Linear Regression Metamodeling as a Tool to Summarize and Present Simulation Model Results

Hawre Jalal, MD, MSc, Bryan Dowd, PhD, François Sainfort, PhD, Karen M. Kuntz, ScD

Background/Objective. *Modelers lack a tool to systematically and clearly present complex model results, including those from sensitivity analyses. The objective was to propose linear regression metamodeling as a tool to increase transparency of decision analytic models and better communicate their results.* ***Methods***. *We used a simplified cancer cure model to demonstrate our approach. The model computed the lifetime cost and benefit of 3 treatment options for cancer patients. We simulated 10,000 cohorts in a probabilistic sensitivity analysis (PSA) and regressed the model outcomes on the standardized input parameter values in a set of regression analyses. We used the regression coefficients to describe measures of sensitivity analyses, including threshold and parameter sensitivity analyses. We also compared the results of the PSA to deterministic full-factorial and one-factor-at-a-time designs.* ***Results***. *The regression intercept represented the estimated base-case outcome, and the other coefficients described the relative parameter uncertainty in the model. We defined simple relationships that compute the average and incremental net benefit of each intervention. Metamodeling produced outputs similar to traditional deterministic 1-way or 2-way sensitivity analyses but was more reliable since it used all parameter values.* ***Conclusions***. *Linear regression metamodeling is a simple, yet powerful, tool that can assist modelers in communicating model characteristics and sensitivity analyses.* ***Key words:*** *sensitivity analysis; cost-effectiveness; economic evaluation; decision analysis; design of experiments; metamodel; regression.* ***(Med Decis Making 2013;33:880–890)***

Sensitivity analysis communicates decision robustness, or lack thereof, and is an important component of decision analysis. A wide array of tools are available to conduct sensitivity analyses, ranging from changing 1 parameter at a time to value-of-information analyses. New methods of sensitivity analyses often address the shortcomings of previous approaches, but they introduce new challenges. To date, modelers lack a single tool that can clearly and systematically report sensitivity analyses.

One-way sensitivity analyses are recommended as a first step in evaluating uncertainty, where 1 input parameter value is varied within a meaningful range (e.g., 95% confidence interval) while the rest of the parameters are held equal to their mean values.[1] An analyst also may be interested in identifying threshold values at which the preferred intervention changes.[2] These approaches provide a simple overall picture of the model behavior. However, deterministic sensitivity analysis suffers from two important limitations. First, it strikes a compromise between clarity and completeness. An analyst can vary only a limited number of parameters simultaneously without losing clarity. Thus, the model outcomes do not adequately reflect the combined uncertainty in all parameters. Second, deterministic sensitivity analysis only partially illustrates uncertainty because it ignores the parameters' distributions.

Probabilistic sensitivity analysis (PSA) addresses these limitations by assigning distributions to all parameter uncertainties and sampling from these

Received 4 June 2012 from the Division of Health Policy and Management, School of Public Health, University of Minnesota, Minneapolis, MN (HJ, BD, FS, KMK). This research was supported by the Agency for Healthcare Research and Quality Dissertation Grant (R36HS020868) and the National Institutes of Health (RC1AR058601). Revision accepted for publication 6 May 2013.

Address correspondence to Karen M. Kuntz, Division of Health Policy and Management, School of Public Health, University of Minnesota, MMC 729, 420 Delaware Street SE, Minneapolis, MN 55455; telephone: (612) 625-9333; fax: (612) 624-2196; e-mail: kmkuntz@umn.edu.

DOI: 10.1177/0272989X13492014

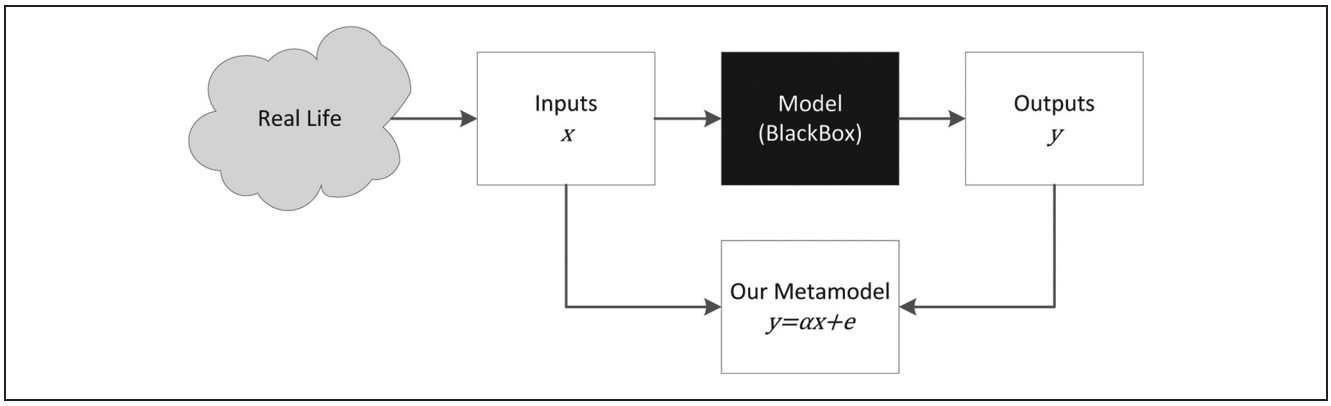

Figure 1 Metamodel as an add-on to simulation models. Simulation models are approximations of real-life complexity. Metamodels simplify and increase model transparency by summarizing the relationship between model inputs and outputs.

distributions simultaneously.[3] Thus, guidelines often recommend using PSA in decision analysis because it is a better representation of the model outcomes under the overall parameter uncertainties.[4] However, PSA does not show how uncertainties in estimating the individual parameters affect the outcome. Expected value of information (EVI) methods are used to quantify the value of reducing model uncertainty in some or all of the parameters; however, EVI computations can be challenging.[5] Thus, no single measure of sensitivity analyses is perfect and decision analysts often are encouraged to perform a combination of deterministic and probabilistic sensitivity analyses.[6]

A metamodel, or model of the model, is a second model that simplifies the relationship between the inputs and outputs of a simulation model. Metamodels have been used successfully as surrogates of simulation models for nearly half a century. Blanning[7] and Kleijnen[8] were among the first to operationalize the term *metamodel*. The original motivation of metamodeling was to reduce computing costs by defining a simpler mathematical relationship between the model outputs and inputs than the actual model. Once sufficient model results are gathered to define and validate a metamodel, the metamodel is used to replace the model upon which it is built. Although model replacement continues to be a valid application of metamodeling, we focus on the advantages of using metamodeling to increase model transparency and to improve the presentation of model results.

METHODS

Linear Regression Metamodeling

Figure 1 illustrates a metamodel with respect to model inputs and outputs. Linear regression metamodeling involves defining a model outcome as a linear function of the model input parameters. For example, in a simple model with 2 input parameters, we can define the outcome as

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + e \quad (1)$$

where y is a model outcome (e.g., the simulated lifetime cost of a surgical intervention), x_1 and x_2 are 2 input parameters (e.g., the probability of failing surgery and the cost of the surgical procedure, respectively), α_0 is the intercept, α_1 and α_2 are the coefficients of the model input parameters x_1 and x_2, respectively, and e is the residual term.

Using linear regression, one can easily calculate the α coefficients. However, raw regression coefficients are sensitive to scale, which makes them more difficult to interpret in relation to one another. Model parameters can be rescaled to overcome the above limitations. One scaling method is to standardize, or normalize, the parameters. The standardized value of the j^{th} parameter z_j is defined as

$$z_j = \frac{x_j - \bar{x}_j}{\sigma_j},$$

where $\bar{x}_j$ and σ_j are the mean and standard deviation (s) values for the input parameter x_j, respectively. Using the standardized input parameters (but not standardizing the outcome), produces

$$y = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + e, \quad (2)$$

where β_0 is the new intercept, and β_1 and β_2 are the coefficients of the standardized parameters z_1 and z_2, respectively.

The intercept β_0 is equal to the base-case outcome when all the parameters are set equal to their mean

values. In addition, β_1 and β_2 represent the change in the outcome due to 1 *s* change in x_1 and x_2, respectively. Thus, these coefficients communicate parameter uncertainty. Furthermore, the absolute values of these coefficients can be used to rank the inputs by their relative uncertainty.

Cost-effectiveness analyses (CEAs) require a composite measure of an intervention's expected cost and benefit. One such quantity is the net health benefit (NHB), which is defined as

$$\mathrm{NHB} = \mathrm{E} - \frac{\mathrm{C}}{\lambda}, \quad (3)$$

where E and C are the computed benefit and the cost of the intervention, respectively, and λ is the willingness to pay (WTP) threshold per unit of health. NHB is measured in units of health and is calculated as the difference between the benefit gained investing in an intervention and what could have been obtained if the resources were instead invested in a marginally cost-effective alternative.[9]

We can express each of the 3 outcomes (NHB, C, and E) as a separate linear function of the standardized input parameters such that

$$\theta_0 + \theta_1 z_1 + \theta_2 z_2 = \gamma_0 + \gamma_1 z_1 + \gamma_2 z_2 - \frac{\delta_0 + \delta_1 z_1 + \delta_2 z_2}{\lambda}, \quad (4)$$

where θ_j, γ_j, and δ_j are the regression coefficients of standardized parameter z_j for the outcomes NHB, E, and C, respectively. Differentiating Equation 4 with respect to each standardized parameters z_j results in

$$\theta_j = \gamma_j - \frac{\delta_j}{\lambda}, \quad (5)$$

where the variables are as previously described. Equation 5 describes the effect of 1 *s* change in a particular parameter *j* on the NHB. It also illustrates that the θ_j coefficients can be derived directly from the γ_j and δ_j parameters, which allows the analyst to incorporate varying λ values. We can also calculate the predicted change in the incremental NHB (ΔNHB) of intervention T_1 relative to T_0 from a 1 *s* change in parameter *x* as

$$\Delta\theta_j = \Delta\gamma_j - \frac{\Delta\delta_j}{\lambda}, \quad (6)$$

where $\Delta\theta_j = \theta_j^{T_1} - \theta_j^{T_0}$, $\Delta\gamma_j = \gamma_j^{T_1} - \gamma_j^{T_0}$, and $\Delta\delta = \delta_j^{T_1} - \delta_j^{T_0}$. All 3 expressions ($\Delta\theta_j$, $\Delta\gamma_j$, and $\Delta\delta_j/\lambda$) are measured in units of health (e.g., quality-adjusted life-years, QALYs).

Threshold Analysis

We can use Equation 6 to conduct threshold analyses to identify a parameter's value (if any) at which an intervention T_1 becomes (or ceases to be) preferred relative to another intervention T_0 while holding the other parameters at their mean values. A decision maker is indifferent between T_1 and T_0 when they break even (i.e., ΔNHB = 0), or

$$\Delta\theta_0 + \Delta\theta_1 z_1 + \Delta\theta_2 z_2 = 0. \quad (7)$$

In a threshold analysis, except for the j^{th} parameter for which we are interested in calculating the threshold, the rest of the standardized parameters are equal to zero. Thus, the break-even value for a parameter z_j can be calculated as

$$z_j^* = -\frac{\Delta\theta_0}{\Delta\theta_j},$$

where z_j^* is the threshold value at which a decision maker becomes indifferent between T_0 and T_1, $\Delta\theta_0$ is the base-case ΔNHB where all parameters are equal to their mean values, and $\Delta\theta_j$ is the rate at which ΔNHB changes per unit change in z_j.

The value z_j^* divides z_j into 2 regions: $z_j > z_j^*$ and $z_j < z_j^*$. The region of z_j where T_1 is preferred over T_0 depends on how the incremental NHB (ΔNHB) changes with respect to z_j, and it is measured by $\Delta\theta_j$. If this relationship is positive ($\Delta\theta_j > 0$), T_1 is preferred when $z_j > z_j^*$. The reverse is also true. In the case of a negative relationship between z_j and ΔNHB, T_1 is preferred when $z_j < z_j^*$. Obviously, as $\Delta\theta_j$ approaches zero, z_j^* approaches infinity, indicating that z_j must be infinitely large (or small) to change the preferred intervention.

To present the threshold values on their original scales,

$$\begin{aligned} x_j &> \bar{x}_j + z_j^*\sigma_j \quad \text{if} \quad \Delta\theta_j > 0, \quad \textit{and} \\ x_j &< \bar{x}_j + z_j^*\sigma_j \quad \text{if} \quad \Delta\theta_j < 0, \end{aligned}$$

where all the variables are as defined previously.

It is also possible to use metamodel results to examine the impact on ΔNHB of T_1 relative to T_0 while changing 2 standardized parameters (e.g., z_1 and z_2) simultaneously and holding the rest of the parameters equal to their mean estimates (i.e., 2-way threshold analysis). The concept of interaction in regression analysis is closely related to 2-way sensitivity analysis. Adding an interaction (cross-product) term to Equation 7 and redefining the relationship as an inequality produces

$$\Delta\theta_0 + \Delta\theta_1 z_1 + \Delta\theta_2 z_2 + \Delta\theta_3 z_1 z_2 > 0, \quad (8)$$

where $\Delta\theta_3$ is the coefficient of the interaction term $z_1 z_2$. This coefficient captures the additional effect of z_1 on ΔNHB when z_2 is increased by 1. This coefficient can also be interpreted as the additional effect of z_2 on ΔNHB when z_1 is increased by 1. In a perfectly linear model, adding interaction terms does not change the main effects of z_1 and z_2 because all the parameters are standardized.

The constraint shown in Equation 8 compares 2 interventions at a time. To compare K interventions, one must recompute this constraint for all $K(K-1)/2$ intervention pairs possible. Thus, for a particular intervention T to be optimal at specific values of z_1 and z_2, it must satisfy all $(K-1)$ constraints that involve T. The result of this analysis also can be presented graphically.

A Simplified Example

We demonstrate our approach with a simplified decision problem using TreeAge Pro (TreeAge Software, Inc., Williamstown, MA) that evaluates the treatment options for 50-year-old patients with a newly diagnosed, potentially curable cancer. The decision tree compares 3 interventions: chemotherapy, radiation, and surgery. Each intervention is represented by 3 Markov health states shown in Figure 2. These states are cured (cancer-free), cancer, and dead. If cured, quality of life increases from 0.8 to 1. The probability of cure depends on the type of intervention. Once cured, the patient stays cancer-free until dying of causes other than cancer. The risk of death in the cured state is obtained from standard US life tables and depends on the patient's age. The risk of death increases for those in the cancer state. We assumed a cycle length of 1 year and used a half-cycle correction. Furthermore, we used a willingness to pay threshold (λ) of $50,000 per QALY and discounted the interventions' costs and benefits by 5% annually.

An intervention may be successful and the patient may be cured, or the treatment may fail and the patient may stay in the cancer state. Chemotherapy and radiation can be tried only for a few cycles (5 and 4 years, respectively). During this period, the probability of success is assumed to be constant. If treatment fails after this period, the patient stays in the cancer state. Surgery can be tried only once and its benefit is assumed to be instantaneous; however, it also carries an immediate risk of death during the procedure.

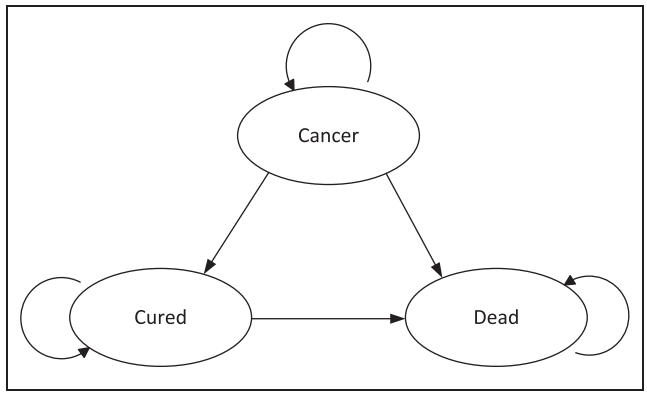

Figure 2 Markov-state diagram. The ovals represent the Markov states and the arrows indicate the allowed transitions.

Table 1 summarizes the parameter distributions used in this model. All model parameters are fictitious and are not based on evidence. The sole purpose of this exercise is to demonstrate the method outlined above using a simple transparent model.

Model Runs

We conducted both probabilistic sampling and deterministic designs. In the probabilistic parameter sampling (i.e., PSA), we simulated each intervention's outcome for 10,000 hypothetical cohorts, the characteristics of which were sampled from 8 independent random distributions described in Table 1. We also generated parameter sets using deterministic designs, where we changed the values of each of the 8 parameters deterministically among the mean, lower bound (LB), and upper bound (UB) shown in Table 1. For simplicity, we set the LB and UB equal to $-2\,s$ and $+2\,s$ from the mean, respectively. We used 2 types of deterministic designs, one-factor-at-a-time and full-factorial designs. We performed regression metamodeling on the output generated from each design.

In the one-factor-at-a-time design, we assumed that only 1 parameter changes at a time and the rest of the parameters were equal to their mean values. For example, in the first model run, only the value of the first parameter is equal to its LB, but the rest of the parameters are equal to their mean values. In the second run, the first parameter is changed to its UB. In the third simulation, the value of the first parameter is returned to its mean, and the value of the second parameter is changed to its LB, and so

Table 1 Input Parameter Descriptions

Parameter	Description	Distribution	Mean	*s*	LB	UB
pFailChemo	Annual probability of failing chemotherapy	Beta(45,55)	0.45	0.050	0.353	0.547
pFailRadio	Annual probability of failing radiotherapy	Beta(50,50)	0.5	0.050	0.402	0.598
pFailSurg	Probability of failing surgery	Beta(5,95)	0.05	0.022	0.007	0.093
pDieSurg	Probability of dying because of surgery	Beta(10,90)	0.1	0.030	0.041	0.159
μCancer	Cancer-specific mortality rate	LN(−1.69,0.4)	0.2	0.083	0.037	0.363
cChemo	Annual chemotherapy cost ($)	LN(9.9,0.1)	20,000	2005	16,066	23,934
cRadio	Annual radiotherapy cost ($)	LN(9.17,0.3)	10,000	3069	3979	16,021
cSurg	Surgery cost ($)	LN(10.05,0.4)	25,000	10,414	4568	45,432

Note: LN(μ,σ) = log normal distribution with parameters μ and σ; Beta(α,β) = beta distribution with shape parameters α and β; *s* = standard deviation; LB = lower bound (−2 *s*); UB = upper bound (+ 2 *s*).

on. This design resulted in 16 (=8 parameters × 2 runs/parameter) model runs.

In the full-factorial design, all possible LB and UB values for each parameter were sampled. In this design, no parameter was held at its mean value for any simulation. This design is particularly useful to detect interaction between parameters with a limited number of simulations. This design resulted in 256 ($=2^8$) scenarios. Similar to the PSA design, metamodeling following full-factorial design also explores the entire parameter space.[10,11]

Figure 3 illustrates input sampling from PSA, full-factorial, and one-factor-at-a-time designs with 2 inputs: the cancer mortality rate (μCancer) and the probability of failing chemotherapy (pFailChemo). The full-factorial design explores the edges, or corners, of the parameter space by equating each parameter to its UB or LB, whereas in the one-factor-at-a-time design only 1 parameter is changed to either its UB or its LB. In PSA, each simulation represents a hypothetical cohort. For each set of input samples, whether stochastic or deterministic, we computed the lifetime cost and benefit for that particular parameter set. One-factor-at-a-time, full-factorial, and PSA produced 16, 256, and 10,000 model runs, respectively.

Statistical Analysis

The 3 simulation data sets were exported into Stata (Stata Statistical Software, Inc., College Station, TX) for regression analyses. Each data set consisted of 14 variables (8 input parameters and 6 outcomes that represent the cost and benefit of the 3 interventions). We primarily used the PSA data set in our analyses. First, we regressed each intervention's outcome on the unstandardized parameters. Then, we repeated the analyses on the standardized parameters. In addition, we performed further regressions to compare the results from the PSA with the one-factor-at-a-time and full-factorial designs.

RESULTS

Regression Analysis of the Example

Table 2 presents the results of regressing the unstandardized and standardized parameters on the cost, benefit, and NHB of the chemotherapy intervention. The intercept of the unstandardized regression is the expected outcome when all parameters are set to zero. The coefficients of the unstandardized regressions are interpreted relative to a 1-unit change in each parameter on the original scales. For example, changing the cost of chemotherapy (cChemo) by $1 does not have a noticeable impact on the NHB of chemotherapy, and increasing the chance of failing chemotherapy (pFailChemo) from 0 to 1 increases the cost of chemotherapy by nearly $36,000 and reduces the benefit by 7 QALYs.

In the standardized regressions, the intercept is the expected outcome when all the parameters are equal to their mean values (i.e., their base-case estimates). The other coefficients are interpreted in units of standard deviations. For example, 1 *s* change in the cost of chemotherapy (cChemo) increases the cost of this intervention by $2060, has no noticeable impact on the benefit, and reduces the NHB by 0.04 QALYs. Furthermore, the absolute value of the coefficients of the standardized parameters can be used to rank parameters by their importance. For example, the uncertainty in the mortality rate of cancer (μCancer) has the greatest impact on the NHB of chemotherapy, followed by the probability of failing chemotherapy (pFailChemo) and the cost of chemotherapy (cChemo). Table 3 shows the results of regressing

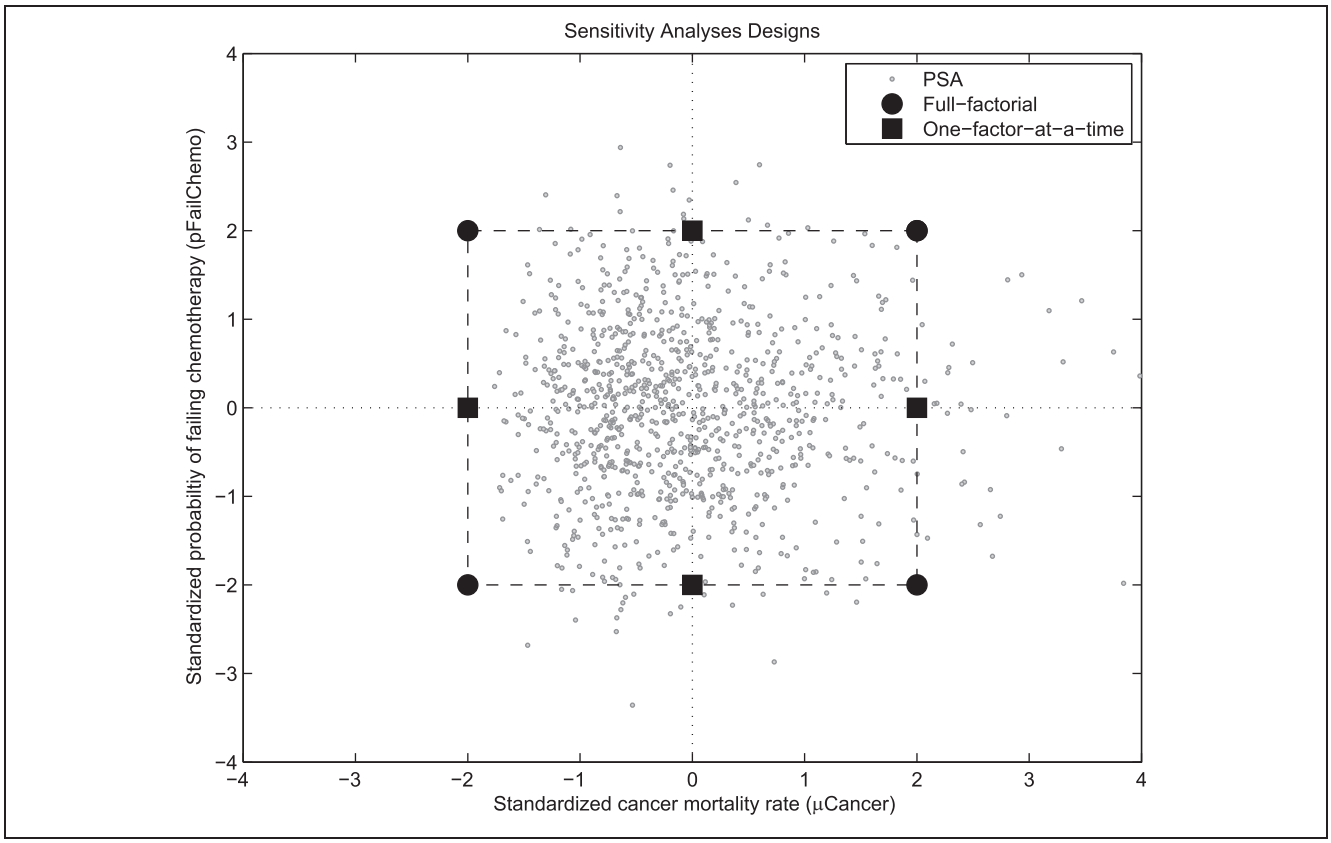

Figure 3 Sensitivity analysis design. This figure shows how 2 parameter values are sampled in different types of simulation designs. Cancer mortality rate (μCancer) is plotted on the X axis and the probability of failing chemotherapy (pFailChemo) is plotted on the Y axis. In probabilistic sensitivity analysis (PSA), 1000 parameter pairs are sampled randomly from their distributions (Table 1). In deterministic designs, the analyst chooses the input values. In 2^K full-factorial design, the inputs are sampled from the edges, or corners, of the parameter space. In contrast, in one-factor-at-a-time design, only 1 parameter is varied at a time, the other parameter is held at its mean value.

Table 2 Regression Coefficients from Metamodels Using Unstandardized and Standardized Model Parameters on the Outcomes of Chemotherapy

	Dependent Variables (*Y*)					
Predictors (*X*)	**Cost**	**Benefit**	**NHB**	**Cost**	**Benefit**	**NHB**
(Parameters)	**With Unstandardized Predictors**			**With Standardized Predictors**		
Intercept	−13,066	18.01	18.27	20,796	13.49	13.08
pFailChemo	35,545	−7.116	−7.827	1742	−0.349	−0.384
pFailRadio	1	0.001	0.001	0	0.000	0.000
pFailSurg	91	0.009	0.008	2	0.000	0.000
pDieSurg	156	0.081	0.078	5	0.002	0.002
μCancer	−14,678	−6.599	−6.305	−1226	−0.551	−0.527
cChemo	1	0.000	0.000	2060	−0.001	−0.042
cRadio	0	0.000	0.000	0	0.001	0.001
cSurg	0	0.000	0.000	−2	0.000	0.000
Observations	10,000	10,000	10,000	10,000	10,000	10,000
R^2	0.98	0.97	0.98	0.98	0.97	0.98

Note: NHB = net health benefit.

Table 3 Standardized Parameters and Their Interactions on All Model Outcomes

	Chemotherapy			Radiation			Surgery		
	Cost	Benefit	NHB	Cost	Benefit	NHB	Cost	Benefit	NHB
Model base-case	20,595	13.46	13.05	10,983	12.94	12.72	25,000	13.44	12.94
Metamodel parameter									
Intercept	20,796	13.49	13.08	11,058	12.98	12.76	24,912	13.46	12.96
pFailChemo	1744	−0.35	−0.38	2	0	0	0	0	0
pFailRadio	0	0	0	880	−0.45	−0.47	0	0	0
pFailSurg	2	0	0	1	0	0	0	−0.23	−0.23
pDieSurg	0	0	0	1	0	0	0	−0.45	−0.45
μCancer	−1228	−0.55	−0.53	−676	−0.59	−0.58	0	−0.04	−0.04
cChemo	2061	0	−0.04	−2	0	0	0	0	0
cRadio	0	0	0	3430	0	−0.07	0	0	0
cSurg	0	0	0	2	0	0	10,285	0	−0.21
μCancer*pFailChemo	−236	−0.07	−0.06	7	0	0	0	0	0
μCancer*pFailRadio	8	0	0	−117	−0.06	−0.06	0	0	0
μCancer*pFailSurg	1	0	0	1	0	0	0	−0.02	−0.02
Observations	10,000	10,000	10,000	10,000	10,000	10,000	10,000	10,000	10,000
R^2	1	0.98	0.98	1	0.98	0.98	1	1	1

Note: NHB = net health benefit.

the standardized parameters and their interactions on the cost, benefit, and NHB of all 3 interventions. Notice that the main effects and the intercept for chemotherapy are similar to those in Table 2. This is because all the parameters are centered around zero and the model is nearly linear. We also show the model's base-case estimates when all parameters are equal to their mean values. The metamodel's intercept reproduced the base-case results extremely well confirming that the effects were locally near-linear.

In Table 4 we summarize the regression results of the standardized parameters and their interactions on the incremental NHB of all the possible comparisons of the 3 interventions. The intercept in this table represents the base-case incremental NHB, and the other coefficients represent the effect of changing the parameters by 1 *s* on the incremental NHB. These coefficients can also be used to conduct threshold analyses for each strategy pair.

Threshold Analysis

In Table 5 we compute the 1-way threshold values for each parameter at which the preferred intervention changes in the pairwise comparisons. These threshold values are shown on the parameters' original and standardized scales. For example, chemotherapy is preferred to radiation as long as the annual cost of chemotherapy (cChemo) is less than \$36,000 (7.8 *s*) while the other parameters are held at their means. In addition, if the cost of chemotherapy is less than \$26,000 (2.9 *s*), chemotherapy is preferred over surgery. As would be expected, the cost of chemotherapy does not have a direct impact on the choice between radiation and surgery. Calculating these threshold values from the model directly produces threshold values of \$34,000 and \$24,000, respectively, which are slightly lower than the metamodel results. The metamodel results are more reliable since they average the outcome over all

Table 4 Results of Regressing the Standardized Parameters and Their Interactions on the ΔNHB

Parameter	Chemo> Radio	Chemo> Surgery	Radio> Surgery
Intercept	0.320	0.120	−0.200
pFailChemo	−0.384	−0.384	0.000
pFailRadio	0.466	0.000	−0.466
pFailSurg	0.000	0.234	0.234
pDieSurg	−0.001	0.446	0.447
μCancer	0.054	−0.484	−0.538
cChemo	−0.041	−0.041	−0.001
cRadio	0.068	0.001	−0.067
cSurg	0.001	0.206	0.205
μCancer*pFailChemo	−0.065	−0.062	0.003
μCancer*pFailRadio	0.061	0.002	−0.059
μCancer*pFailSurg	0.000	0.017	0.017

Note: NHB = net health benefit.

Table 5 One-Way Threshold Analysis

Parameter	Chemo>Radio	Chemo>Surgery	Radio>Surgery
pFailChemo	<0.49 (0.83)	<0.47 (0.31)	NA
pFailRadio	>0.47 (−0.69)	NA	<0.48 (−0.43)
pFailSurg	NA	>0.04 (−0.51)	>0.07 (0.85)
pDieSurg	NA	>0.09 (−0.27)	>0.11 (0.45)
μCancer	>−0.29 (−5.93)	<0.22 (0.25)	<0.17 (−0.37)
cChemo	<35,717 (7.84)	<25,803 (2.89)	NA
cRadio	>−4382 (−4.69)	NA	<904 (−2.96)
cSurg	NA	>18,928 (−0.58)	>35,146 (0.97)

Note: Results are presented for the unstandardized and standardized parameters (in parentheses).

parameter values instead of assuming their mean values.

The results from Table 4 were used to construct 2-way threshold analyses. Figure 4 illustrates a 2-way threshold analysis of cancer mortality rate (μCancer) and the probability of failing chemotherapy (pFailChemo). Each region satisfies 2 (i.e., $K-1$) constraints. For example, chemotherapy is the optimal choice when chemotherapy is preferred over radiation where

$$0.320 + 0.054z_1 - 0.384z_2 - 0.065z_1z_2 > 0,$$

and chemotherapy is preferred over surgery where

$$0.120 - 0.484z_1 - 0.384z_2 - 0.062z_1z_2 > 0,$$

where z_1 and z_2 represent the standardized values of μCancer and pFailChemo, respectively. The coefficients are obtained from the corresponding variables from Table 4.

Table 6 shows the results of using different parameter selection designs on the NHB of chemotherapy. Sixteen scenarios were used in the metamodeling of the one-factor-at-a-time design, 256 were used in the metamodeling of the full-factorial design, and 10,000 were used in the metamodeling of the PSA design. The one-factor-at-a-time design was unable to detect interactions; nevertheless, the beta coefficients from all the designs were very close and the input ranking did not change, indicating that the relationship between the NHB of chemotherapy and the parameters are nearly linear.

DISCUSSION

We developed a simple method to present sensitivity analysis results from CEA models using NHBs and linear regression metamodeling. We illustrated our approach with a simplified example that involves treating cancer patients with chemotherapy, radiation, or surgery.

Today, many practical metamodeling applications exist in the engineering and computer science literature.[12,13] However, only a few applications exist in medical decision making. For example, Merz and Small[14] used logistic regression to identify important variables that can change the decision of anticoagulation in pregnant women with deep vein thrombosis, Tappenden and others[15] investigated several metamodeling techniques for value of information calculations in CEA of interferon-β and glatiramer acetate for multiple sclerosis, and Stevenson and others[16] used Gaussian process modeling in the CEA of treating established osteoporosis.

Our metamodeling approach summarizes CEA results in a transparent manner and reveals important model characteristics. The intercept presents the base-case model outcome (i.e., when all the parameters are equal to their mean values). The other coefficients associated with each standardized parameter represent the change in the outcome due to 1 *s* change of that parameter. Thus, these regression coefficients describe the relative importance of uncertainty in each parameter. In addition to regression coefficients, the importance of input parameters can be ranked using other correlation statistics.[11,17] Linear regression metamodeling may perform better than these statistics, especially when the input parameters are themselves correlated as shown by Merz and Small.[14] We also illustrated how to conduct 1-way and 2-way threshold analyses using the metamodeling coefficients.

An important distinction between our approach and traditional deterministic sensitivity analyses is that metamodeling summarizes the results over the entire parameter space. Traditional deterministic sensitivity analyses vary 1 or 2 parameters and hold

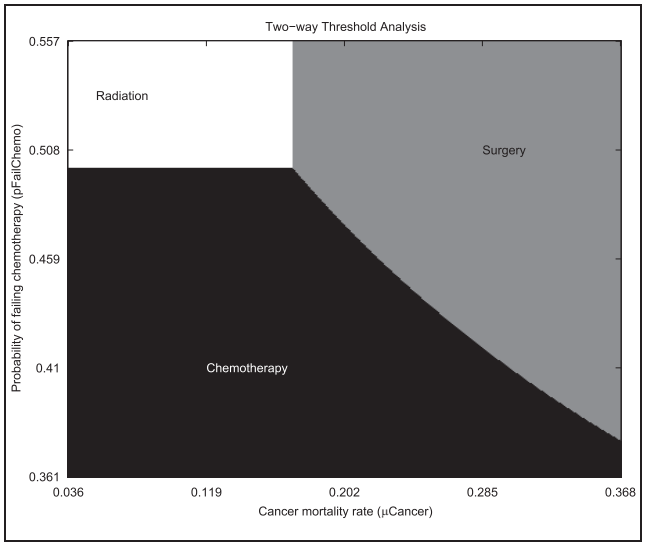

Figure 4 Two-way threshold analysis from regression results. Cancer mortality rate (μCancer) is plotted on the X axis and the probability of failing chemotherapy (pFailChemo) is plotted on the Y axis. The parameters are allowed to range between their lower bound (-2s) and upper bound (+ 2s). The graph shows the optimal strategies for each combination of input parameters, while the rest of the parameters in the model are equal to their mean values.

the rest of the parameters at their mean values. Regression analyses can isolate the correlation between model parameters that vary simultaneously, thus exploring the entire parameter space.

We considered NHB more suitable for our approach than the incremental cost-effectiveness ratio (ICER). Calculating the ICER requires ranking of the interventions that may change in parameter sensitivity analysis. In NHB analyses, there is no ranking involved. Incremental NHB of 2 interventions is always the difference between their average NHBs. Our approach can be used to calculate the ICER in special circumstances. Specifically, our method can be applied if the ranking of the interventions does not change over the ranges of the parameters. This may be the case when all the interventions are compared with a common base-case and ΔC and ΔE are positive.

In NHB analysis, however, a WTP threshold must be specified. One should conduct sensitivity analyses using different WTP thresholds. Fortunately, such analysis does not involve performing extra simulations or reestimating the regression results because a WTP threshold is only needed in the final steps in our approach.

We compared the results from the PSA design to 2 types of deterministic designs: a full-factorial (2^K) design and a one-factor-at-a-time design. In theory, all these designs produce the same results when the model is perfectly linear. In our model, the results from all the designs were similar but not identical. One-factor-at-a-time design is perhaps the simplest form of sensitivity analysis. 2^K design is a popular approach in design of experiments which assumes that the model outcome is linearly related to the parameters. Higher order designs, such as 3^K, 4^K, and 5^K designs, are recommended if an analyst expects a less linear relationship.

Only full-factorial and PSA designs can detect interactions, which commonly occur among a model's input parameters as they define the outcomes. Conditional probabilities are perfect examples of interactions in decision models. For example, we define pDieCancer, which is a function of μCancer, to be conditional on pFailChemo in the chemotherapy intervention. As a result, we expect an interaction between these 2 parameters. In metamodeling, a parameter with a small main effect is generally expected to have a small interaction with the other parameters. In addition, higher level interactions (e.g., 3-way interaction) are generally considered unimportant.[13]

One disadvantage of PSA is the computing cost. The PSA data set that we used consisted of 10,000 scenarios. Obtaining this simulation set using TreeAge Pro took less than 1 minute on a standard desktop PC, and performing the regression analyses in Stata was virtually instantaneous. However, our model was relatively simple, and conducting a similar analysis on a more complex model may still be computationally expensive. For a more complex model, deterministic full-factorial designs, which generally require fewer simulations, may be considered. Partial factorial designs are also available when there are a large number of parameters in the model; for a detailed description, see Kleijnen.[13]

Validation is an important step following metamodel specification.[18] Validation is especially important when metamodels are used to replace computationally costly models (e.g., output prediction in what-if scenarios, and optimization). However, we limited our metamodel specification to enhance our understanding of model results. We considered the R^2 statistics and visual plotting (not shown) adequate measures of overall metamodel fit.[13] Both measures indicated a nearly linear relationship between model inputs and outputs.

Table 6 Results of Regressing the Standardized Parameters on the Outcome of Chemotherapy

Parameter	One-Factor-at-a-Time (2K)	Full-Factorial (2^K)	PSA
Intercept	13.064	13.170	13.080
pFailChemo	−0.393	−0.345	−0.384
μCancer	−0.578	−0.584	−0.526
cChemo	−0.041	−0.042	−0.042
μCancer*pFailChemo	NA	−0.078	−0.062
cChemo*pFailChemo	NA	−0.004	−0.003
cChemo*μCancer	NA	0.003	−0.002
Observations	16	256	10,000
R^2	0.98	1	0.98

Notes: Results are shown for 3 designs: one-factor-at-a-time, full-factorial, and probabilistic sensitivity analysis (PSA).

We only specified a first-order polynomial relationship. Extending our metamodel specification to higher order polynomials is simple using higher order inputs (e.g., $x^2, x^3, \ldots$, etc.). However, including these higher order terms may complicate the interpretation of the regression coefficients as they become functions of the parameter values.

Our approach has some limitations. First, loss of detail is inevitable. State-transition models are nonlinear in general. In these models, linear metamodeling is at best an approximation. Metamodeling trades detail for simplicity and ease of presentation. However, simplicity may outweigh detail in the current presentation of CEA results.

Linear regression is the most widely used metamodeling approach by many researchers for its simplicity and ease of use.[13] Our linear approximation will perform best when the relationship between model inputs and outputs is approximately linear, or locally linear. In most circumstances, the linear metamodel may be a favorable initial analysis, but linear approximation may become problematic when a high level of accuracy is desired in a nonlinear model. Specifically, if the preferred intervention is highly sensitive to a parameter's standardized value, one may consider more advanced sensitivity analyses methods (e.g., EVI).

In this study, we only explored simple threshold values at which a pair of strategies are evaluated using linear regression. A potential extension of our approach is to calculate decision sensitivity, the change in the likelihood of a strategy being optimal because of changing a parameter value. In this case, a categorical dependent variable can be defined as the index of the strategy with the maximum NHB. Then, one can use multinomial logistic regression to measure the relative decision sensitivity of each parameter.

Finally, R^2 statistics may not be a good indicator of metamodels' validity in deterministic designs. R^2 values were generally high in the deterministic designs because of the small sample size (number of model runs in each design) and because each input typically has 2 data points. When the purpose of a metamodel is model replacement (e.g., to perform quick what-if scenarios and reduce computing cost), lack-of-fit statistics (e.g., F-statistics) and cross-validation methods may be more appropriate.[18,19]

CONCLUSION

Linear regression metamodeling can reveal important characteristics of CEA models including the base-case results, relative parameter importance, interaction, and threshold and sensitivity analyses. We recommend using metamodeling as an important step of sensitivity analyses in economic evaluation. It may also prove to be the only analysis needed.

ACKNOWLEDGMENTS

The authors thank two anonymous reviewers for their suggestions and feedback.

REFERENCES

1. Hunink M, Glasziou P, Siegel J, et al. Decision Making in Health and Medicine: Integrating Evidence and Values. Cambridge (UK): Cambridge University Press; 2001.

2. Pauker S, Kassirer J. The threshold approach to clinical decision making. N Engl J Med. 1980;302:1109–17.

3. Doubilet P, Begg CB, Weinstein MC, Braun P, McNeil BJ. Probabilistic sensitivity analysis using Monte Carlo simulation: a practical approach. Med Decis Making. 1985;5:157–77.

4. Gold M, Siegel J, Russell L, Weinstein M. Cost-Effectiveness in Health and Medicine. Cambridge (UK): Cambridge University Press; 1996.

5. Coyle D, Oakley J. Estimating the expected value of partial perfect information: a review of methods. Eur J Health Econ. 2008;9:251–9.

6. Briggs A, Fenwick E, Karnon J, Paltiel AD, Sculpher M, Weinstein M. Model parameter estimation and uncertainty analysis: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group-6. Med Decis Making. 2012;32:722–32.

7. Blanning RW. The sources and uses of sensitivity information. Interfaces. 1974;4:32–8.

8. Kleijnen JPC. A comment on Blanning's ''Metamodel for sensitivity analysis: The regression metamodel in simulation.'' Interfaces. 1975;5:21–3.

9. Stinnett AA, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. Med Decis Making. 1998;18:S68–80.

10. Kleijnen JPC. An overview of the design and analysis of simulation experiments for sensitivity analysis. Eur J Oper Res. 2005; 164:287–300.

11. Duintjer Tebbens RJ, Thompson KM, Hunink MGM, et al. Uncertainty and sensitivity analyses of a dynamic economic evaluation model for vaccination programs. Med Decis Making. 2012; 28:182–200.

12. Friedman LW, Pressman I. The metamodel in simulation analysis: can it be trusted? J Oper Res Soc. 1988;39:939–48.

13. Kleijnen JPC. Design and Analysis of Simulation Experiments. New York: Springer Science + Business Media; 2008.

14. Merz JF, Small MJ. Measuring decision sensitivity: a combined Monte Carlo-logistic regression approach. Med Decis Making. 1992;12:189–96.

15. Tappenden P, Eggington S, Oakley J, Mccabe C. Methods for expected value of information analysis in complex health economic models: developments on the health economics of interferon-β and glatiramer acetate for multiple sclerosis. Health Technol Assess. 2004;8:1–78.

16. Stevenson MD, Oakley J, Chilcott JB. Gaussian process modeling in conjunction with individual patient simulation modeling: a case study describing the calculation of cost-effectiveness ratios for the treatment of established osteoporosis. Med Decis Making. 2004;24:89–100.

17. Coyle D, Buxton MJ, O'Brien BJ. Measures of importance for economic analysis based on decision modeling. J Clin Epidemiol. 2003;56:989–97.

18. Friedman LW, Friedman HH. Validating the simulation metamodel: some practical approaches. Simulation. 1985;45:144–6.

19. Kleijnen JPC, Sargent RG. A methodology for fitting and validating metamodels in simulation. Eur J Oper Res. 1997;120:14–29.