

Is my model good enough? Deriving the expected value of model improvement via specifying model discrepancy

Mark Strong^{1,§}, Jeremy E. Oakley²

1. School of Health and Related Research (ScHARR), University of Sheffield, UK.
2. School of Mathematics and Statistics, University of Sheffield, UK.

§ Corresponding author m.strong@sheffield.ac.uk

September 2011

Abstract

Health economic models are representations of judgements about the functional relationships between the model's input parameters and the costs and health effects that the model is aiming to predict. We recognise that we can rarely define with certainty a 'true' model for a particular decision problem. Building an 'incorrect' model will result in an uncertain prediction error, which we denote 'structural uncertainty'. Structural uncertainty can be quantified within a Bayesian framework via the specification of a series of discrepancy terms, each representing at a sub function level within the model the difference between the sub function output and the true value of the intermediate parameter implied by the sub function. By using value of information analysis we can then determine the expected value of learning the discrepancy terms, which we interpret as the expected value of model improvement (EVMI). We illustrate the method using a Markov model case study.

KEYWORDS: Economic model uncertainty; Structural uncertainty; Bayesian decision theory; Expected value of perfect information; Gaussian process

1 Introduction

In the context of health economic evaluation, models that predict expected costs and health outcomes under competing decision options are typically ‘law-driven’ (based on our knowledge of the system) rather than ‘data-driven’ (fitted to data), following the distinction given in Saltelli et al. (2008). Indeed, we usually build such models because of a lack of data on long term costs and health outcomes. A law driven model can be thought of as a representation of judgements about the relationship between the model inputs and the model outputs. If we are uncertain what this ‘true’ structural relationship is, then even if we were to run the model at its true inputs, there would be an uncertain ‘structural error’ in the model prediction. We denote this uncertain error ‘structural uncertainty’. Note that we use the term ‘true’ value of the input to mean that which we would estimate in some perfect study with infinite sample size, and ‘true’ structural relationship to mean a (possibly non-unique) functional relationship that would result in the correct output given any set of ‘true’ values of the inputs.

Unless we are able to build a model that is true in the sense above we should expect an uncertain structural error in the model prediction. What are our judgements about this error? Given that we are likely to be uncertain about some or all of the model inputs is the uncertainty about the model structure important? Is the imperfect model good enough for the decision?

Some authors have considered the problem of uncertain health economic model structure from a statistical perspective and suggested model selection or averaging approaches based on the fit of the model to some set of observations (Jackson et al., 2009, 2010, 2011; Bojke et al., 2009). These methods clearly have a role in guiding structural choices for parts of the model where intermediate outputs can be fitted to data, but they cannot guide choices about the whole model structure since we do not observe future costs and health effects under each of our competing decision options. Our approach is therefore quite different. Rather than consider some measure of the ‘correctness’ of our model, we instead incorporate within the model our judgements about the structural error that arises from its imperfection.

If we denote the true, but unknown, target of our model prediction as Z (this could be for example the vector of net benefits of some set of decision options), and the output of our built model as Y then we write

$$Z = Y + \delta, \tag{1}$$

where the structural error is denoted by an uncertain ‘discrepancy’ term δ . We

define structural uncertainty as uncertainty about Z due to uncertainty about the structural error.

In a previous paper we illustrated, using a decision tree model case study, a method for making judgements about model structural error based on decomposing the model into a series of sub-functions that reveal a set of ‘intermediate’ parameters (Strong et al., 2011). For each sub function we asked the question ‘if we knew true values of the inputs to this sub function do we believe with certainty that the output of the sub function would equal the true value of the corresponding intermediate parameter?’. We introduced a discrepancy term δ_i , $i = 1, \dots, I$ at each sub function where we judged there to be potential structural error and expressed our beliefs about the size (and direction) of the error via the joint distribution, $p(\mathbf{X}, \delta_1, \dots, \delta_I)$, between inputs \mathbf{X} and discrepancies. Given $p(\mathbf{X}, \delta_1, \dots, \delta_I)$ we then used variance-based sensitivity analysis to determine the relative importance of the discrepancy terms in driving the uncertainty in the output (Saltelli et al., 2008). Alternatively, we can adopt a decision theoretic approach and compute the expected value of learning the true values of the discrepancies. If this is small compared with the expected value of learning the true values of inputs, then this offers reassurance that our current model is good enough for the decision at hand.

To illustrate, figure 1a shows a hypothetical model with ten inputs, $Y = f(X_1, \dots, X_{10})$, that aims to predict a quantity Z . The model has been decomposed into a series of sub functions, for example $Y_1 = f_1(X_1, X_2, X_3)$ and $Y_2 = f_2(X_4, X_5)$, revealing a set of six ‘intermediate’ parameters Y_1, \dots, Y_6 that have ‘true’ values Z_1, \dots, Z_6 . In three of the sub functions there is judged to be structural error. Figure 1b shows the incorporation of three uncertain discrepancy terms added to the model to correct the structural error.

Figure 1: (a) Hypothetical model with ten inputs and one output, decomposed to reveal six intermediate parameters. (b) We suppose that there is structural error in the sub functions that result in Y_1 , Y_5 and Y_6 . Three discrepancy terms are added to correct the error, i.e. $Z_1 = Y_1 + \delta_1$, $Z_5 = Y_5 + \delta_2$ and $Z_6 = Y_6 + \delta_3$.

In this paper we illustrate the application of the discrepancy method for managing structural uncertainty in another common type of health economic decision model, the Markov model. We imagine a scenario where we have built a relatively simple Markov model, but recognise that reality is more complex. We do not believe that even if we were to learn the ‘true’ values of the Markov transition probabilities and all other uncertain inputs in the model, that the predicted costs and health outcomes would equal their true values. We know that the model is a simplification, and we seek to answer the question ‘is it good enough?’.

The paper is organised as follows. In section 2 we introduce a case study based on a simple Markov model designed to predict the costs and health effects of two competing treatment options for HIV/AIDS. In section 3 we apply the discrepancy analysis method to the case study model in three scenarios that represent plausible

sets of assumptions regarding the structural error. In section 4 we present results including the ‘expected value of model improvement’ (EVMI) in each scenario. In a final section we discuss limitations and draw conclusions.

2 Case study model

In order to illustrate the method we introduce a case study that is based on a four state Markov model first described in Chancellor et al. (1997) and subsequently used for illustrative purposes in Drummond et al. (2005) and Briggs et al. (2006). The purpose of the model is to predict costs and health outcomes (life years) under two decision options, zidovudine monotherapy versus zidovudine plus lamivudine combination therapy, in people with HIV. Allowable transitions between the four health states are shown in figure 2.

Figure 2: Structure of the case study Markov model

The authors of the original paper chose time steps of 1 year and ran the model to a time horizon of 20 years.

2.1 Notation

We index the monotherapy and combination therapy decision options $d = 1, 2$ respectively, the four mutually exclusive health states as $n = 1, \dots, 4$, and the time steps in years as $t = 0, \dots, 20$. If we imagine a cohort of people exposed to decision option d , we denote h_{dnt} as the proportion of the cohort who are in health state n during time step t (alternatively, h_{dnt} represents, under decision option d at time step t , the probability that a single individual exists in health

state n versus the other states). We call $\mathbf{h}_{dt} = (h_{d1t}, \dots, h_{d4t})'$ the *state vector* for decision option d at time step t , and note the constraint $\sum_{n=1}^4 h_{dnt} = 1 \forall d, t$.

We denote the costs and health effects accrued for health state n during time step t under decision d as c_{dnt} and e_{dnt} respectively. Costs and outcomes are time dependent to allow the discounting of costs and effects accrued in the future (Krahn and Gafni, 1993). We can therefore write costs and effects at time step t in terms of costs and effects at time zero via $c_{dnt} = c_{dn0}(1 + r_c)^{-t}$ and $e_{dnt} = e_{dn0}(1 + r_e)^{-t}$, where r_c and r_e are the per-year discount rate for costs and health effects. The health effect of interest for this decision problem is life years, so $e_{dn0} = 1$ for health states $n = 1, 2, 3$, and zero for the death state $n = 4$. We denote the vector of costs for all health states at time step t under decision d as $\mathbf{c}_{dt} = (c_{d1t}, \dots, c_{d4t})'$, and the vector of health effects as $\mathbf{e}_{dt} = (e_{d1t}, \dots, e_{d4t})'$.

2.2 The Markov model

The authors assumed a simple time-homogeneous Markov process (i.e. transition probabilities remain fixed for all time steps). Under this assumption the probability that an individual will move from health state x to health state y under decision d is given by p_{dxy} , and we note the constraints that $p_{dxy} \geq 0 \forall d, x, y$ and $\sum_{y=1}^4 p_{dxy} = 1 \forall d, x$.

Transition from a worse health state to a better health state is considered impossible in this decision scenario. The transition matrix for the monotherapy ($d = 1$) option is therefore of the form,

$$\mathbf{M}_1 = \begin{pmatrix} p_{111} & p_{112} & p_{113} & p_{114} \\ 0 & p_{122} & p_{123} & p_{124} \\ 0 & 0 & p_{133} & p_{134} \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad (2)$$

where the lower diagonal elements are zero. Death is an absorbing state.

The matrix \mathbf{M}_1 is modified by the incorporation of a combination therapy treatment effect parameter, RR , to give the transition matrix for the combination therapy ($d = 2$) option,

$$\mathbf{M}_2 = \begin{pmatrix} 1 - RR(p_{112} + p_{113} + p_{114}) & RR \cdot p_{112} & RR \cdot p_{113} & RR \cdot p_{114} \\ 0 & 1 - RR(p_{123} + p_{124}) & RR \cdot p_{123} & RR \cdot p_{124} \\ 0 & 0 & 1 - RR \cdot p_{133} & RR \cdot p_{134} \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (3)$$

Given the transition matrix \mathbf{M}_d and state vector \mathbf{h}_{dt} , we can generate $\mathbf{h}_{d,t+1}$ via the evolution equation

$$\mathbf{h}'_{d,t+1} = \mathbf{h}'_{dt} \mathbf{M}_d, \quad (4)$$

and we can therefore express \mathbf{h}_{dt} in terms of the state vector at time step 0, i.e. $\mathbf{h}'_{dt} = \mathbf{h}'_{d0} \mathbf{M}_d^t$, where $\mathbf{M}_d^t = \prod_{l=1}^t \mathbf{M}_d$.

If we value (in cost units) one unit of health outcome at λ , our final model for the net monetary benefit associated with decision option d is

$$NB_d = \lambda e_d^{tot} - c_d^{tot} = \lambda \sum_{t=0}^{20} \mathbf{h}'_{d0} \mathbf{M}_d^t \mathbf{e}_{dt} - \sum_{t=0}^{20} \mathbf{h}'_{d0} \mathbf{M}_d^t \mathbf{c}_{dt}. \quad (5)$$

Assuming that we are uncertain about some or all the inputs into the model, our optimum decision is that which maximises the expected net benefit.

2.3 Base case input parameter values

Transition probabilities, costs and the treatment effect parameter are all considered uncertain in the base case model, with distributions shown in Tables 1 and 2.

Table 1: Transition probability distributions for $d = 1$

$$\begin{aligned} (p_{111}, p_{112}, p_{113}, p_{114}) &\sim \text{Dirichlet} (1251, 350, 115, 14) \\ (p_{121}, p_{122}, p_{123}, p_{124}) &\sim \text{Dirichlet} (0, 731, 512, 15) \\ (p_{131}, p_{132}, p_{133}, p_{134}) &\sim \text{Dirichlet} (0, 0, 1312, 437) \\ (p_{141}, p_{142}, p_{143}, p_{144}) &= (0, 0, 0, 1) \end{aligned}$$

Table 2: Cost and relative risk distributions

Label	Description	Distribution	Mean	SD
cc_1	Undiscounted care costs of 1 time step in state 1 (£)	normal	2756	400
cc_2	Undiscounted care costs of 1 time step in state 2 (£)	normal	3052	437
cc_3	Undiscounted care costs of 1 time step in state 3 (£)	normal	9007	1449
RR	Treatment effect (combi vs monotherapy)	lognormal	$\log(0.509)$	0.05

Drug treatment costs are considered fixed and known, as are discount rates (Table 3). We assume that the combination therapy is effective throughout the whole of the modelled 20 year period, rather than just for the first year (this is presented as an alternative scenario rather than the base case in Chancellor et al., 1997).

Table 3: Fixed inputs

Label	Description	Value
c_Z	Zidovudine cost (£)	2278
c_L	Lamivudine cost (£)	2087
r_c	Discount rate for costs	3.5% per year
r_e	Discount rate for outcomes	3.5% per year

3 Discrepancy analysis

3.1 Incorporating judgements about model structural error into the Markov model

We believe that the transition of individuals through health states is not adequately described by a simple time-homogeneous Markov model and therefore expect there to be error in the prediction of our base case model. We wish to quantify this structural error to determine whether we need to build a more complex model. In particular we wish to determine the expected value of improving the model. We restrict ourselves in this paper to considering only structural error that relates to the Markov model itself. In many applications a Markov model is part of a larger model that may also include, for example, a decision tree element where we may also judge there to be structural error.

We introduce a series of discrepancy terms, each of which represents the difference between the output of a sub function in the built model and the true value of that output quantity. Discrepancy terms are incorporated in the model at the level of the evolution of the health state vector, replacing Eq. (4) with

$$\mathbf{h}'_{dt} = \mathbf{h}'_{d,t-1} \mathbf{M}_d + \boldsymbol{\delta}_{dt}, \quad (6)$$

where $\boldsymbol{\delta}_{dt}$ is a vector of discrepancy terms that quantifies the error in the state vector at time t for decision option d .

In the analysis for our case study we have found it more intuitive to think about discrepancies as applying to the transition matrix rather than to the state vector, writing $\boldsymbol{\delta}_{dt} = \mathbf{h}'_{d,t-1} \boldsymbol{\Delta}_{dt}$ and expressing judgements about the model error via $\boldsymbol{\Delta}_{dt}$, a matrix of discrepancy terms of the same dimensionality as \mathbf{M}_d . We

re-express Eq. (6) as

$$\begin{aligned}
\mathbf{h}'_{dt} &= \mathbf{h}'_{d,t-1} \mathbf{M}_d + \delta_{dt}, \\
&= \mathbf{h}'_{d,t-1} \mathbf{M}_{dt} + \mathbf{h}'_{d,t-1} \Delta_{dt}, \\
&= \mathbf{h}'_{d,t-1} (\mathbf{M}_d + \Delta_{dt}).
\end{aligned} \tag{7}$$

The matrix $(\mathbf{M}_d + \Delta_{dt})$ must obey the same constraints as \mathbf{M}_d , i.e. all elements must lie within the interval $[0, 1]$ and each row must sum to one. We can ensure this if each element of Δ_{dt} , δ_{dtxy} , is constrained to lie in the interval $[-p_{dxy}, 1 - p_{dxy}]$, and if each row of Δ_{dt} sums to zero.

Given the transition probability matrices (equations 2 and 3), there are potentially six such unconstrained discrepancy terms per decision option per time step, and we denote these $\delta_{d1t}, \dots, \delta_{d6t}$. The discrepancy matrix Δ_{dt} is therefore

$$\Delta_{dt} = \begin{pmatrix} -(\delta_{d1t} + \delta_{d2t} + \delta_{d3t}) & \delta_{d1t} & \delta_{d2t} & \delta_{d3t} \\ 0 & -(\delta_{d4t} + \delta_{d5t}) & \delta_{d4t} & \delta_{d5t} \\ 0 & 0 & -\delta_{d6t} & \delta_{d6t} \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{8}$$

We may judge that structural error relates only to a subset of the transitions in the model. Where we judge there to be no structural error the corresponding discrepancy term will be zero.

3.2 Case study scenario 1 - time dependent transition probabilities

In the first scenario of our case study we judge that there is an important time dependent relationship between age and the probability of death that is not captured in the simple time homogeneous model. We therefore introduce three discrepancy terms (per time step per decision), one for each transition from an alive state to the death state. Given the general expression for the discrepancy matrix in Eq. (8) we expect that δ_{dit} is non-zero for $i = 3, 5, 6$, and zero for $i = 1, 2, 4$. Given three discrepancy terms per decision option per time step there are $3 \times 2 \times 21 = 126$ discrepancy terms in total. Specifying judgements about the model discrepancy via the joint distribution of such a large number of terms clearly requires a parsimonious parametrisation that reflects the dependencies between discrepancy terms.

To illustrate our approach to this specification problem we consider the discrepancy term, δ_{d6t} , that describes the structural error in the built model with

respect to the probability of transition from AIDS to death. We judge that the probability of this transition increases monotonically over time rather than being constant in the base case model, but we are unsure as to the exact nature of the relationship between the probability of death and time. This belief implies that the uncertain discrepancy term δ_{d6t} must also increase monotonically with respect to time. We judge that at $t = 0$ the probability of death may be approximately 20% lower than the constant value (0.25) in the built model, and at $t = 20$ may be approximately 20% higher, but we have considerable uncertainty. Figure 3 represents some plausible realisations of the discrepancy δ_{d6t} as a function of time for $d = 1$.

Figure 3: Four plausible realisations of the discrepancy term δ_{d6t} for $d = 1$ in scenario 1

3.3 Parametrising the discrepancy using a Gaussian process

We wish to find a convenient and parsimonious parametrisation for the joint distribution of the 126 discrepancy terms $\delta_{d,i,t}$, $d = 1, 2$, $i = 3, 5, 6$, and $t = 0, \dots, 20$. We begin by noting that the reason for choosing a Markov model structure for our built model was to reflect a dynamic time dependent process, so it seems reasonable to consider discrepancy as a function of time step, i.e.

$\delta_{dit} = f_{di}(t)$. We then assume that the functions $f_{di}(t)$ follow a Gaussian process, i.e. that $\{f_{1,1}(0), \dots, f_{2,6}(20)\}$ has a multivariate normal distribution with mean function, $E\{f_{di}(t)\} = m(d, i, t)$ and covariance function $\text{Cov}\{f_{di}(t), f_{d^*i^*}(t^*)\} = c(d, i, t, d^*, i^*, t^*)$.

This highly flexible and parsimonious parametrisation of set of unknown functions allows us to specify not only our uncertainty about each δ_{dit} , but also the correlation structure of discrepancies through time, the correlations between the three non-zero discrepancy terms per decision, and the correlation between the discrepancy terms for the $d = 1, 2$ decisions for each transition i .

3.3.1 Specifying the mean function

We specify the mean for each discrepancy, $E\{f_{di}(t)\}$, as a function $m(d, i, t)$. For scenario 1 a linear form, $E\{f_{di}(t)\} = m(d, i, t) = \beta_{0,di} + \beta_{1,di}t$, adequately reflects our judgements, but depending on the decision problem alternative choices might be higher order polynomial, $E\{f_{di}(t)\} = m(d, i, t) = \beta_{0,di} + \dots + \beta_{n,di}t^n$, exponential, $E\{f_{di}(t)\} = m(d, i, t) = \beta_{0,di} + \beta_{1,di} \exp(\beta_{2,di}t)$, or stepped, $E\{f_{di}(t)\} = m(d, i, t) = \beta_{0,di} + \beta_{1,di}I(t > \beta_{2,di})$. We placed normal distributions on the linear mean function parameters $\beta_{0,di}$ and $\beta_{1,di}$ with hyperparameters shown in Table 4.

Table 4: Hyperparameters to specify GP mean function

Hyperparameter	Scenario 1	Scenario 2	Scenario 3	Transition
Intercept ($\beta_{0,di}$)	Mean (sd) $\times 10^{-3}$	Mean (sd) $\times 10^{-3}$	Mean (sd) $\times 10^{-3}$	state x to y ; decision
$\beta_{0,11}$	0 (0)	0 (0)	0 (0)	1 to 2; monotherapy
$\beta_{0,12}$	0 (0)	0 (0)	0 (0)	1 to 3; monotherapy
$\beta_{0,13}$	-1.0 (0)	0 (0)	0 (0)	1 to 4; monotherapy
$\beta_{0,14}$	0 (0)	0 (0)	0 (0)	2 to 3; monotherapy
$\beta_{0,15}$	-1.2 (0)	0 (0)	0 (0)	2 to 4; monotherapy
$\beta_{0,16}$	-25.0 (0)	0 (0)	0 (0)	3 to 4; monotherapy
$\beta_{0,21}$	0 (0)	0 (0)	0 (0)	1 to 2; combi therapy
$\beta_{0,22}$	0 (0)	0 (0)	0 (0)	1 to 3; combi therapy
$\beta_{0,23}$	-0.5 (0)	0 (0)	0 (0)	1 to 4; combi therapy
$\beta_{0,24}$	0 (0)	0 (0)	0 (0)	2 to 3; combi therapy
$\beta_{0,25}$	-0.61 (0)	0 (0)	0 (0)	2 to 4; combi therapy
$\beta_{0,26}$	-12.7 (0)	0 (0)	0 (0)	3 to 4; combi therapy
Slope ($\beta_{1,di}$)	Mean (sd) $\times 10^{-4}$	Mean (sd) $\times 10^{-4}$	Mean (sd) $\times 10^{-4}$	state x to y ; therapy
$\beta_{1,11}$	0 (0)	0 (0)	0 (0)	1 to 2; monotherapy
$\beta_{1,12}$	0 (0)	0 (0)	0 (0)	1 to 3; monotherapy
$\beta_{1,13}$	1.0 (0)	0 (0)	0 (0)	1 to 4; monotherapy
$\beta_{1,14}$	0 (0)	0 (0)	0 (0)	2 to 3; monotherapy
$\beta_{1,15}$	1.2 (0)	0 (0)	0 (0)	2 to 4; monotherapy
$\beta_{1,16}$	25.0 (0)	0 (0)	0 (0)	3 to 4; monotherapy
$\beta_{1,21}$	0 (0)	24.8 (13.78)	0 (0)	1 to 2; combi therapy
$\beta_{1,22}$	0 (0)	8.2 (4.57)	0 (0)	1 to 3; combi therapy
$\beta_{1,23}$	0.51 (0)	1.2 (0.68)	0 (0)	1 to 4; combi therapy
$\beta_{1,24}$	0 (0)	50.0 (27.8)	0 (0)	2 to 3; combi therapy
$\beta_{1,25}$	0.61 (0)	1.5 (0.82)	0 (0)	2 to 4; combi therapy
$\beta_{1,26}$	12.7 (0)	30.7 (17.0)	0 (0)	3 to 4; combi therapy

3.3.2 Specifying the covariance function

We make a number of simplifying assumptions when specifying the covariance function, but note that all of these assumptions may be relaxed at the cost of specifying a greater number of hyperparameters. We assume in scenario 1 that the variance of each discrepancy δ_{dit} remains constant for all t , requiring the specification of $2 \times 3 = 6$ variances, which we denote σ_{di}^2 . We state beliefs about the within-decision, between-transition term correlation through a parameter $\phi_{i,i^*} = \text{cor}(\delta_{dit}, \delta_{di^*t})$, assuming that this is constant over time t and across decisions. We state beliefs about the between-decision correlation through a parameter $\psi_{d,d^*} = \text{cor}(\delta_{dit}, \delta_{d^*it})$, assuming that this is constant over time t and across transitions i .

Finally we state beliefs about the correlation of the discrepancies through time by defining a correlation function $\rho(\cdot, \cdot)$ that depends on the distance between time steps, assuming this holds for all d and i . For the purposes of scenario 1 we use

the ‘Gaussian form’

$$\rho(t, t^*) = \exp \left\{ - \left(\frac{t - t^*}{\omega} \right)^2 \right\}, \quad (9)$$

where ω is the correlation length. The correlation length determines the degree of correlation between discrepancy terms at any particular ‘distance’, where distance is the number of Markov time steps between the terms. See Neal (1999) for a discussion of alternatives to this simple Gaussian form of correlation function.

The overall covariance function is therefore

$$\begin{aligned} \text{Cov}\{f_{di}(t), f_{d^*i^*}(t^*)\} &= c(d, i, t, d^*, i^*, t^*), \\ &= \sigma_{di}\sigma_{d^*i^*}\psi_{d,d^*}\phi_{i,i^*}\rho(t, t^*), \\ &= \sigma_{di}\sigma_{d^*i^*}\psi_{d,d^*}\phi_{i,i^*}\exp \left\{ - \left(\frac{t - t^*}{\omega} \right)^2 \right\}. \end{aligned} \quad (10)$$

Finally we specify a correlation structure for the discrepancies as they evolve through time via the correlation function with parameter ω (Eq. 9). Values chosen are shown in Table 5.

Table 5: Hyperparameters to specify GP covariance function

Variance hyperparameters (σ_{id})	Scenario 1 ($\times 10^{-3}$)	Scenario 2 ($\times 10^{-3}$)	Scenario 3 ($\times 10^{-3}$)	Transition
σ_{11}	0	0	28.9	A to B monotherapy
σ_{12}	0	0	9.6	A to C monotherapy
σ_{13}	1.0	0	1.4	A to D monotherapy
σ_{14}	0	0	58.1	B to C monotherapy
σ_{15}	1.2	0	1.7	B to D monotherapy
σ_{16}	25.0	0	35.7	C to D monotherapy
σ_{21}	0	5.1	14.7	A to B combi therapy
σ_{22}	0	1.7	4.9	A to C combi therapy
σ_{23}	0.51	0.25	0.7	A to D combi therapy
σ_{24}	0	10.4	29.6	B to C combi therapy
σ_{25}	0.61	0.31	0.9	B to D combi therapy
σ_{26}	12.7	6.4	18.1	C to D combi therapy
Correlation hyperparameters	Scenario 1	Scenario 2	Scenario 3	Description
ϕ	0.8	0.9	0	Between discrepancy term correlation
ψ	0.9	0	0	Between decision correlation
ω	32	7	7	Correlation length parameter

3.3.3 Monotonicity constraint for $f_{di}(t)$

We wish to ensure that $f_{di}(t)$ is monotone with respect to t to reflect our belief that the probability of death increases with time. However, realisations of a Gaussian

process tend to be ‘wiggly’ non-monotone functions, with the degree of ‘wiggleness’ controlled by the ω parameter. Increasing values of ω will result in an increasingly smooth functions, so by carefully choosing ω we can ensure that the realisations of the Gaussian process reflect a plausible relationship between the discrepancy and time.

Monotonicity with respect to t implies that, for a once differentiable function $f_{di}(t)$ that $\partial f_{di}(t)/\partial t > 0 \forall t$, or $\partial f_{di}(t)/\partial t < 0 \forall t$. Informally then, we can ensure monotonicity by choosing hyperparameters for the mean and covariance functions such that this holds with some probability α .

It is a property of an n times differentiable Gaussian process $f(x) \sim GP\{m(x), c(x, x^*)\}$ with n times differentiable mean and covariance functions, that $\partial^n f(x)/\partial x^n$ is also a Gaussian process with mean function

$$E \left\{ \frac{\partial^n}{\partial x^n} f(x) \right\} = \frac{\partial^n}{\partial x^n} m(x), \quad (11)$$

and covariance function

$$\text{cov} \left\{ \frac{\partial^n}{\partial x^n} f(x) \Big|_{x=x}, \frac{\partial^n}{\partial x^n} f(x) \Big|_{x=x^*} \right\} = \frac{\partial^{2n}}{\partial x^n \partial x^{*n}} c(x, x^*). \quad (12)$$

See O’Hagan (1992) for further details.

This implies that $\partial f_{di}(t)/\partial t$ is the Gaussian process,

$$\frac{\partial}{\partial t} f_{di}(t) \sim GP \left\{ \frac{\partial}{\partial t} m(d, i, t), \frac{\partial^2}{\partial t \partial t^*} c(d, i, t, d, i, t^*) \right\}, \quad (13)$$

and we can ensure monotonicity of $f_{di}(t)$ with some pre-specified probability α by choosing parameters of $m(\cdot)$ and $c(\cdot, \cdot)$ such that

$$\left| \frac{\partial}{\partial t} m(d, i, t) \right| - \Phi^{-1}(\alpha) \sqrt{\frac{\partial^2}{\partial t \partial t^*} c(d, i, t, d, i, t^*)} > 0, \quad (14)$$

where $\Phi^{-1}(\alpha)$ is the inverse normal cumulative distribution function.

Given a linear mean function, $m(d, i, t) = \beta_{0,di} + \beta_{1,di}t$, and a Gaussian form for the correlation function with respect to time, $c(d, i, t, d, i, t^*) = \sigma_{di}^2 \exp \left\{ - \left(\frac{t-t^*}{\omega} \right)^2 \right\}$, Eq. (14) becomes

$$\left| \beta_{1,di} \right| - \Phi^{-1}(\alpha) \sqrt{\frac{2\sigma_{di}^2}{\omega^2}} > 0, \quad (15)$$

which by solving for ω gives

$$\omega > \frac{\Phi^{-1}(\alpha) \sqrt{2} \sigma_{di}}{\beta_{1,di}}. \quad (16)$$

We can therefore, given $\beta_{1,di}$ and σ_{di}^2 , ensure with some probability α that $f_{di}(t)$ is monotone through a choice of correlation length parameter ω that obeys (16). For scenario 1 we set $\alpha = 0.95$ and chose ω accordingly.

Ten samples from the Gaussian process for discrepancy term $\delta_{1,6,t}$ are shown in figure 4. Note the variation in functional form generated by the Gaussian process, reflecting our uncertainty about the relationship between probability of death and time, but with the constraint that the relationship between discrepancy and time should be monotone.

Figure 4: Ten samples from the distribution on discrepancy term $\delta_{1,6,t}$ in scenario 1

3.4 Sensitivity analysis to determine whether the discrepancies make any difference to the decision

Given our specification of discrepancy for our built model we can determine whether we should build a more complex model by examining the sensitivity of the decision to the discrepancy. We calculate, using standard Monte Carlo methods, the expected value of learning the true value of the discrepancy terms via the partial expected value of perfect information (EVPI),

$$\text{EVPI}(\boldsymbol{\delta}) = E_{\boldsymbol{\delta}}\{\max_d E_{\mathbf{X}|\boldsymbol{\delta}}(NB_d)\} - \max_d E(NB_d).$$

where \mathbf{X} is the vector of model inputs, and $\boldsymbol{\delta}$ is the vector of discrepancy terms. If $\text{EVPI}(\boldsymbol{\delta})$ is large compared with the value of learning the inputs, $\text{EVPI}(\mathbf{X})$, then we conclude that the potential structural error in adopting the simple Markov model is important.

The expected value of learning the discrepancy terms, $\text{EVPI}(\boldsymbol{\delta})$, is the ‘expected value of model improvement’ (EVMI) under the assumption that any new input parameters that are introduced into the model during the structural improvement are known with certainty. It is likely however that model improvement will involve the addition of new *uncertain* input parameters. In this case the $\text{EVPI}(\boldsymbol{\delta})$ provides an upper bound for the EVMI. If the $\text{EVPI}(\boldsymbol{\delta})$ is small this offers us some reassurance that the model is good enough for the decision, whereas if it is large we know our uncertainty about the model structure is resulting in decision uncertainty. In the latter case improving the model may be worthwhile, but this will depend on the degree of decision uncertainty induced by any newly introduced uncertain inputs.

3.5 Case study scenario 2 - an uncertain relationship between efficacy and time since treatment commencement

The duration of effect of the combination therapy was a key uncertainty at the time of publication of Chancellor et al. (1997), and the authors presented results for three alternative scenarios: effectiveness lasting one year, two years and 20 years. We ask the following question: if our built model assumes that the combination therapy is effective over 20 years, but we are uncertain whether this is true, do we need to build a more complex model that incorporates an uncertain relationship between efficacy and time from commencement of treatment?

The treatment effect acts on six unconstrained terms in the transition matrix for the combination therapy (Eq. 3), but does not act on the transition matrix for the monotherapy, therefore resulting in six non-zero discrepancies per time step, $\delta_{2,1,t}, \dots, \delta_{2,6,t}$. This specification of discrepancy is equivalent to incorporating a time varying treatment effect parameter (RR), but with the additional flexibility that allows the treatment effect to vary across the different transitions in the model (e.g. HIV to AIDS versus HIV to death).

We believe that efficacy falls over time, and therefore that the discrepancy between our built model and reality increases over time. We again chose a linear

mean function $E(\delta_{2it}) = \beta_{0,i} + \beta_{1,i}t$ with uncertain slope. The intercept parameter $\beta_{0,i}$ is zero in this case to reflect our judgement that during time step 1 the treatment effect parameter RR correctly determines the effectiveness of the combination therapy. We placed normal distributions on the six $\beta_{1,i}$ parameters with hyperparameters $\mu_{\beta_{1,i}}$ and $\sigma_{\beta_{1,i}}^2$ shown in Table 4.

Next, we specify the covariance function. Our uncertainty about the six discrepancies $\delta_{2,1,t}, \dots, \delta_{2,6,t}$ is controlled through variance terms $\sigma_{2,1}^2, \dots, \sigma_{2,6}^2$, assumed to hold for all t . We specify our judgement about the dependency between the discrepancy terms for the six transitions through a single correlation parameter $\phi_{i,i^*} = \phi \forall i \neq i^*$ which we assume constant for all t . Since there is no discrepancy for the monotherapy option $d = 1$ in this scenario we do not need to specify between-decision correlations (i.e. there is no ψ_{d,d^*} correlation parameter). Finally we specify a correlation structure for the discrepancies as they evolve through time via a Gaussian form correlation function with parameter ω (Eq. 9), ensuring via Eq. (16) that discrepancy as a function of time is monotone with probability $\alpha = 0.95$. Values for all covariance function parameters are shown in Table 5.

Ten samples from the Gaussian process for discrepancy term $\delta_{2,1,t}$ are shown in figure 5. Note the variation in functional form generated by the Gaussian process, reflecting our uncertainty about the relationship between efficacy and time.

Figure 5: Ten samples from the distribution on discrepancy term $\delta_{2,1,t}$ in scenario 2

3.6 Case study scenario 3 - relaxation of the memoryless property

In scenario 3 we judge that the probability of transition to state y at time step $t+1$ is dependent not only on the state x at time t but on the states occupied at time steps $\leq t-1$. We therefore want to relax the Markov assumption and consider more complex time dependencies that would necessitate a more flexible modelling framework (for example using a discrete event or agent based approach). In order to judge whether this is necessary we add relatively unstructured discrepancy to allow for a wide range of possible deviations from the simple memoryless Markov process. Hyperparameters are shown in Tables 4 and 5. Ten samples from the Gaussian process for discrepancy term $\delta_{2,1,t}$ are shown in figure 6.

Figure 6: Ten samples from the distribution on discrepancy term $\delta_{2,1,t}$ in scenario 3

4 Results

4.1 Base case model

We implemented the model in R (R Development Core Team, 2011). We sampled from the base case model input parameters and ran the model 10,000 times. The mean incremental cost of combination therapy over monotherapy was £45,402 and the mean incremental benefit was 3.86 life years, representing a cost per life year gained of £11,749. Value of information analysis with $\lambda = \text{£}12,000$ per life year¹ suggests that decision uncertainty is being driven by uncertainty in the treatment effect parameter with $\text{EVPI}(\text{RR}) = \text{£}169.91$ (EVPI index, 46.5%), and uncertainty in the cost parameters with $\text{EVPI}(\text{costs}) = \text{£}194.41$ (53.2%). See Table 6.

¹We assumed for the purposes of this case study a willingness to pay of $\lambda = \text{£}12,000$ per life year to ensure that we were in the region of decision uncertainty. This is lower than the value that would be used for decisions in many Western health economies.

Table 6: Partial EVPI results

Parameter	Partial EVPI (EVPI index [†])			
	Base case	Scenario 1	Scenario 2	Scenario 3
Transition probabilities	£0 (0%)	£0 (0%)	£0 (0%)	£1.17 (0.1%)
Relative risk	£169.91 (46.5%)	£193.09 (48.1%)	£64.63 (19.4%)	£164.55 (17.2%)
Costs	£194.41 (53.2%)	£201.72 (50.2%)	£65.17 (19.55%)	£167.53 (17.5%)
Discrepancy terms	-	£7.86 (2.0%)	£110.21 (33.1%)	£699.06 (73.0%)
Overall EVPI	£365.42	£401.53	£333.43	£957.28

[†] The partial EVPI as a proportion of the overall EVPI

4.2 Scenario 1

After the addition of discrepancy to reflect the judgements about model error due to the time homogeneity assumption, the mean incremental cost of combination therapy over monotherapy was £44,697 and the mean incremental benefit was 3.80 life years, representing a cost per life year gained of £11,769. Value of information analysis suggests that the decision uncertainty is still dominated by the uncertainty in the inputs with $EVPI(RR) = £193.09$ (48.1%) and $EVPI(costs) = £201.72$ (50.2%). There is little value in learning δ with $EVPI(\delta) = £7.86$ (2.0%), indicating that building a more complex model is not advisable at a willingness to pay for one life year of $\lambda = £12,000$.

It appears that uncertainty regarding the model error that results from the time homogeneity assumption is not a significant driver of decision uncertainty. This would suggest that our simple built model is ‘good enough’ for the decision in this scenario.

4.3 Scenario 2

After the addition of discrepancy terms to reflect the judgements about model error due to the constant treatment efficacy assumption, the mean incremental cost of combination therapy over monotherapy was £39,741 and the mean incremental benefit was 3.20 life years, representing a cost per life year gained of £12,409. Value of information analysis suggests that although there is still some value in learning the treatment effect and cost parameters, it is the discrepancy terms that are now most important in driving decision uncertainty. In this scenario there is value in improving the model such that it better reflects our judgements about the decision problem, as well as value in reducing parameter uncertainty.

4.4 Scenario 3

After the addition of discrepancy terms to reflect the judgements about model error due to the Markovian assumption of memorylessness, the mean incremental cost of combination therapy over monotherapy was £45,111 and the mean incremental benefit was 3.84 life years, representing a cost per life year gained of £11,744. Value of information analysis suggests that the decision is again sensitive to the discrepancy terms, and that building a more complex model to better represent non-Markovian transitions between health states may be worthwhile.

5 Discussion

We have demonstrated a method for incorporating within a model judgements about the structural error that results from building an ‘incorrect’ model. The method allows us to determine the expected value of building a better model, given the uncertainty in the input parameters. This approach will be most valuable in cases where the decision problem is complex, but due to difficulties in obtaining input parameter estimates or lack of time or resources we have built a simple model. We feel that this may be of particular relevance in the emerging field of economic evaluation of public health interventions where decision problems generally have many complex elements, but models are often relatively simple (for good examples see descriptions of the models that have been used by the National Institute for Health and Clinical Excellence to support public health intervention resources allocation decisions in England, <http://www.nice.org.uk/Guidance/PHG/Published>).

The most important potential practical limitation of the method lies on our ability to meaningfully specify a distribution on the discrepancies, $p(\boldsymbol{\delta})$. This only needs to be done fairly crudely, as long as we are ‘generous’ with our specification of uncertainty. The expected value of learning $\boldsymbol{\delta}$ will then provide an upper bound on the value of better modelling. If $EVPI(\boldsymbol{\delta})$ is small compared with the value of learning the inputs, even with the generous estimate of uncertainty about the structural error, then we can be reassured that the current model is ‘good enough’. In contrast, if $EVPI(\boldsymbol{\delta})$ dominates $EVPI(\mathbf{X})$ then we conclude that it will be worthwhile to rebuild the model so that it better reflects our beliefs about the relationships between the inputs and the target quantities we wish to predict.

We believe that the discrepancy approach to managing structural uncertainty

offers significant advantages over other methods proposed in the literature for addressing the problem of health economic model adequacy. Current alternative methods fall into two categories: those based on scenario analysis (NICE, 2009; Kim et al., 2010), and those based on model averaging (Bojke et al., 2009; Jackson et al., 2009, 2010). Both rely on the (implicit or explicit) synthesis of the results of a set of plausible models that differ in their structural assumptions. See Bilcke et al. (2011) and Jackson et al. (2011) for a general discussion of this broad approach to managing structural uncertainty in health economic evaluation.

Scenario analysis based methods generate a series of model outputs under a discrete set of assumptions about the model structure, leaving the decision maker to implicitly weight the outputs according to the plausibility of the structural assumptions. Although the simplicity of this approach is attractive, it is not clear how a decision maker should construct these implicit weights.

Model averaging is a more refined version in which the outputs from a set of models are weighted according to an explicit criterion. This criterion could be the posterior probability that the model is ‘correct’ conditional on observations on the model output if these are available (see Kadane and Lazar, 2004, for a review of model averaging in general), or some measure of the predictive ability of the model given a replicate data set (Jackson et al., 2010). Ultimately, however, model averaging in the context of health economic evaluation suffers from the fundamental problem of an absence of data against which to measure the adequacy of the model in its entirety. We do not measure overall costs and health effects over extended time periods under competing decision options. We can in theory elicit judgements about model adequacy, but making probability statements about models, which are by definition abstract non-observables is likely to be very difficult. The sub-function discrepancy terms identified in our method are, by contrast, defined such that they relate to observables, precisely so that judgements about them are easier to elicit.

References

- Bilcke, J., Beutels, P., Brisson, M. and Jit, M. (2011). Accounting for methodological, structural, and parameter uncertainty in decision-analytic models: A practical guide, *Medical Decision Making*, DOI: 10.1177/0272989X11409240.
- Bojke, L., Claxton, K., Sculpher, M. and Palmer, S. (2009). Characterizing structural uncertainty in decision analytic models: A review and application of methods, *Value in Health*, **12** (5): 739–749.
- Briggs, A., Claxton, K. and Sculpher, M. (2006). *Decision modelling for health economic evaluation*, Oxford: Oxford University Press.
- Chancellor, J. V., Hill, A. M., Sabin, C. A., Simpson, K. N. and Youle, M. (1997). Modelling the cost effectiveness of lamivudine/zidovudine combination therapy in hiv infection, *PharmacoEconomics*, **12** (1).
- Drummond, M. F., Sculpher, M. J., Torrance, G. W., O'Brien, B. and Stoddart, G. (2005). *Methods for the economic evaluation of health care programmes*, Oxford: Oxford University Press, third edn.
- Jackson, C. H., Bojke, L., Thompson, S. G., Claxton, K. and Sharples, L. D. (2011). A framework for addressing structural uncertainty in decision models, *Medical Decision Making*, **31** (4): 662–674.
- Jackson, C. H., Sharples, L. D. and Thompson, S. G. (2010). Structural and parameter uncertainty in bayesian cost-effectiveness models, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **59** (2): 233–253.
- Jackson, C. H., Thompson, S. G. and Sharples, L. D. (2009). Accounting for uncertainty in health economic decision models by using model averaging, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **172** (2): 383–404.
- Kadane, J. B. and Lazar, N. A. (2004). Methods and criteria for model selection, *Journal of the American Statistical Association*, **99** (465): 279–290.
- Kim, S.-Y., Goldie, S. J. and Salomon, J. A. (2010). Exploring model uncertainty in economic evaluation of health interventions: The example of rotavirus vaccination in vietnam, *Medical Decision Making*, **30** (5): E1–E28.

- Krahn, M. and Gafni, A. (1993). Discounting in the economic evaluation of health care interventions, *Medical Care*, **31** (5): 403–418.
- Neal, R. M. (1999). Regression and classification using gaussian process priors (with discussion), *Bayesian Statistics*, **6**: 475–501.
- NICE (2009). Methods for the development of nice public health guidance (second edition), Tech. rep., NICE.
- O’Hagan, A. (1992). Some Bayesian numerical analysis, in *Bayesian Statistics*, edited by Bernardo, J. M., Berger, J. O., Dawid, A. P. and M. Smith, A. F., vol. 4, pp. 345–363, Oxford University Press.
- R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.
- Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M. and Tarantola, S. (2008). *Global Sensitivity Analysis: The Primer*, Chichester: John Wiley and Sons Ltd.
- Strong, M., Oakley, J. E. and Chilcott, J. (2011). Managing structural uncertainty in health economic decision models: a discrepancy approach, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, In Press.