Uncertainty and Patient Heterogeneity in Medical Decision Models

Bas Groot Koerkamp, MD, MSc, Milton C. Weinstein, PhD, Theo Stijnen, PhD, M. H. Heijenbrok-Kal, PhD, M. G. Myriam Hunink, MD, PhD

Parameter uncertainty, patient heterogeneity, and stochastic uncertainty of outcomes are increasingly important concepts in medical decision models. The purpose of this study is to demonstrate the various methods to analyze uncertainty and patient heterogeneity in a decision model. The authors distinguish various purposes of medical decision modeling, serving various stakeholders. Differences and analogies between the analyses are pointed out, as well as practical issues. The analyses are demonstrated with an example comparing imaging tests for patients with chest pain. For complicated analyses step-by-step algorithms are provided. The focus is on Monte Carlo simulation and value of information analysis. Increasing model complexity is a major challenge for probabilistic sensitivity analysis and value of information analysis. The authors discuss nested analyses that are required in patient-level models, and in nonlinear models for analyses of partial value of information analysis. Key words: uncertainty; patient heterogeneity; decision making; Markov models; Monte Carlo method; probabilistic sensitivity analysis; value of information analysis. (Med Decis Making 2010;30:194–205)

Uncertainty and patient heterogeneity are receiving increasing attention in medical decision modeling. The National Institute of Health and Clinical Excellence (NICE) in the United Kingdom has advocated the use of probabilistic sensitivity analysis (PSA) to assess parameter uncertainty. At the same time, decision models have become increasingly complex, often requiring patient-level simulation, which introduces stochastic uncertainty. Uncertainty about the model structure further complicates decision modeling. Finally, patient heterogeneity is relevant, for example, to identify subgroups for which a new intervention is cost effective.

The objective of this study is to demonstrate the various methods to analyze uncertainty and patient heterogeneity in a decision model. We illustrate these methods with an example comparing imaging tests for patients with chest pain. Tutorials on uncertainty or patient heterogeneity in decision modeling have typically focused on a single methodology: for example, PSA, or the expected value of perfect information. A broader view avoids confusion about similar methods, such as first-order and second-order Monte Carlo simulation. Moreover, appreciating value of information analysis, for example, is facilitated when its relation with other methods addressing parameter uncertainty is understood. Our focus is on Monte Carlo simulation and value of information analysis. The scope of this study also includes various nested simulations to accommodate multiple levels of uncertainty and patient heterogeneity.
We will distinguish various purposes of medical decision modeling, serving various stakeholders. Before the advent of cost-effectiveness analysis, the patient and his doctor were the main stakeholders of medical decision modeling. Initially decision models were developed to improve the care of specific patients. Nowadays, many models in health care consider costs in addition to health effects, and policy makers have become stakeholders too. NICE uses the cost-effectiveness outcomes of typically a base case cohort analysis to inform reimbursement decisions. We will demonstrate how analyses, other than the base case cohort analysis, using the same decision model, can serve additional purposes or stakeholders.

In the next section we give a brief description of the decision problem, including a deterministic analysis, used as the primary example for the methods presented in this study. In the sections that follow, we consider stochastic uncertainty, parameter uncertainty, model uncertainty, and patient heterogeneity. We provide verbal explanations of the analyses and present equations in the Web appendix. For complicated analyses, step-by-step algorithms are provided in boxes. We assume familiarity with Markov cohort analysis and probability distributions.

REAL LIFE EXAMPLE

Decision Problem: Diagnostic Strategies for Patients with Chest Pain

Conventional catheter coronary angiography (CA) is considered the reference standard test that can distinguish patients with coronary heart disease (CHD) from patients without CHD. Unfortunately, CA has considerable drawbacks: it has a risk of mortality and morbidity, and it is expensive. Multidetector computed tomographic angiography (CTA) is less expensive and has a minimal risk. Its test characteristics (sensitivity and specificity), however, are imperfect: CTA misclassifies both patients with CHD and those without CHD. The initial risk and cost of CA v. the harm and cost of misclassifying patients with chest pain is the main trade-off when choosing between these imaging tests.

Each year 400,000 patients in the United States newly present with chest pain that may be caused by CHD. It is important to identify patients with CHD, inasmuch as they can benefit from a coronary arterial bypass graft (CABG) or a percutaneous coronary intervention (PCI). The current decision problem is to find the optimal imaging test to diagnose CHD in patients with chest pain.

The available evidence regarding the costs and effects of imaging tests for diagnosing CHD was synthesized into a Markov model from the health care system perspective. The model extrapolated the evidence on costs and effects over the entire remaining lifetime of patients. Although earlier versions of the model compared various imaging strategies, for illustrative purposes we will now only compare CA with CTA. For comparison we will consider a third strategy, in which patients with chest pain receive medical therapy without an imaging test.

Decision Model

In the model we assume that if a CTA test result was positive or uninterpretable, a CA followed. Patients with a positive CA receive a PCI for 1- or 2-vessel disease and a CABG for 3-vessel disease and left main disease. Both treatments have an associated disutility and a short-term risk of mortality and myocardial infarction. The beneficial effects of treatment are 3-fold: reduction in long-term mortality, reduction in long-term risk of myocardial infarction, and reduction of chest pain severity. Age- and gender-specific life tables were used to model the subsequent lifetime outcomes using Markov models. Three chest pain states were distinguished in the model: no, mild, and severe chest pain. Moreover, each year patients may suffer a myocardial infarction or undergo a CABG or PCI, depending on the extent of the CHD and treatment history. We modeled the cost of tests and treatments, as well as the annual cost depending on chest pain severity and left ventricular ejection fraction—for patients with CHD. The model outcomes of each imaging test were quality-adjusted life expectancy (QALE), expressed in quality-adjusted life years (QALYs), and expected lifetime costs, expressed in US dollars. We applied a half-cycle correction for all analyses. Tables 1 to 3 in the Web appendix present all model parameters with their 95% uncertainty interval and sources.

Deterministic Analysis

We performed a Markov cohort analysis for 55-year-old men (representing men aged 50 to 59) with atypical chest pain. Table 1 presents the results of this deterministic analysis, including incremental cost-effectiveness ratios (ICERs). Policy makers can conclude that for a willingness to pay (WTP) of less
than $31,000/QALY, both imaging tests are not cost effective. CA is cost effective if the WTP is $85,000/QALY or more. In between these threshold values, CTA is cost effective.

### STOCHASTIC UNCERTAINTY

Stochastic uncertainty—also known as first-order uncertainty or individual patient variability—represents the uncertainty in patient-level outcomes.\(^{17}\) This uncertainty is entirely due to chance.\(^{18}\) In decision models each chance node contributes to this uncertainty. For example, at a specific chance node a patient has an estimated probability of 3.2% that he will die from surgery. Stochastic uncertainty reflects the uncertainty related to the actual outcome—a patient may or may not fall within the 3.2% of patients that die—which should be distinguished from uncertainty around the 3.2% because of the limited sample of patients in which the value was estimated (i.e., parameter uncertainty) and from uncertainty about whether the 3.2% applies to this particular type of patient (i.e., patient heterogeneity).

We evaluated stochastic uncertainty using first-order Monte Carlo analysis (also known as microsimulation).\(^{10,11}\) This should not be confused with second-order Monte-Carlo analysis, dealing with parameter uncertainty (see below). A first-order Monte Carlo analysis simulates subjects one by one. Probabilities at chance nodes and a random number generator result in a subject’s path along the chance nodes. This path is called a random walk or a “trial.” Counters (also known as tracker variables) can record the accumulated (quality-adjusted) lifetime and costs, as well as events along the subject’s path. When the subject dies, the simulation restarts with a new subject. We performed 10,000 random walks in the example model. Using the results of this analysis we can calculate, for example, that the probability that a 55-year-old man with atypical chest pain lives at least another 10 years is 82%. Patients are typically interested in such outcomes in addition to expected outcomes.

The most common rationale for the analysis of stochastic uncertainty is the use of patient-level models, in which a first-order Monte Carlo simulation (i.e., microsimulation) is necessary to estimate expected outcomes.\(^{19}\) In so-called patient-level models, a first-order Monte Carlo simulation (i.e., microsimulation) is necessary to estimate expected outcomes.\(^{20}\) First-order Monte Carlo simulation allows modeling of the influence of patient history on subsequent events. For example, when a subject suffers a myocardial infarction, his future mortality rate will increase. Markov cohort models can model the influence of patient history only by including it in the definition in health states, which can result in an unwieldy number of states.\(^{11}\)

If a stakeholder is uninterested in individual patient outcomes, then the analysis of stochastic uncertainty in patient-level models only contributes noise to the expected outcomes. Instead of a single Markov cohort analysis, many trials are required to obtain a precise estimate of the expected outcomes. Evaluating more trials of the model will improve precision but also requires more computing time. The precision of these expected outcomes could be assessed by the standard error of the mean (SEM). The SEM is calculated as the standard deviation of the trial outcomes divided by the square root of the number of trials. Gaussian process modeling has been suggested as a time-efficient alternative to first-order Monte Carlo simulation,\(^{21}\) but such methodology is still in the developmental stage. Griffin and others\(^{19}\) discussed how patient-level modeling can be avoided in certain circumstances.

### PARAMETER UNCERTAINTY

#### Deterministic Sensitivity Analysis

Deterministic sensitivity analysis evaluates the influence of uncertainty in 1 or more parameters on the expected outcomes. In univariable (1-way) sensitivity analysis, the outcome of each strategy is calculated over a justifiable range of 1 parameter, for example, the 95% uncertainty interval based on the results of a study. A tornado diagram presents the results of many 1-way sensitivity analyses. In a tornado diagram a horizontal bar represents the range of expected outcomes at the decision node (across

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Cost</th>
<th>Incr Cost</th>
<th>Effect</th>
<th>Incr Effect</th>
<th>ICER</th>
</tr>
</thead>
<tbody>
<tr>
<td>No imaging test</td>
<td>26953</td>
<td>12.96567</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTA</td>
<td>34997</td>
<td>8044</td>
<td>13.22813</td>
<td>0.262</td>
<td>30649</td>
</tr>
<tr>
<td>CA</td>
<td>35153</td>
<td>156</td>
<td>13.22997</td>
<td>0.002</td>
<td>84836</td>
</tr>
</tbody>
</table>

Note: All costs are in US$, all effects in QALYs. Incr, incremental; ICER, incremental cost-effectiveness ratio; CTA, multidetector computed tomographic angiography; CA, conventional catheter coronary angiography.
all strategies) given the range of each selected estimated parameter (e.g., the 95% uncertainty interval). The tornado shape arises by ordering the bars by width, starting with the widest at the top. A mark can indicate where the optimal strategy changes across the range of a parameter.

We built a tornado diagram of all estimated parameters, with the exception of parameters that are correlated, such as Dirichlet distributions. We used net monetary benefit (NMB) as the expected outcome, combining the outcomes cost and effect: 

\[
\text{NMB} = \text{effect} \times \text{WTP} - \text{cost}. 
\]

Figure 1A presents a tornado diagram for the 10 estimated parameters with the largest impact on the expected outcome (i.e., NMB) at the decision node, given a WTP of $50,000/QALY. The black mark in the top bar demonstrates that only this parameter—the utility of non-specific chest pain—causes a change in optimal strategy at the lower end of its range. Three other parameters changed the optimal strategy within their range. These are presented in the tornado diagram of Figure 1B.

**Probabilistic Sensitivity Analysis**

PSA, or second-order Monte Carlo analysis, evaluates the joint effect of uncertainty about all estimated parameter values in the model. The uncertainty about the parameter values is represented by probability distributions and propagated in the model resulting in a probability distribution of the expected outcome for each strategy. The probability distributions of the parameters are often obtained using traditional parametric statistical methods. Briggs and others give guidance on selecting the appropriate distributions. Alternatively, bootstrapping has the advantage that the analyst does not have to make parametrical assumptions about the parameter distribution. However, standard bootstrapping methods may lead to misleading inferences, for example, when cost data exhibit highly skewed distributions. Correlations between parameters can be modeled by drawing values from joint distributions of the correlated parameters. Bayesian Markov chain Monte Carlo methods are increasingly used to model correlations, but require dedicated software such as WinBUGS.

We performed PSA by randomly drawing a value for each parameter from its probability distribution. This set of values is commonly referred to as a “sample.” The model was then recalculated for this sample using a Markov cohort analysis. We repeated this for 10,000 samples. Table 2 (columns 1–4) presents 10 samples of the PSA. In column 5, for each sample we identified the optimal alternative as the strategy with the highest net benefit. The probability that CTA is the “true” optimal strategy is the percentage of samples in which it has the highest net benefit. Table 2 shows that CTA is optimal in 50% of the 10 samples. Of 10,000 samples, CTA was optimal in 54%.

Acceptability curves (Figure 2) demonstrate that the probability that CTA is cost effective is less than 55% for any reasonable value of the WTP. Note also that the probability that CA is cost effective is smaller than the probability that CTA is cost...
effective for any WTP, even when the expected net benefit of CA exceeds the expected net benefit of CTA (i.e., for WTP > $85,000/QALY; see Table 1). The cost-effectiveness acceptability frontier (CEAF) has been introduced within acceptability curves to indicate the intervention with the highest expected net benefit.36

Next, we estimated the 95% uncertainty interval for the incremental net benefit of CTA v. CA from the results of the PSA. For each sample we calculated the difference in net benefit between the two strategies. The 95% uncertainty interval for the incremental net benefit of CTA v. CA is –$430 to +$516, with an expected difference of $64. The equivalent values in quality-adjusted life-days are –3 to +4 days, with an expected difference of half a day. These values are small, but typical for incremental benefits of diagnostic tests and screening programs.

The use of PSA also has a more technical justification. A deterministic analysis is valid only if the expected outcome of the model (f) equals the model outcome when evaluated in the expected values of the parameters (x): E[f(x)] = f(E[x]). Correlated parameters in which the model is linear do not require sampling from their probability distribution to obtain unbiased expected outcomes. A linear function of parameters x₁, x₂, ..., xₙ is defined as a function of the form a₁ * x₁ + a₂ * x₂ + ... + aₙ * xₙ for certain constants a₁, a₂, ..., aₙ. Uncorrelated parameters in which the model is multilinear also do not require sampling from their probability distribution to obtain unbiased expected outcomes. A function is multilinear if it is a linear function of each parameter when the other parameters are given fixed values. Nonlinearity is the rule rather than exception: in Markov models the transition probabilities are multiplied by themselves repeatedly in the calculation of expected outcomes. The same issue arises in the section on partial expected value of perfection information (pEVPI) when we estimate the pEVPI. In Web Appendix A we present a numerical example. Although the expected outcomes of

<table>
<thead>
<tr>
<th>Sample</th>
<th>NMB (CTA)</th>
<th>NMB (no test)</th>
<th>NMB (CA)</th>
<th>Sample Best</th>
<th>Sample Max</th>
<th>Baseline Max</th>
<th>Opportunity Loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$598,104</td>
<td>$588,775</td>
<td>$597,540</td>
<td>CTA</td>
<td>$598,104</td>
<td>$598,104</td>
<td>$0</td>
</tr>
<tr>
<td>2</td>
<td>$608,478</td>
<td>$606,464</td>
<td>$608,743</td>
<td>CA</td>
<td>$608,743</td>
<td>$608,478</td>
<td>$265</td>
</tr>
<tr>
<td>3</td>
<td>$609,489</td>
<td>$606,648</td>
<td>$602,758</td>
<td>CTA</td>
<td>$609,489</td>
<td>$609,489</td>
<td>$0</td>
</tr>
<tr>
<td>4</td>
<td>$624,283</td>
<td>$624,815</td>
<td>$624,614</td>
<td>No test</td>
<td>$624,185</td>
<td>$624,283</td>
<td>$532</td>
</tr>
<tr>
<td>5</td>
<td>$636,335</td>
<td>$626,805</td>
<td>$633,813</td>
<td>CTA</td>
<td>$636,035</td>
<td>$636,035</td>
<td>$0</td>
</tr>
<tr>
<td>6</td>
<td>$635,034</td>
<td>$625,744</td>
<td>$636,382</td>
<td>CA</td>
<td>$636,382</td>
<td>$635,034</td>
<td>$1348</td>
</tr>
<tr>
<td>7</td>
<td>$638,400</td>
<td>$630,043</td>
<td>$637,928</td>
<td>CTA</td>
<td>$638,400</td>
<td>$638,400</td>
<td>$0</td>
</tr>
<tr>
<td>8</td>
<td>$637,761</td>
<td>$628,196</td>
<td>$632,105</td>
<td>CTA</td>
<td>$637,761</td>
<td>$637,761</td>
<td>$0</td>
</tr>
<tr>
<td>9</td>
<td>$655,188</td>
<td>$640,486</td>
<td>$655,815</td>
<td>CA</td>
<td>$655,815</td>
<td>$655,188</td>
<td>$627</td>
</tr>
<tr>
<td>10</td>
<td>$622,801</td>
<td>$617,226</td>
<td>$623,182</td>
<td>CA</td>
<td>$623,182</td>
<td>$622,801</td>
<td>$381</td>
</tr>
<tr>
<td>Average</td>
<td>$626,557</td>
<td>$619,520</td>
<td>$625,288</td>
<td>50% CTA</td>
<td>$626,873</td>
<td>$626,557</td>
<td>Total EVPI = $315</td>
</tr>
</tbody>
</table>

Table 2  Results of the Probabilistic Sensitivity Analysis (Columns 1–5) and Expected Value of Information Analysis (Columns 1–8)

Note: NMB, net monetary benefit; WTP, willingness to pay = $50,000/QALY; CTA, multidetector computed tomographic angiography; No test, no imaging test; CA, conventional catheter coronary angiography; Sample Best, the strategy with the highest net benefit of the sample; Baseline Max, the outcome of the strategy with the overall optimal outcome; Sample Max, the outcome of the strategy with the highest net benefit of the sample; Total EVPI, the expected benefit per patient of a hypothetical study with an infinite sample size that would eliminate all parameter uncertainty.

Figure 2  Cost-effectiveness acceptability curves. The probability is presented that each strategy is the “true” optimal strategy across a range of values for the willingness to pay. For each value of the willingness to pay the total probability adds to 1. CTA, computed tomographic angiography; No test, no imaging test; CA, coronary angiography.
the Markov cohort analysis in the section describing the real life example are biased in theory, we could not detect this bias: a PSA of 100,000 samples found expected outcomes similar to the deterministic analyses.

A PSA becomes more complicated in patient-level models (see the section on stochastic uncertainty) that do not allow for Markov cohort analyses. Instead, the recalculation for each sample requires an entire first-order Monte Carlo simulation (e.g., 1000 trials). The PSA represents an outer loop of \(M\) samples; the first-order Monte Carlo simulation represents an inner loop of \(N\) trials. This is sometimes called the “\(M\) by \(N\) problem” and it is a major obstacle to performing PSA because of the time-consuming calculations required.\(^{19,20,37,38}\) Confusion often exists about performing a PSA of a patient-level model. Box 1 presents a step-by-step algorithm.

**Total EVPI**

Information obtained in future quantitative research—for example, a randomized controlled trial—can reduce parameter uncertainty. A decrease in parameter uncertainty may avoid reimbursement of suboptimal interventions, and consequently is expected to benefit patients and/or reduce costs. Value of information analyses explicitly estimate the expected benefit of collecting information in future research. The value of information is not the actual value of future research—which we will only learn after performing future research—but the “expected” value of future research. It is expressed in the same units as the model outcome, typically net monetary benefit. Value of information analysis was introduced by Grundy and others\(^{39}\) in the late fifties and developed by Raiffa and Schlaifer\(^{40,41}\). Since the late eighties it has received increasing attention in the risk analysis literature and more recently in health care\(^{2,8,42,43}\).

The total EVPI is the expected benefit per patient of a hypothetical study with an infinite sample size that would eliminate all parameter uncertainty. It is estimated by the average opportunity loss of the samples of the PSA (Table 2).\(^{44}\) The opportunity loss of a sample is defined as the difference between the maximum expected benefit of that sample (sample max) and the sample’s expected benefit of the baseline optimal strategy (baseline max).\(^{45}\) For example, the maximum expected benefit of the second sample in Table 2 is the expected benefit of CA: $608,743. CTA is the baseline optimal strategy; the expected benefit of CTA in sample 2 is $608,478. The opportunity loss of sample 2 is the difference between these values: $608,743 – $608,478 = $265. The final column of Table 2 presents the opportunity loss for each sample. The average opportunity loss of all samples is $315 per patient and is an estimate of the total EVPI per patient. See Box 2 for a step-by-step algorithm. Using 10,000 samples instead of the 10 samples of Table 2, we obtained a more precise estimate of the total EVPI: $294 per patient. This means that after eliminating uncertainty we can expect an improvement in net monetary benefit of $294 per patient. The probability that the actual value is zero is 54%, identical to the current probability that CTA is the optimal strategy. This result also implies that the current expected harm due to uncertainty is $294 per patient, with a health equivalent of 2 quality-adjusted life-days (WTP = $50,000/QALY).

More research to decrease uncertainty is justified if the expected benefit to future patients exceeds the cost of research. The population EVPI represents the

---

**Box 1**

**Second-Order Monte Carlo Simulation (Probabilistic Sensitivity Analysis) in Patient-Level Model**

1. Draw a sample from all parameter distributions.
2. Plug in the sampled values and perform a first-order Monte Carlo simulation with \(N\) trials.
3. Calculate the expected outcome of the \(N\) trials for each strategy.
4. Repeat steps 1 to 3 \(M\) times.
5. The \(M\) expected outcomes obtained at step 3 represent the distribution of the expected outcome characterizing parameter uncertainty.

**Box 2**

**Total Expected Value of Perfect Information (EVPI)**

1. Perform a second-order Monte Carlo simulation, sampling all estimated parameters.
2. For each sample of the probabilistic sensitivity analysis, calculate the opportunity loss as the difference between the maximum expected benefit of the sample and the sample’s expected benefit of the baseline optimal strategy.
3. The total EVPI is the average opportunity loss.
expected benefit to all future patients. It is estimated as the product of the total EVPI per patient and the population that is expected to benefit from future research, discounting expected benefit in future years. The population EVPI is a ceiling level for the expected return on investment of research. If research is more expensive than the population EVPI, it is a bad investment: the uncertainty is not important enough to be resolved. The annual population to benefit is usually ambiguous, because it is not obvious whether we should consider the local setting, one country, or all patients worldwide. Moreover, the period that patients will benefit from the proposed data collection is typically uncertain because of future improvements, novel interventions, or new insights. These ambiguities, however, are not drawbacks of value of information analysis in itself, but inherent to the problem of allocating resources wisely.

We estimated the annual population to benefit (males aged 50–59 years with atypical chest pain) for the United States at 44,000 patients (i.e., 11% of 400,000). Assuming a period of 5 years and a discount rate of 3%, we found a total EVPI for the population to benefit of $61 million.

Partial EVPI

Instead of estimating the total EVPI of all parameters, we can estimate the partial EVPI of 1 or more parameters (i.e., the parameters of interest). First, we demonstrate the so-called 1-level algorithm. Analogous to the estimation of total EVPI, a second-order Monte Carlo simulation is performed, but only the parameters of interest are sampled from their distributions. The parameters not of interest (i.e., all other parameters) remain fixed at their mean values. For each sample the model is recalculated and the opportunity loss is calculated analogous to the algorithm of total EVPI. The average opportunity loss of many samples is an estimate of the partial EVPI of the parameters of interest. See Box 3 for a step-by-step algorithm of the 1-level algorithm.

Keeping the parameters not of interest fixed at their mean value may result in the same bias that we discussed in the section on probabilistic sensitivity analysis (technical justification of PSA). In a 2-level algorithm, a second-order Monte Carlo simulation is performed for each sample of the parameters of interest, to avoid this bias. All parameters of interest are sampled in the outer loop second-order Monte Carlo simulation; selected parameters not of interest are sampled in the inner loop second-order Monte Carlo simulation. The criteria to decide if a parameter not of interest requires sampling in the inner loop are the same as the criteria of the section on probabilistic sensitivity analysis, to avoid a bias when obtaining expected outcomes of a model. If correlations exist between parameters of interest and a parameter not of interest, the latter should also be sampled in the inner loop when an analytic expression for the conditional mean value is not available. See Box 4 for a step-by-step algorithm of the 2-level algorithm.

The outer loop—sampling the parameters of interest—determines the precision of the estimate of the partial EVPI. More samples in the inner loop—recalculating the model by sampling the parameters not of interest—yields less biased results. Brennan and others recommend a 1 to 5 ratio of samples of the outer v. the inner loop. The correct order of magnitude was found with a minimum of 100 samples in the outer loop.

In a tornado diagram we identified 4 parameters that have the capability to change the optimal strategy somewhere along their range of likely values. We first estimated the partial EVPI of these parameters because they should exceed zero. We divided the remaining parameters into several groups. Most groups have a zero partial EVPI, and therefore all constituent parameters have a zero partial EVPI. The tornado diagram identified 4 of 5 parameters with a nonzero partial EVPI. The ranking in the tornado diagram, however, is not a good predictor of the importance of uncertainty as reflected by the partial EVPI. Table 3 presents the ranked
results for partial EVPIs with their rankings in the tornado diagram. Note that the sum of the partial EVPIs of individual parameters generally does not equal the total EVPI.45,46

To assess whether more research is justified we should estimate the partial EVPI of a set of parameters that could be measured in a specific study. Parameters with a partial EVPI of zero may seem useless to consider in future research. A combination of such parameters, however, may jointly have a non-zero partial EVPI. We analyzed the partial EVPI of 8 study designs. Six of these were observational studies, measuring, respectively: test characteristics of CTA, complications of coronary angiography, utilities of chest pain states, costs of interventions, complications of PCI, and complications of CABG. The other two study designs were randomized trials: medical treatment v. PCI in 1- or 2-vessel disease and medical treatment v. CABG in 3-vessel disease or left main disease. For the observational study measuring utilities we found the highest partial EVPI: $91 per patient, with a population EVPI of $19 million. The cost study had a partial EVPI of $48 per patient, and the diagnostic study for test characteristics of CTA had a partial EVPI of $31 per patient. All other study designs, including the two randomized controlled trials had a (near) zero partial EVPI.

The (partial) expected value of sample information (EVSI) estimates the expected value of obtaining information for finite sample sizes.47–50 With increasing sample size, the partial EVSI will reach a ceiling: the partial EVPI, representing an infinite sample size. At the same time, the cost of research increases with increasing sample size. The expected net benefit of sampling (ENBS) is defined as the difference between the EVSI and the cost of research. The maximum ENBS is associated with the optimal sample size of a proposed study design. We refer to Ades and others for an extensive coverage of EVSI.

Table 3 Individual Model Parameters Ranked by Partial Expected Value of Perfect Information (pEVPI) as a Measure for the Importance of Uncertainty

<table>
<thead>
<tr>
<th>Parameter</th>
<th>pEPVI</th>
<th>Tornado Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utility of nonspecific chest pain</td>
<td>42</td>
<td>1</td>
</tr>
<tr>
<td>Cost CA</td>
<td>32</td>
<td>15</td>
</tr>
<tr>
<td>Cost CTA</td>
<td>25</td>
<td>34</td>
</tr>
<tr>
<td>TPR CTA</td>
<td>14</td>
<td>39</td>
</tr>
<tr>
<td>Utility of mild chest pain</td>
<td>0.001</td>
<td>2</td>
</tr>
<tr>
<td>Annual costs of mild chest pain pat with normal LVEF</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Cost CABG</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>Cost PCI</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>Reduction in mortality after CABG in LMD</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>RR of dying with 1 or 2 VD (compared with no CAD)</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>RR of dying with 3 VD (compared with no CAD)</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>RR of dying with LMD (compared with no CAD)</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Utility of severe chest pain</td>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>

Note: The Tornado rank in the third column demonstrates that it is not a valid proxy for the importance of uncertainty. All parameters not in this table had a zero pEVPI. CA, coronary angiography; CTA, computed tomographic angiography; TPR, true positive rate; LVEF, left ventricular ejection fraction; CABG, coronary arterial bypass graft; PCI, percutaneous coronary intervention; LMD, left main disease; RR, relative risk; VD, vessel disease; CAD, coronary artery disease.

Box 4

Partial Expected Value of Perfect Information (EVPI)—2-Level Algorithm

1. Draw a sample from the distributions of the parameters of interest.
2. Draw a sample from the distributions of the following parameters not of interest:
   - Correlated parameters not of interest in which the model is not linear
   - Uncorrelated parameters not of interest in which the model is not multilinear
   - Parameters not of interest correlated with parameters of interest, for which an analytic expression for the conditional mean value is not available
All other parameters not of interest may be fixed at their mean values.
3. Recalculate the model performing a cohort analysis for the values of steps 1 and 2.
4. Repeat step 2 and 3 K times—inner loop.
   Calculate the expected benefit of each strategy.
5. Calculate the opportunity loss as the difference between the maximum expected benefit in step 4 and the expected benefit in step 4 of the baseline optimal strategy.
6. Repeat step 1 to 5 J times—outer loop.
7. The partial EVPI is the average opportunity loss.
MODEL STRUCTURE UNCERTAINTY

The structure of a model typically depends on assumptions about the natural course of a disease and how medical interventions may influence this course. The amount and form of available data will also determine the model structure. Moreover, analysts can opt for more or less model complexity: for example, a simple decision tree v. a Markov model. This may depend on the research question, available time, and the required validity and precision of the results. Consequently, different teams of experts may come up with different models to represent the same decision problem. Campbell and others demonstrated how this can result in large variation in the outcomes.\textsuperscript{51} Evaluating only the plausible single best model may result in an underestimation of uncertainty.\textsuperscript{5} Ideally, analysts should build a model for every imaginable set of assumptions regarding a decision. In real life most analysts settle for using their single best model. Claxton and others pointed out that more research is needed regarding the trade-off between the realism of the model and the available time.\textsuperscript{37} If alternative modeling assumptions may affect the decision, more evidence to justify the use of one assumption instead of the others should be found. If this is impossible, the results of each model can be presented, and the policy maker can decide. Sometimes model structure uncertainty can be dealt with using parameterization.\textsuperscript{52} Finally, some sources of uncertainty are not covered by parameter or structural uncertainty: for example, uncertainty about the appropriate evidence sources, and uncertainty about the selection of interventions.

We performed a structural sensitivity analysis by assuming an additive mortality function. The CTA strategy remained the optimal strategy at a WTP of \$50,000/QALY, with an ICER of \$28,000 compared with no imaging test. The ICER differed about 10\% from the ICER using the multiplicative model.

PATIENT HETEROGENEITY

Patient heterogeneity is usually analyzed to identify differences in the optimal strategy for subgroups of patients.\textsuperscript{53} Clinical guidelines and reimbursement decisions reflect these differences between subgroups. Moreover, a strategy could have a high incremental cost-effectiveness ratio for the total population, but a low ratio for a certain subgroup; the mean value may obscure the cost effectiveness of a strategy for a subgroup. Baseline patient characteristics can influence each estimated parameter in the model: for example, we can distinguish heterogeneity in treatment effects, heterogeneity in costs, and heterogeneity in utilities. In practice it is often difficult to determine whether a difference between subgroups is genuine or simply reflects noise in the data. Criteria are being developed to decide when it is justified to model heterogeneity in a parameter.\textsuperscript{54}

Sensitivity analysis may evaluate the optimal strategy for various subgroups and involves repeated analysis of a model for, for example, different age groups. If patient heterogeneity is modeled using a continuous or ordinal variable, sensitivity analysis can calculate the expected outcomes over a range of values for the patient characteristic, analogous to deterministic sensitivity analysis of parameters representing parameter uncertainty.

Differences in setting—as opposed to differences in patient characteristics—can also cause heterogeneity in parameter values. This type of heterogeneity arises, for example, when a model developed based on data from one country is used to make inferences for another country. For example, the cost of an appendectomy or the sensitivity of ultrasound for appendicitis can differ across countries or hospitals. A policy maker should be able to assess what the model implies for a situation that may be identical to the base case assumptions of the model except for a limited set of parameter values.

In our example model, gender, age, and type of chest pain are the most relevant patient characteristics. We assessed 30 subgroups: 5 age groups, both genders, and 3 chest pain groups. The higher the prior probability of CHD, the more likely it is that CA is cost effective. The prior probability of CHD is increased by advanced age, male gender, and typical instead of nonspecific or atypical chest pain. CTA is cost effective for intermediate-risk patients, no imaging for low-risk patients.

The analysis of patient heterogeneity is also required when uniform decisions are considered for rather heterogeneous populations. For example, policy makers may want to know the overall ICER of a colon cancer screening program for everyone older than 50 years. For population-level decisions it is important that the heterogeneity of the target population is reflected in the model. Patient heterogeneity can be represented by distributions for each patient characteristic or by bootstrapping “real” subjects from a study population. The expected outcome for the heterogeneous population is the average outcome of many randomly drawn patients performing a random walk in the model. Typically
this requires a patient-level first-order Monte Carlo simulation. Nijhuis and others used this approach to model a heterogeneous population using data from a large study population. Parameter uncertainty in such models adds an additional level of complexity.

DISCUSSION

We demonstrated various methods to analyze uncertainty and patient heterogeneity in decision models. The analyses resulted in outcomes serving various purposes and stakeholders. Policy makers (e.g., NICE) determine the optimal strategy by combining these results with (typically) unmodeled considerations such as ethical viewpoints or the transition costs of a new intervention. More recently, research-funding agencies can use the results of value of information analyses to guide future research. Patients and doctors also combine model results with their preferences regarding health states and risk attitude, as well as specific patient characteristics that are often not accounted for in the model.

The analysis of uncertainty and patient heterogeneity faces several challenges. Increasing model complexity impedes PSA and value of information analysis. Nested analyses are required in patient-level models and for partial value of information analyses in nonlinear models. The required calculation time in decision analytic software using personal computers is prohibitive. Linear cohort models (i.e., no Markov nodes, no patient-level simulation) avoid nested simulations, but could be an unrealistic representation of the decision problem. Accounting for correlations between estimated parameters is another challenge. Correlations are influential on the value of information, but often no data are available to model this. Third, it is not feasible to consider the characteristics and preferences of each individual patient in a decision model. However, doctors rarely see patients that match the base case analysis of a cost-effectiveness study. Instead, doctors care for a variety of patients each with their own unique set of risks and preferences. Finally, model structure uncertainty remains problematic, because more than one structure may be reasonable. We often lack time to build various models, and placing a weight on each model is an arbitrary choice.

Despite these challenges, decisions regarding current implementation and future research must be made. We hope this study will stimulate and help analysts and policy makers evaluating uncertainty and patient heterogeneity to inform these decisions.

ACKNOWLEDGMENTS

We are grateful to a nonrandom draw of our target audience: Johanna Bosch, Linda Everse, Ylian Liem, Rogier Nijhuis, Uwe Siebert, and Karen Visser, for their valuable comments on earlier versions of this manuscript.

REFERENCES


