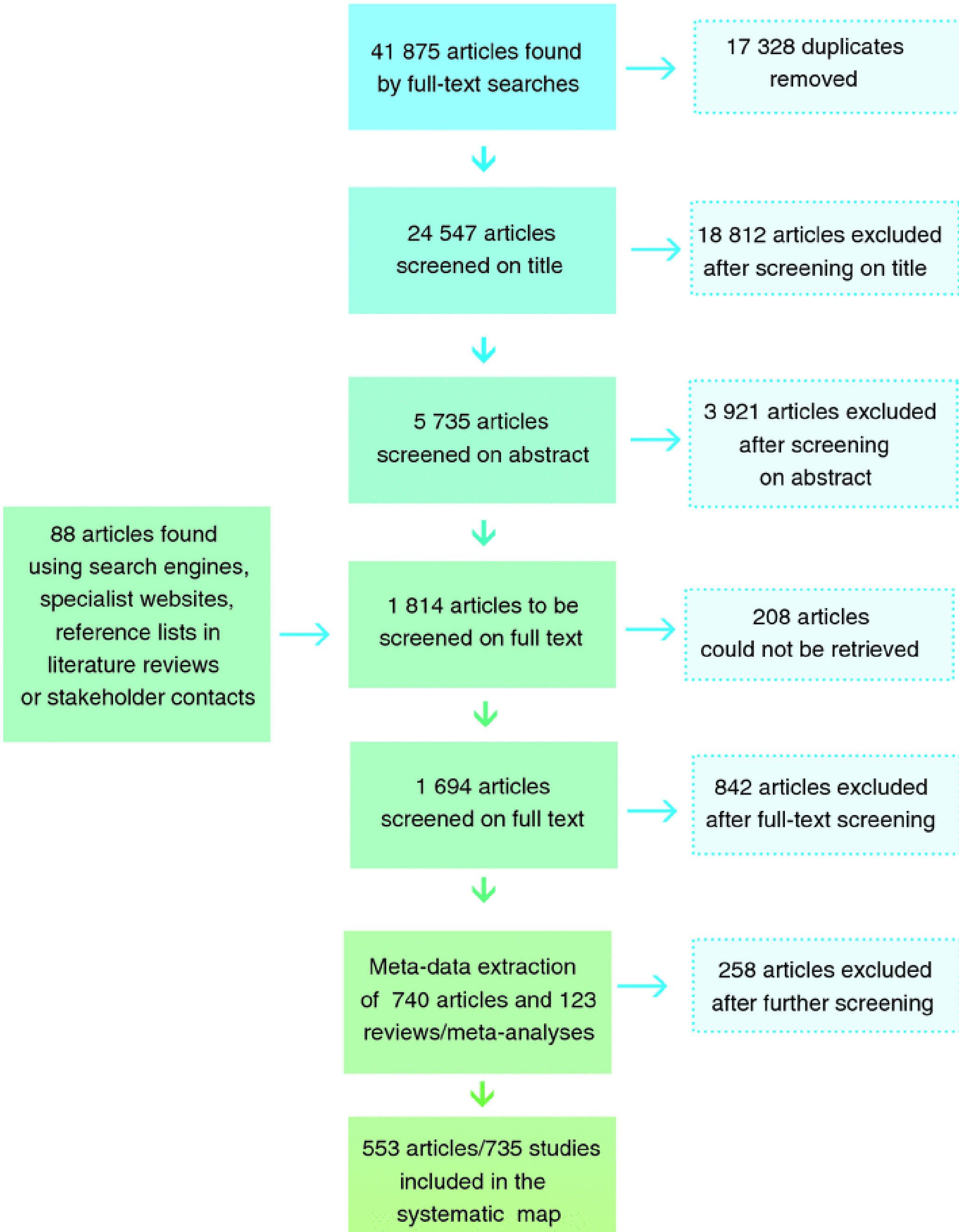
Even if CIs appear satisfactorily narrow, when effects are large and both sample size and number of events are modest, consider the rating down for imprecision.

Example of Evidence Synthesis – managing the soil capital



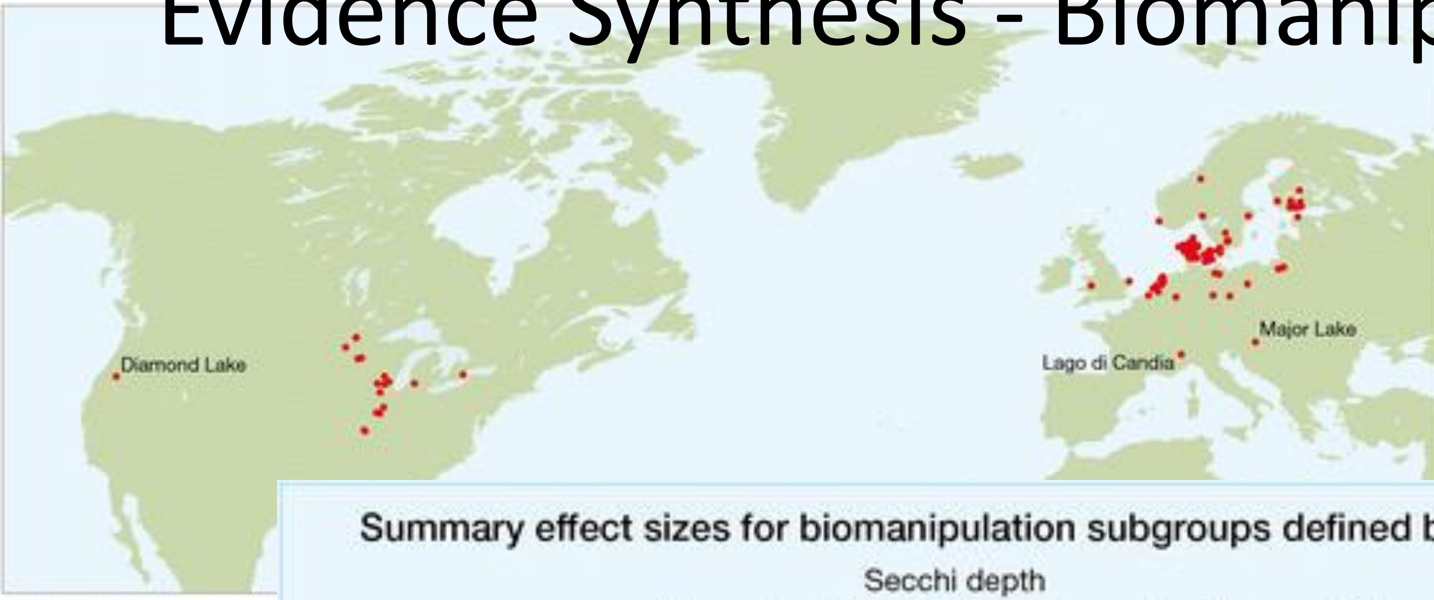
Link to ongoing systematic review

Which in-field interventions work to increase soil organic carbon?

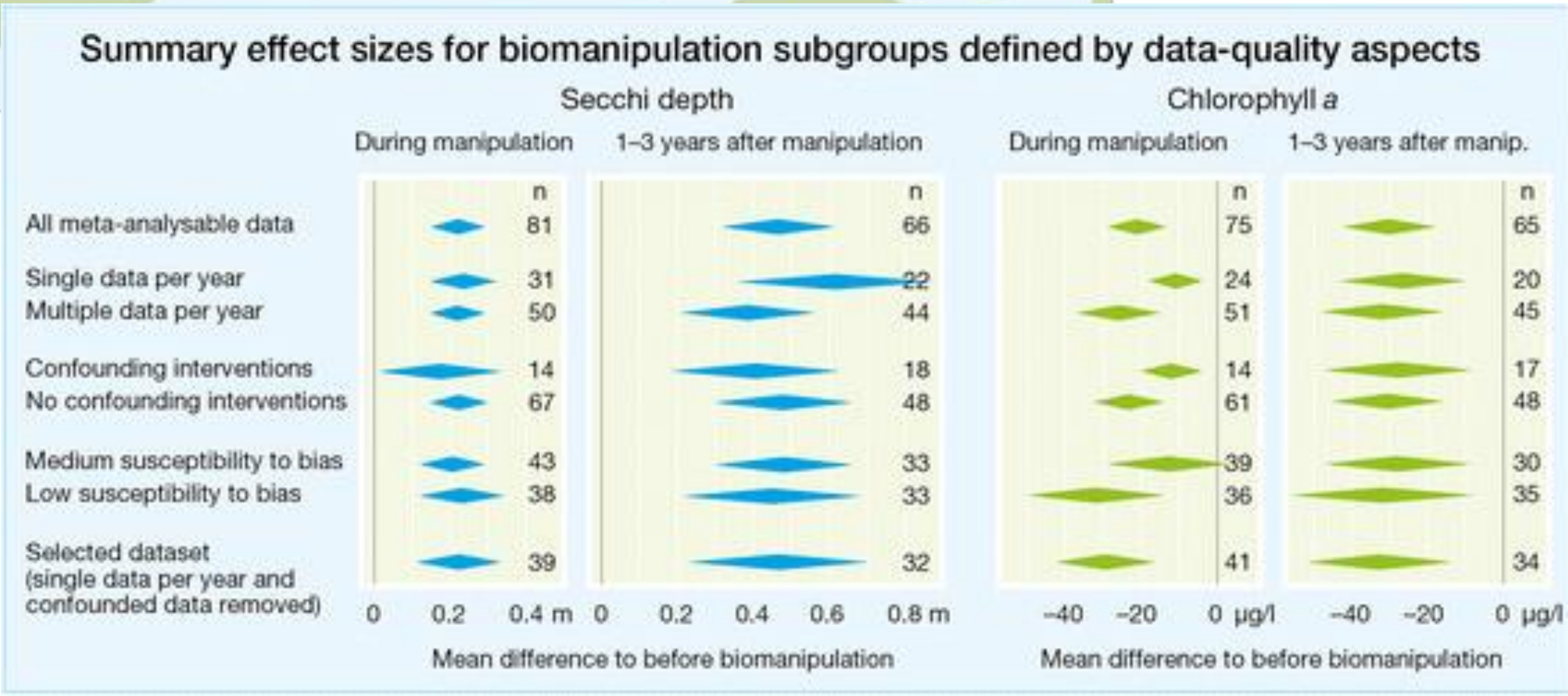


A systematic review starts with a careful literature search

Example of meta-analysis in an Evidence Synthesis - Biomanipulation



link



Back to modeling
Bayesian Evidence Synthesis

STATISTICS IN MEDICINE
Statist. Med. 2003; **22**:3687–3709 (DOI: 10.1002/sim.1586)

Bayesian approaches to multiple sources of evidence and uncertainty in complex cost-effectiveness modelling

David J Spiegelhalter[1,†] and Nicola G Best[2,*,‡,§]

[1] *MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge CB2 2SR, U.K.*
[2] *Department of Epidemiology and Public Health, Imperial College Faculty of Medicine, St. Mary's Campus, Norfolk Place, London W2 1PG, U.K.*

Bayesian Evidence Synthesis

1. **Complex cost-effectiveness models**, in particular discrete-state discrete-time Markov models, which are being increasingly used to make predictions of the consequences of a particular intervention

2. **Probabilistic sensitivity analysis** in cost-effectiveness, in which distributions are put over uncertain parameters

3. **Bayesian approaches** to cost-effectiveness, in particular using Markov chain Monte Carlo (MCMC) methods, to incorporate evidence from a single source (e.g. data arising from a clinical trial) with appropriate propagation of parameter uncertainty;

4. The synthesis of evidence from multiple sources in a form of generalized **meta-analysis**. There will usually be insufficient randomized evidence to fully inform a model that takes into account long-term consequences of an intervention. A generalized synthesis would allow the use of evidence from studies of different designs, possibly including the controversial practice of combining randomized and non-randomized evidence.

Spiegelhalter and Best (2003). Bayesian approaches to mulitple sources of evidence and uncertainty in complex cost-effectiveness modelling. *Statist. Med.*

BES – the statistical model

Charnley

Stanmore

Table IV. Summary of evidence on revision hazards for Charnley and Stanmore prostheses: hazard ratios < 1 are in favour of Stanmore.

	Charnley		Stanmore		Estimated hazard ratio	
Source	Number of patients	Revision rate	Number of patients	Revision rate	(HR)	(95% int.)
					Fixed-effects model	
Registry	28 525	5.9%	865	3.2%	0.55	(0.37–0.77)
RCT	200	3.5%	213	4.0%	1.34	(0.45–3.46)
Case series	208	16.0%	982	7.0%	0.44	(0.28–0.66)

Spiegelhalter and Best (2003). Bayesian approaches to mulitple sources of evidence and uncertainty in complex cost-effectiveness modelling. *Statist. Med.*

BES – the system model

D. J. SPIEGELHALTER AND N. G. BEST

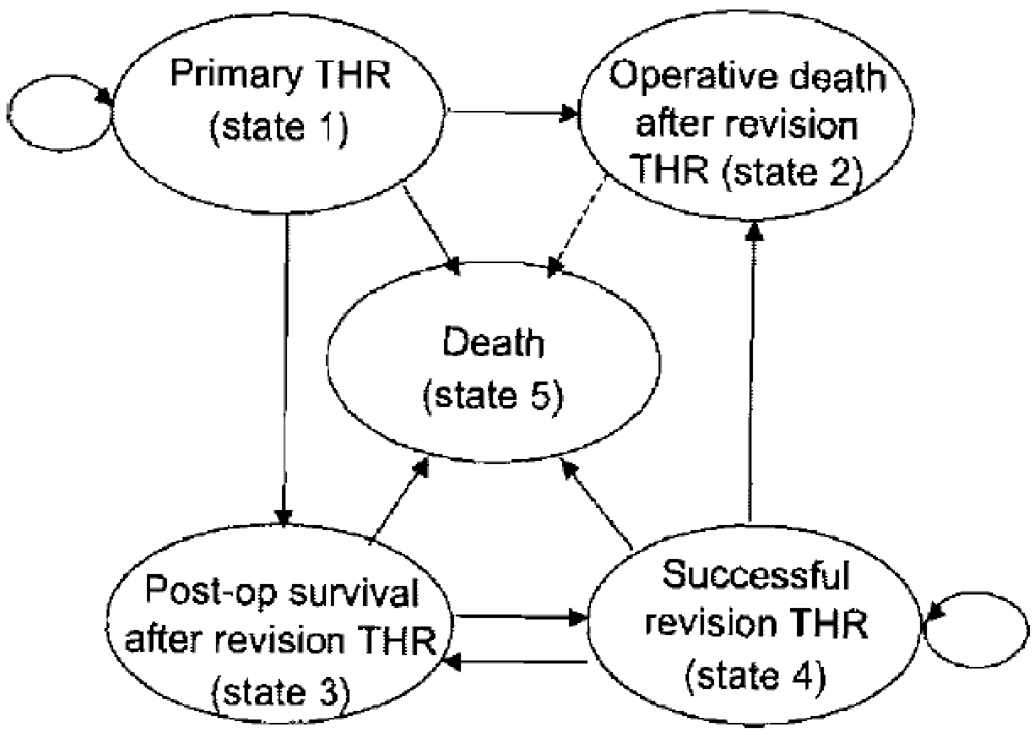


Figure 1. Markov model for outcomes following primary total hip replacement.

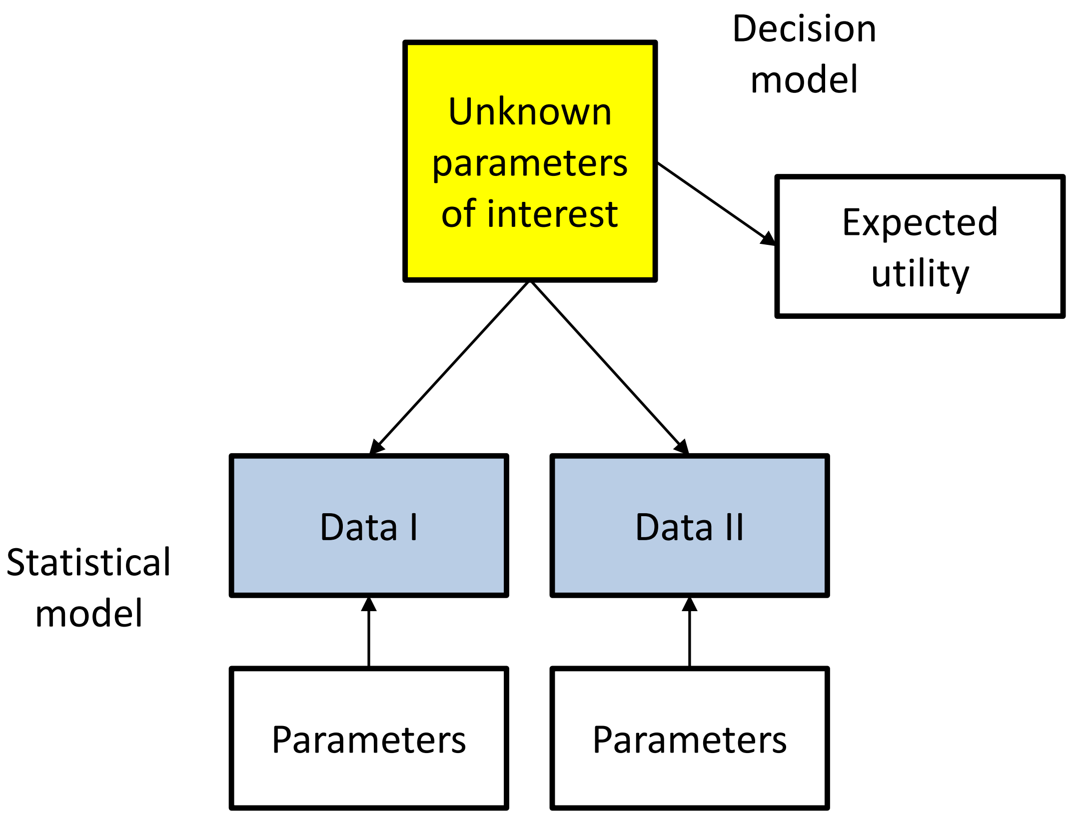
BES – the decision analysis

Table V. Summary of results of comparative analysis of cost-effectiveness for a hypothetical alternative versus the Charnley prostheses, using quality weights of [0.5, 1, 0.2] for weighting the registry, RCT and case study evidence, respectively.

	IC_θ (£)		IQ_θ (QALYs)		ICER		
Subgroup	Mean	SD	Mean	SD	Median	$Q(6000)$	$Q(10\,000)$
Men							
35–44 yr	−90	256	0.136	0.063	−846	0.92	0.94
45–54 yr	−28	216	0.113	0.053	−457	0.91	0.93
55–64 yr	71	156	0.081	0.038	581	0.87	0.92
65–74 yr	216	75	0.038	0.018	5190	0.55	0.77
75–84 yr	279	40	0.020	0.009	13 220	0.04	0.26
>84 yr	303	26	0.013	0.006	21 830	0.00	0.02
Women							
35–44 yr	−63	238	0.127	0.059	−691	0.91	0.94
45–54 yr	−14	206	0.109	0.051	−349	0.90	0.93
55–64 yr	66	161	0.083	0.039	537	0.87	0.92
65–74 yr	209	79	0.040	0.019	4710	0.60	0.80
75–84 yr	274	43	0.021	0.010	12 030	0.07	0.34
>84 yr	297	28	0.015	0.007	18 790	0.00	0.06
Overall	183	90	0.048	0.022	3246	0.73	0.85

Markov model using the model for evidence synthesis based on comparison of Charnley and Stanmore revision rates described in Section 6.2.

BES – integrated model



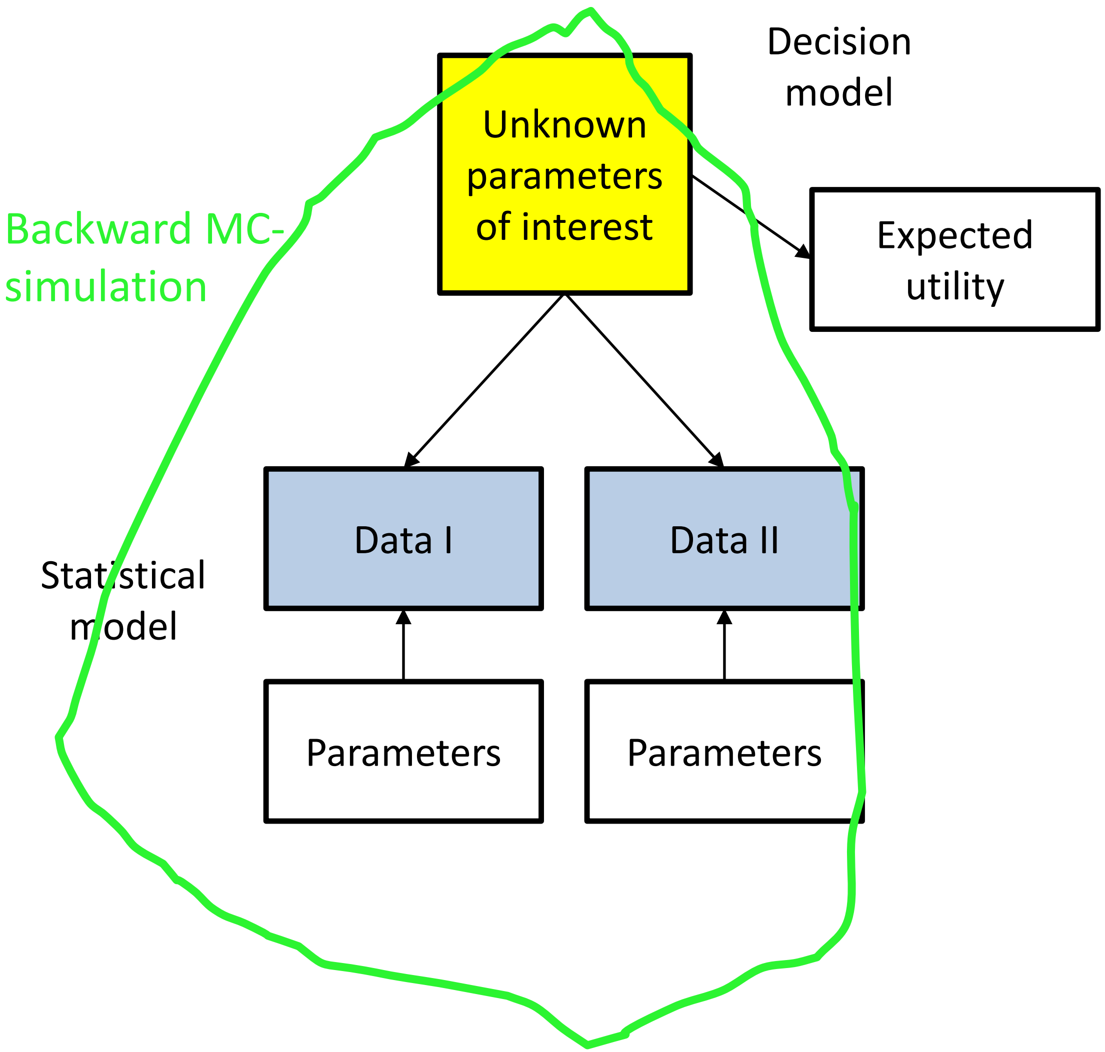
Ades et al. (2006). Bayesian methods for evidence synthesis in cost-effectiveness analysis. *Pharmacoeconomics*

Spiegelhalter and Best (2003). Bayesian approaches to multiple sources of evidence and uncertainty in complex cost-effectiveness modelling. *Stat Med*

Jackson et al. (2015). Calibration of complex models through Bayesian evidence synthesis: a demonstration and tutorial. Med Decis Making

Sahlin and Jiang (2015). Bayesian Evidence Synthesis and the quantification of uncertainty in a Monte Carlo simulation. *J of Risk and Reliability*

BES – integrated model



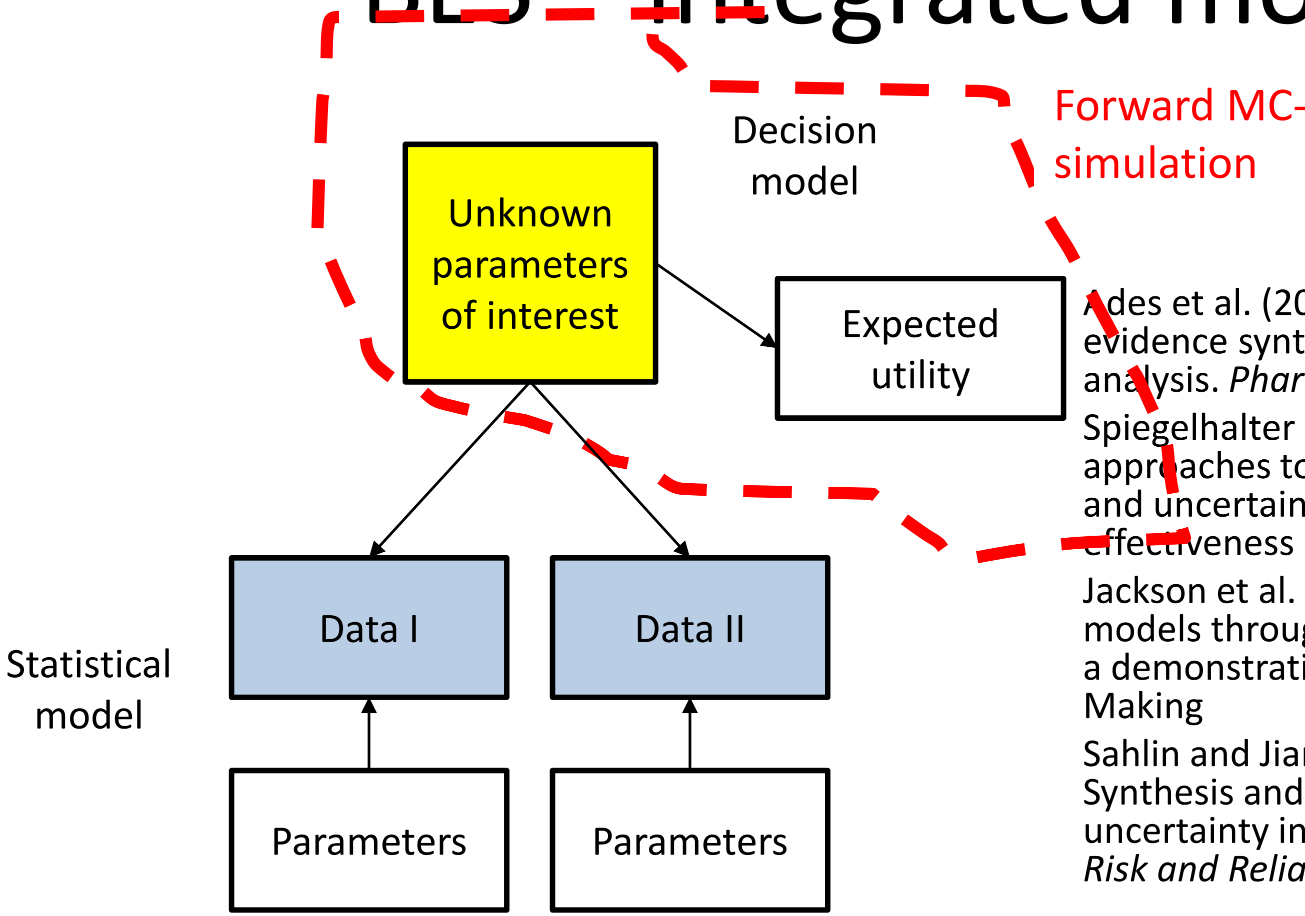
Ades et al. (2006). Bayesian methods for evidence synthesis in cost-effectiveness analysis. *Pharmacoeconomics*

Spiegelhalter and Best (2003). Bayesian approaches to multiple sources of evidence and uncertainty in complex cost-effectiveness modelling. *Stat Med*

Jackson et al. (2015). Calibration of complex models through Bayesian evidence synthesis: a demonstration and tutorial. Med Decis Making

Sahlin and Jiang (2015). Bayesian Evidence Synthesis and the quantification of uncertainty in a Monte Carlo simulation. *J of Risk and Reliability*

BES – integrated model



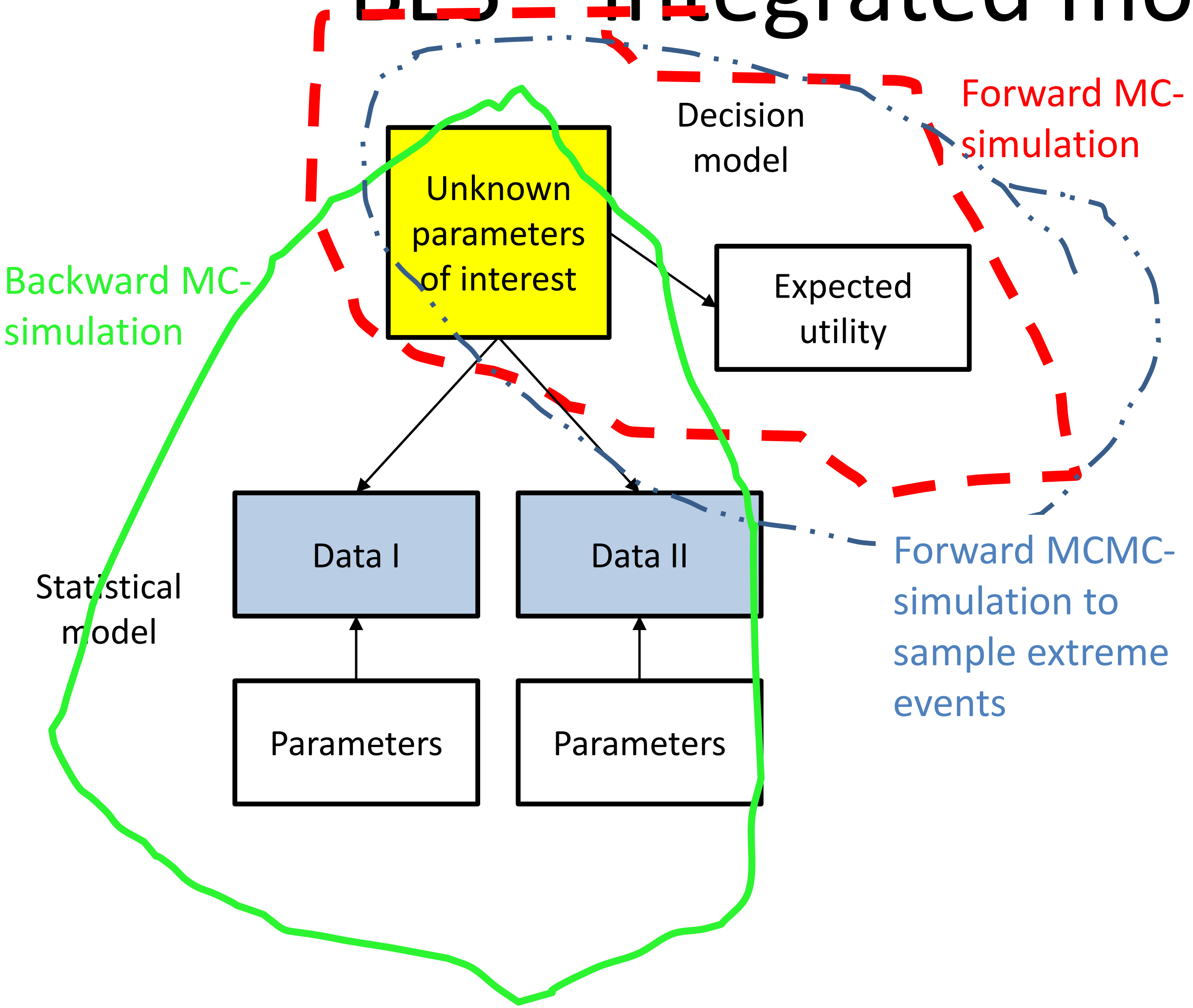
Ades et al. (2006). Bayesian methods for evidence synthesis in cost-effectiveness analysis. *Pharmacoeconomics*

Spiegelhalter and Best (2003). Bayesian approaches to multiple sources of evidence and uncertainty in complex cost-effectiveness modelling. *Stat Med*

Jackson et al. (2015). Calibration of complex models through Bayesian evidence synthesis: a demonstration and tutorial. Med Decis Making

Sahlin and Jiang (2015). Bayesian Evidence Synthesis and the quantification of uncertainty in a Monte Carlo simulation. *J of Risk and Reliability*

BES – integrated model



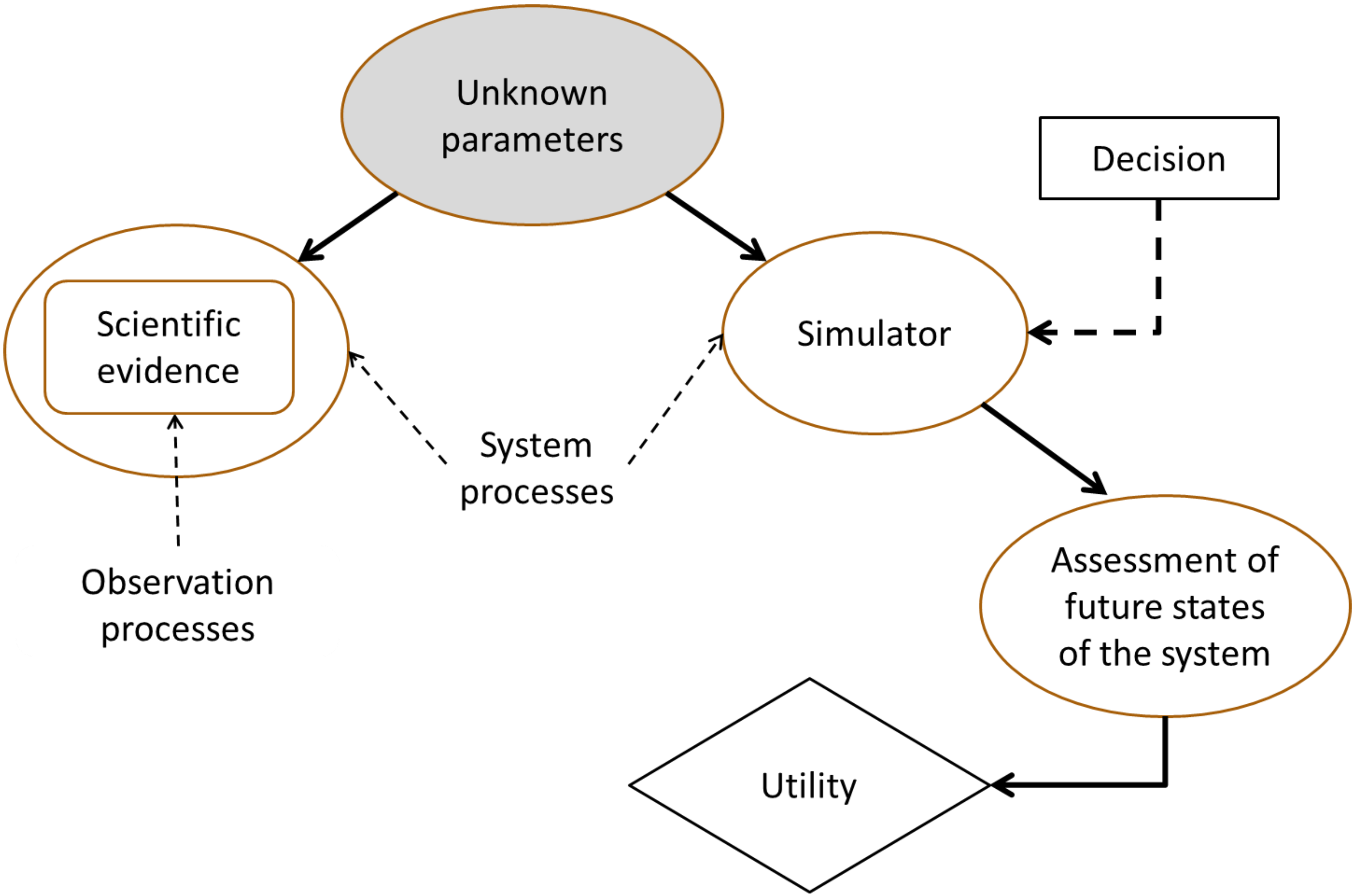
Bayesian Evidence Synthesis is a framework to calibrate complex models

Calibration of Complex Models through Bayesian Evidence Synthesis: A Demonstration and Tutorial

Christopher H. Jackson, PhD, Mark Jit, PhD, Linda D. Sharples, PhD, Daniela De Angelis, PhD

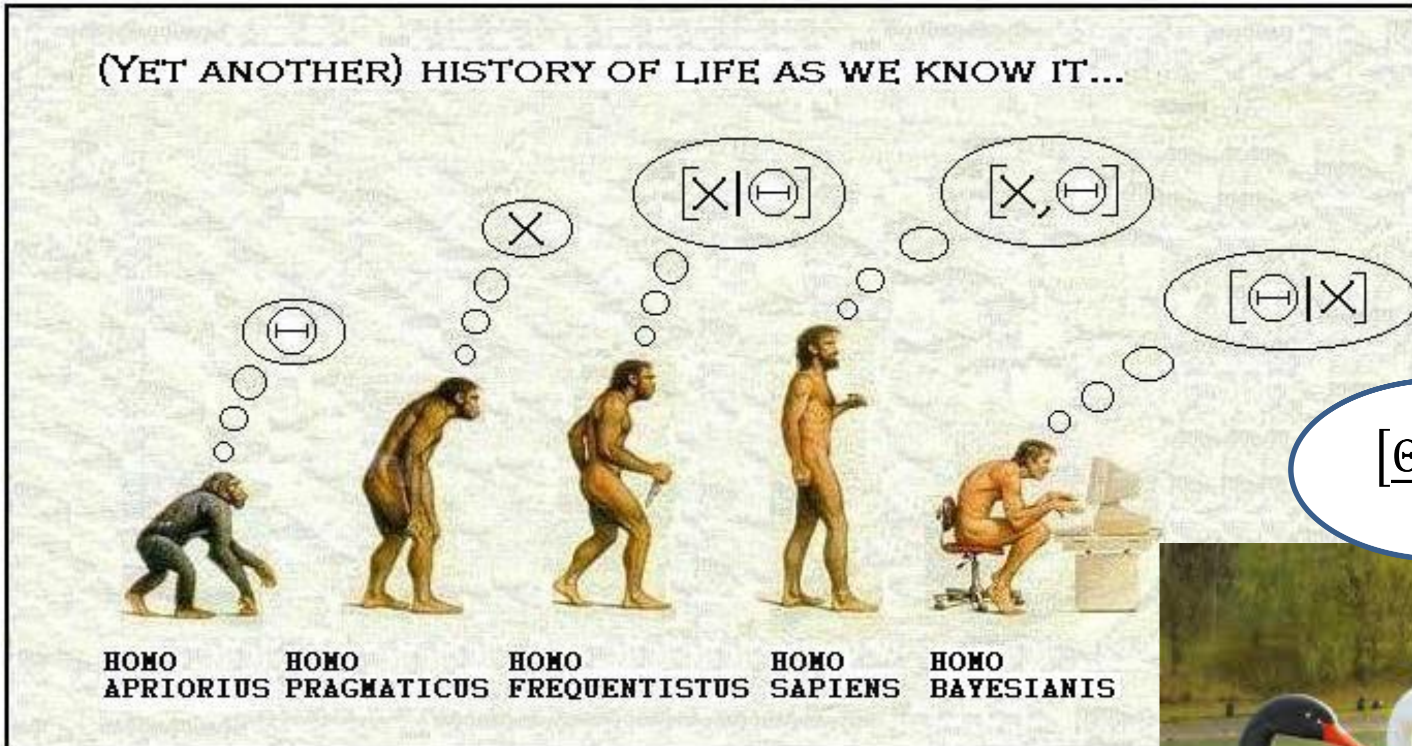
Decision-analytic models must often be informed using data that are only indirectly related to the main model parameters. The authors outline how to implement a Bayesian synthesis of diverse sources of evidence to calibrate the parameters of a complex model. A graphical model is built to represent how observed data are generated from statistical models with unknown parameters and how those parameters are related to quantities of interest for decision making. of human papillomavirus (HPV-16) infection was rebuilt in a Bayesian framework. Transition probabilities between states of disease severity are inferred indirectly from cross-sectional observations of prevalence of HPV-16 and HPV-16–related disease by age, cervical cancer incidence, and other published information. Previously, a discrete collection of plausible scenarios was identified but with no further indication of which of these are more plausible.

BES – another way to illustrate it



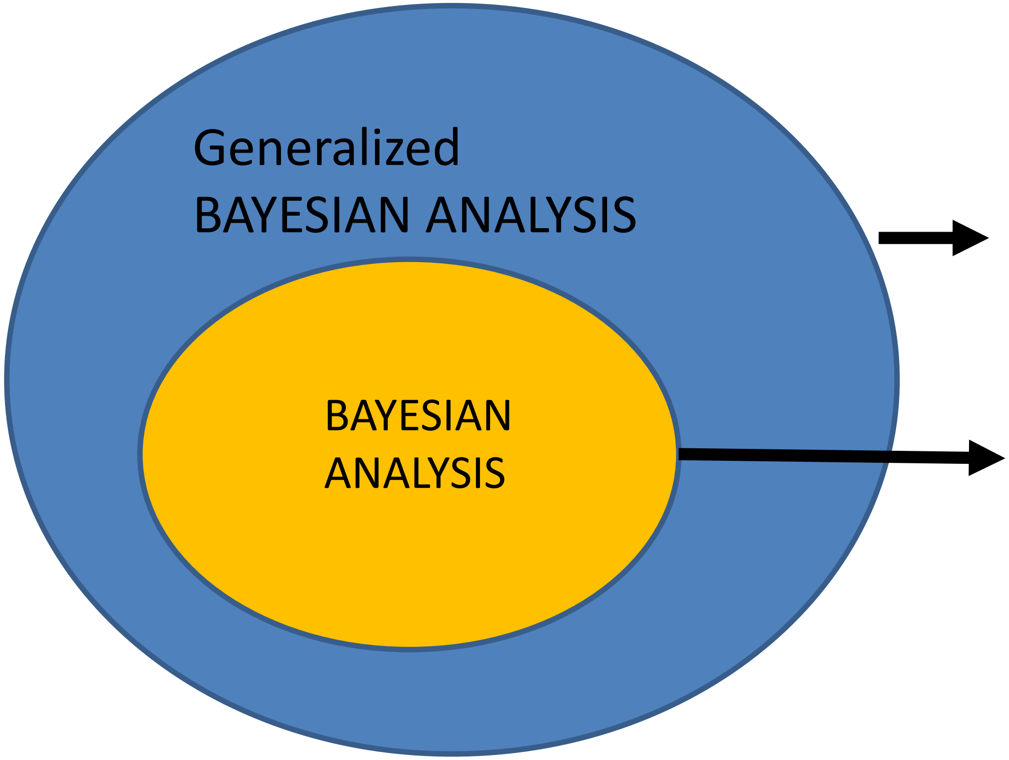
Robust

- Suggestions of the meaning of robust:
- A robust estimate/decsision is insensitive to outliers
- A robust e/d is insensitive to uncertainty
- Consequences of a robust decision remains in a acceptable range
- A robust decision strategy performs well (in a wider context [the meaning of well may include both the outcome and principles of cautiousness] under to widely varying conditions [in the system I pressume]
- A robust decision strategy applies cautionary principles and is sensitive to new knowledge (e.g. adapts to the state of a dynamical system or consider any reductions of uncertainty if that can improve overall performance)


(YET ANOTHER) HISTORY OF LIFE AS WE KNOW IT...
[Θ]
[X]
[X|Θ]
[X,Θ]
[Θ|X]
HOMO APRIORIUS
HOMO PRAGMATICUS
HOMO FREQUENTISTUS
HOMO SAPIENS
HOMO BAYESIANIS


$[\underline{\Theta}\overline{\Theta}|X]$
HOMO DOUBTUS

Robust analysis "=" bound by sensitivity analysis to choise of prior



$\underline{Pr}, \overline{Pr}$ -> $\underline{EU}, \overline{EU}$

Pr -> EU

Robust meta-analysis

Charnley

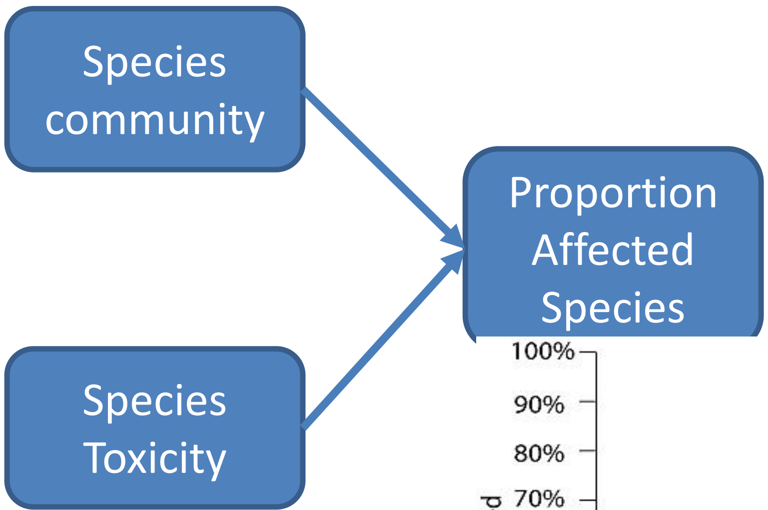
Stanmore

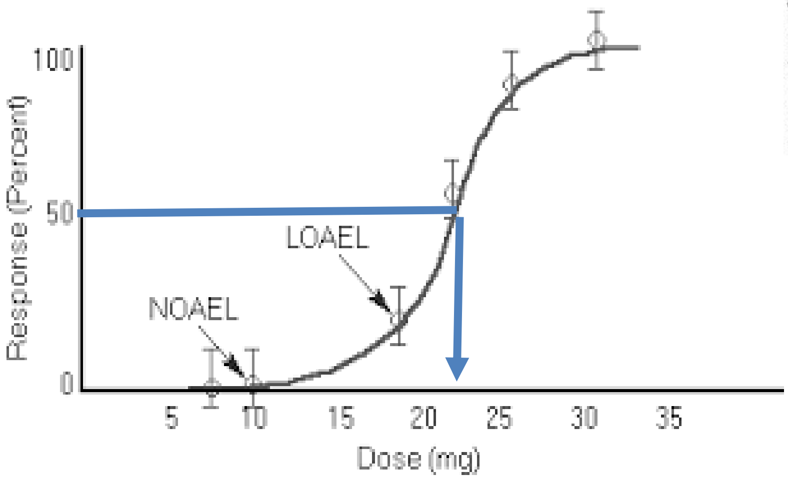
Table IV. Summary of evidence on revision hazards for Charnley and Stanmore prostheses: hazard ratios < 1 are in favour of Stanmore.

Source	Charnley: Number of patients	Charnley: Revision rate	Stanmore: Number of patients	Stanmore: Revision rate	Estimated hazard ratio (HR)	(95% int.)
					Fixed-effects model	
Registry	28 525	5.9%	865	3.2%	0.55	(0.37–0.77)
RCT	200	3.5%	213	4.0%	1.34	(0.45–3.46)
Case series	208	16.0%	982	7.0%	0.44	(0.28–0.66)
					Common-effect model	
					0.52	(0.39–0.67)
Quality weights [registry, RCT, case series]					*Random-effects model*	
				[1, 1, 1]	0.54	(0.37–0.78)
				[0.5, 1, 0.2]	0.61	(0.36–0.98)
				[0.1, 1, 0.05]	0.82	(0.36–1.67)

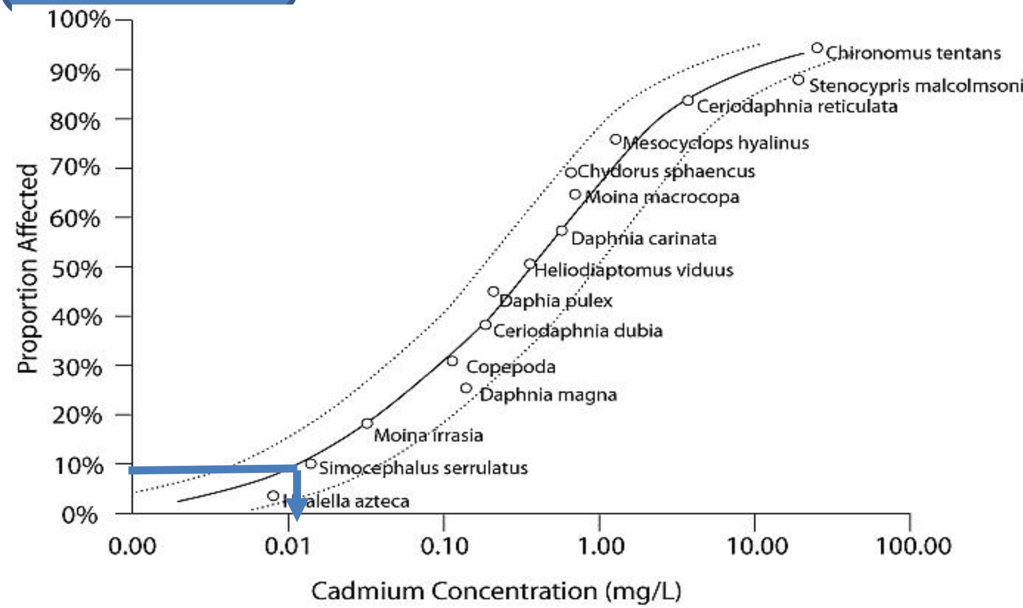
Spiegelhalter and Best (2003). Bayesian approaches to mulitple sources of evidence and uncertainty in complex cost-effectiveness modelling. *Statist. Med.*

Chemical hazard assessment





EC50



Hazardous concentration

A chemical hazard assessment as a Bayesian Evidence Synthesis

- Decision problem
- Utility function
- System model
- Data generating model
- Data
- Priors
- Quality parameter

A chemical hazard assessment as a Bayesian Evidence Synthesis

- Decision problem: Set a treshold - Find the largest acceptable concentration in the environment
- ~~Utility~~ Loss function - LINearEXponential
- System model – Species sensitivity to the substance follows a Normal distribution
- Data generating model – estimates are the result of different ecotoxicoloigcal studies. These are subject to variability which are more similar withing species than between species
- Data – K species, with repeated measurements for some of them
- Priors
- Quality parameter – weight on every toxicity data

LINEX loss function

In this section we discuss the concept of estimating an optimal decision for LHC_p from a completely different loss function. We first start by describing the (modified) LINEX loss function to be

$$L(LHC_p, L\hat{H}C_p) = \beta \left[\exp\left\{ \alpha \frac{\delta - \psi(\boldsymbol{\theta})}{\sqrt{\boldsymbol{\theta}^T \cdot \boldsymbol{I}}} \right\} - \alpha \left\{ \frac{\delta - \psi(\boldsymbol{\theta})}{\sqrt{\boldsymbol{\theta}^T \cdot \boldsymbol{I}}} \right\} - 1 \right] \quad (3)$$

where $\boldsymbol{\theta}^T \cdot \mathbf{I} = \sigma$ is used to scale the difference between the true LHC_p and the estimator $L\hat{H}C_p$ as done by Zieliński (2005) for reasons described later on; and β is a positive constant used to scale the loss function to the correct scale of loss measurement. The LINear-EXponential (LINEX) loss function was first proposed by Varian (1975) which conveyed loss as increasing linearly on one side and exponentially on the other side. That is, not only was

Hickey, G. L., Craig, P. S., & Hart, A. (2009). On the application of loss functions in determining assessment factors for ecological risk. Ecotoxicology and Environmental Safety, 72(2), 293-300.

Outline first exercise

- Study the code in stan_hazardassessment.R
- Draw the DAG of the model
- Generate artficial toxcity data
- Learn about the mean and standard deviation of the SSD by MCMC-sampling from the Bayesian model
- Find hazardous concentration which minimize expected loss
- Use code in the file: environmentalhazardassessment.R

- Use your own seed

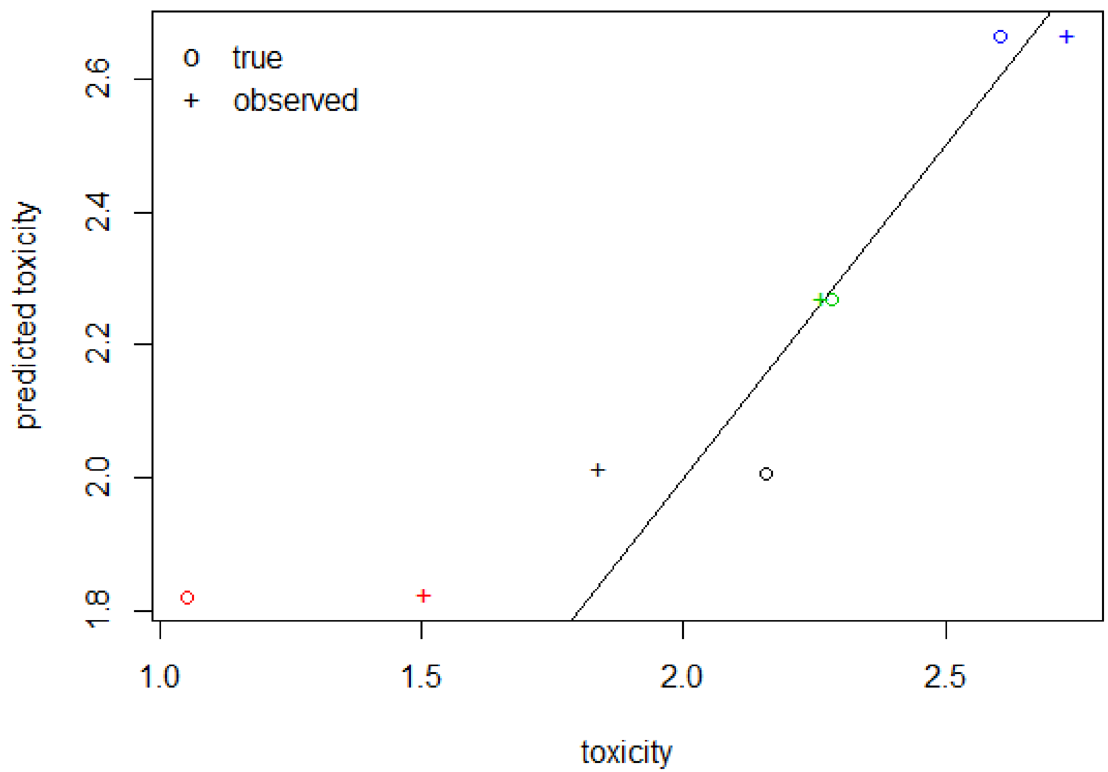
```
## Generate artificial toxicity data from a SSD with mu and sigma
ssd_data <- generate_data(mu = 2,sigma = 1,K = 4,s_sizes=1,seed = 1975)
```

- Run the mcmc sampling using

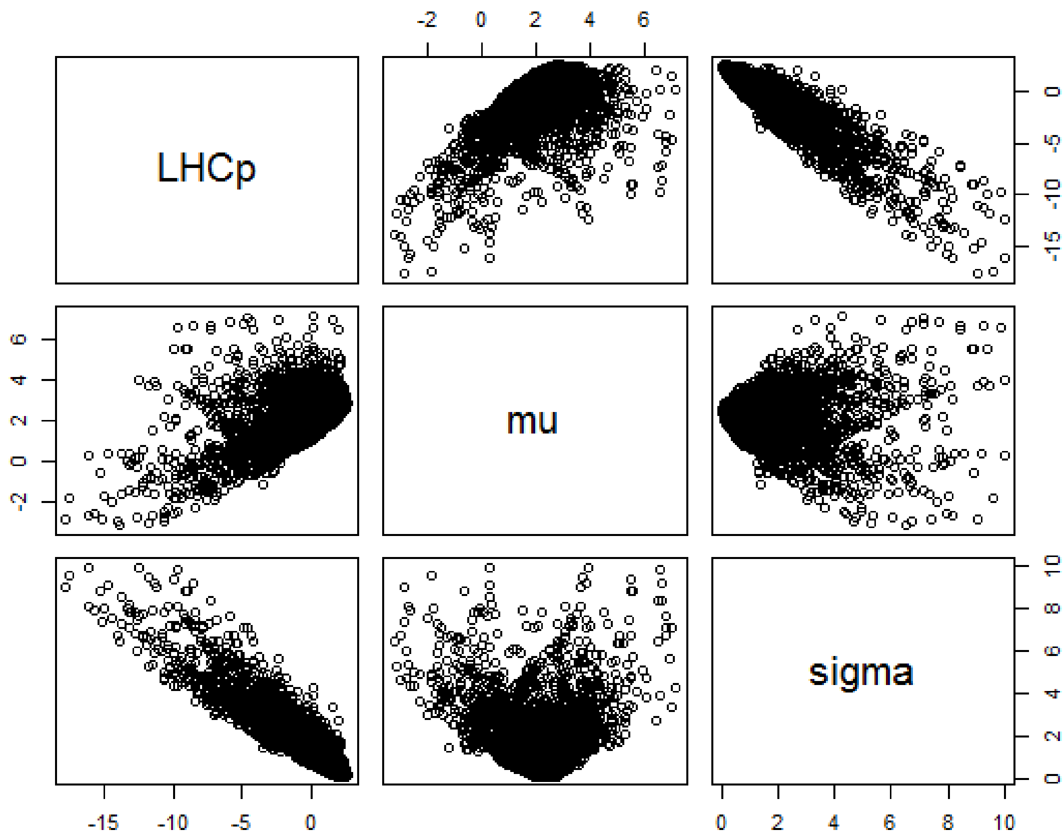
```
model = stan(model_name="model", model_code = code_ssd, data=dat,
             iter = 10000, chains = 4, verbose = FALSE)
```

Are we retrieveing the original parameters?

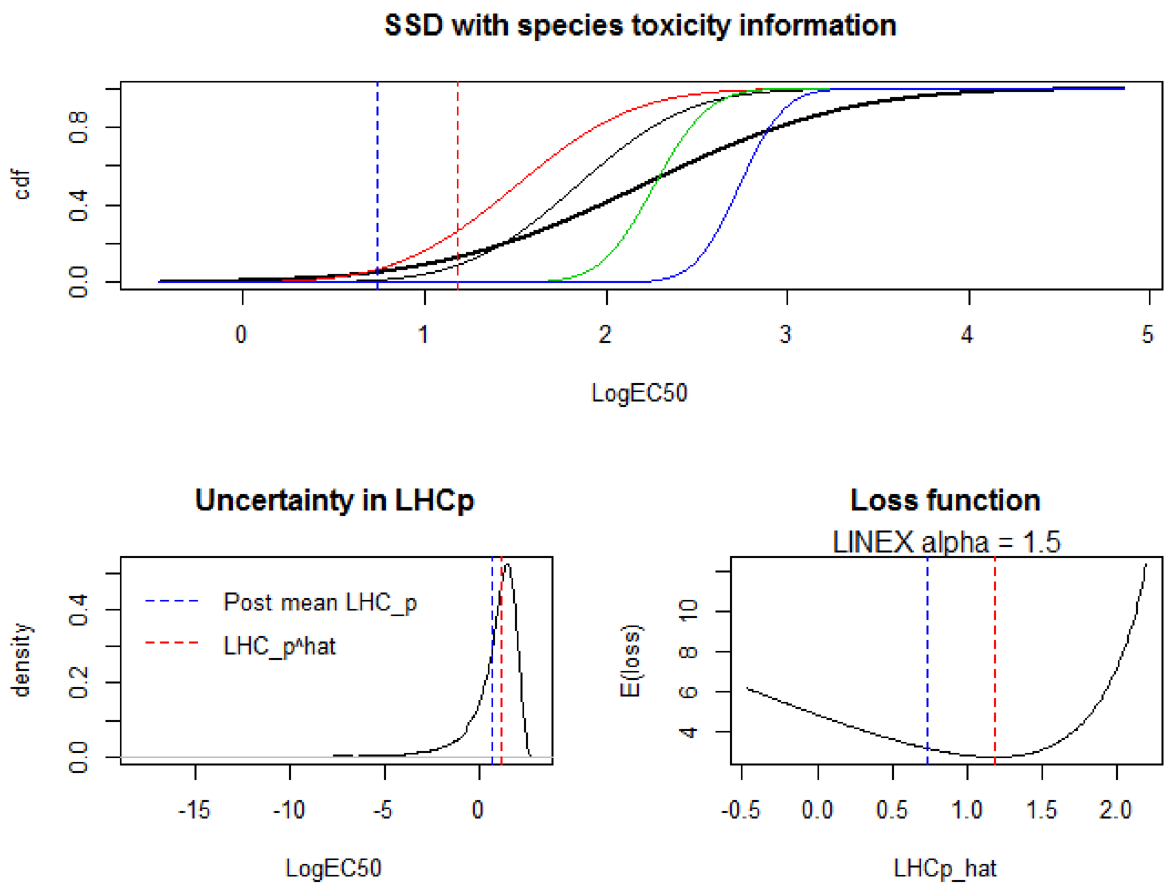
ppcheck_plot_toxicity(stanmodel=model,ssd_data)



The most important variables



System, Decision treshold and Loss



Outline second exercise

- Do a sensitivity analysis agains 1, 2 or 3!
- 1) Priors on mu and sigma (hyperparameters as well as distribution)
- 2) Quality weights on toxicity data (letting a w be close to zero means that it gives that data point very little influence in the model)
- 3) Choices of the alpha in the loss function
- Use e.g. the function robusthazardassessment which is in the R-file ssdcode.R
- How could one find a robust decision (i.e. treshold for the concentration allowed)?
- How would a code for preposterior analysis or prior predictive analysis to find suitable priors look like?

Outline third exercise

- Build your own (Robust) Bayesian Evidence Synthesis
- Simple system
- Use mulitple sources of data with different quality
- Include a decision analysis
- Solve the decision problem