Statistic—Advanced process

	Name	Type	Expression	Report label	Output file
1	Statistic 1	Output	vLY(1)	LY SoC	
2	Statistic 2	Output	vQALY(1)	QALY SoC	
3	Statistic 3	Output	vCost(1)	Cost SoC	
4	Statistic 4	Output	vdLY(1)	Disc LY SoC	
5	Statistic 5	Output	vdQALY(1)	Disc QALY SoC	
6	Statistic 6	Output	vdCost(1)	Disc Cost SoC	
7	Statistic 7	Output	vLY(2)	LY MM	
8	Statistic 8	Output	vQALY(2)	QALY MM	
9	Statistic 9	Output	vCost(2)	Cost MM	
10	Statistic 10	Output	vdLY(2)	Disc LY MM	
11	Statistic 11	Output	vdQALY(2)	Disc QALY MM	
12	Statistic 12	Output	vdCost(2)	Disc Cost MM	
13	Statistic 13	Output	(vCost(2)-vCost(1))/(vQALY(2)-vQALY(1))	ICER	

Double-click here to add a new row.

FIGURE 4.41
Values are to be written out to the default Arena® report. Each statistic needs to be provided with a report label to identify it in the report. The calculation of the ICER has also been added so that it is reported as well. ICER, incremental cost-effectiveness ratio.

5

Analyses

An analysis using a discrete event simulation (DES) involves several steps. First, all the input values are read in and stored in their appropriate locations (e.g., global variables). Next, the entities are created and they are introduced into the simulation where they experience the events specified in the model logic, accruing consequences such as costs, survival time, and quality-adjusted life years (QALYs). This happens either for the specified time horizon or for a fixed number of individuals (i.e., until the last entity dies or otherwise leaves the model). The outcomes of interest are recorded and all the results required to inform the health technology assessment (HTA) are transferred to a suitable medium (e.g., a spreadsheet), if necessary. There may be some additional processing of that information required to produce the final results in the form desired by the decision makers (e.g., an incremental cost-effectiveness ratio, ICER) (Rutter et al. 2011).

As a DES incorporates chance at many junctures (e.g., selecting from distributions, applying risk equations, and making decisions), there is stochastic uncertainty, and this must be dealt with in the analyses. In addition, many of the input values are subject to uncertainty because they are estimated from limited data, and this parameter uncertainty must also be considered at the analytic stage. During the design and construction of a typical DES, the modeler also makes many assumptions and decisions about the applicable logic. Since other choices could also be justifiable, there is always structural uncertainty to take into account. Finally, to address the heterogeneity in the population and variations of interest in other inputs (e.g., the time horizon), the modeler should analyze a range of scenarios.

The individual steps that need to be taken in a DES analysis have already been described in Chapter 3. In this chapter, the focus is on the overall process, particularly the additional processing required for an HTA. First, the base case is defined and the analyses required to produce it are detailed. The handling of stochastic parameter and structural uncertainty is described. Finally, the assessment of sensitivity to changes in input values that reflect heterogeneity and other scenario features is addressed.

5.1 Base Case

For HTA, the analysis of a DES typically involves generating a set of relevant outputs (e.g., costs, rate of clinical events, survival, and QALYs) for each assessed technology. These outputs are generally derived by using what are thought to be the most relevant estimates of the true values of all the inputs. This most-relevant analysis is known in HTA as the *base case*.

To produce the base case, the DES collects information on the experience of each entity and applies valuations to that experience (e.g., the costs and the quality-of-life adjustments). Although the experience is individual, the interest, ultimately, is in the aggregate values across the population (e.g., total cost and total QALYs lived). These aggregate values can be obtained by accumulating the pertinent quantities in global variables, separately by technology. At the end of the simulation, these aggregated values are reported and used to compute the differences between interventions, and these are used to derive additional measures like the ICER. Alternatively, the DES may accumulate the values at the individual level in attributes and carry out the aggregation separately at the end of the simulation, or even export the data for aggregation separately in other software. Although this consumes more computer memory and time, it provides for a fuller description of the results as measures of variation can also be derived (e.g., upper and lower quartiles, and standard deviation) as they would be for any dataset containing individual values.

Either way, the specific approach to carrying out the analyses depends on the software used. In a spreadsheet, all of the calculations can be implemented directly using the built-in functions that are typically available in such programs. If a general programming language is used, then subroutines must be developed to carry out the necessary calculations, and these are called when the simulation has concluded. In DES-specific software, there are often built-in functions that can track these outcomes and produce the required statistics. Many analysts, however, prefer to output the results of the simulation to a spreadsheet and carry out the final calculations there. Regardless of the approach taken, it is important to turn off all animation and other calculation-intensive processes not needed for the runs to ensure that the computations are executed as fast as possible.

5.1.1 Dealing with Chance (Stochastic Uncertainty)

Although the inputs in a simulation are common to all the entities, the results will vary across individuals because the pathways taken through the model are determined by the sequences of random numbers sampled for each entity during the simulation. For example, the time of death for one individual is assigned as 6 months into the simulation because the random number drawn for that individual sampled the twentieth percentile of the time-to-death

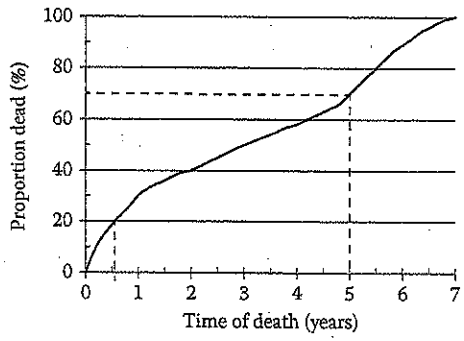


FIGURE 5.1
Example of identical individuals having very different experience in a simulation. The figure shows the cumulative fatality curve for all individuals with the same characteristics and treatment in a simulation. One individual dies at 6 months into the simulation because he was assigned a random number of 0.2 for this event, while another one dies at 5 years was given a random number assignment of 0.7.

cumulative distribution; the sampling for the next individual with identical characteristics yields the seventieth percentile of the distribution, and, therefore, that entity is assigned a time of death of 5 years (Figure 5.1). Across separate analyses of the model, these two individuals who were identical at the start will sample different percentiles of the relevant cumulative distributions each time and, thus, be assigned different times to those events. Hence, even if the analysis is repeated without varying the inputs, the results may still vary since the random numbers used will be different each time (unless steps are taken to force the use of the same set of random numbers) and the number of entities that are simulated is finite. This type of variability due to the play of chance is known as *stochastic uncertainty*.

The set of random numbers used in a particular analysis defines one *replication* of the model (also known as one *trial*). This situation is analogous to that of a randomized controlled trial. Even if the clinical study is repeated with an identical protocol, the second study will produce different results. It is another *replication*—another trial—trying to estimate the same effects. To reduce stochastic uncertainty, a modeler may carry out many replications. For each one, the model is analyzed using the same set of input values, but with different random number sequences for each replication. This is typically accomplished by using different sets for the random number generators in each replication. This leads to the sampling of different values from the input parameter distributions and, consequently, variation in the pathways taken through the model across the replications, despite the use of the same input values. The outputs are reset to zero between replications so that each replication provides a fresh collection of results. By averaging the results across the series of replications (akin to what is done in meta-analysis

across a group of randomized controlled trials), the effect of stochastic variation (i.e., the risk of basing a decision on the results of a single, potentially unrepresentative replication) is reduced. An analysis like this, based on a given set of input values, is considered a *model run*. Since stochastic uncertainty cannot be completely eliminated, no matter how large the model run, it is important to quantify it by expressing the degree of variation observed (Briggs et al. 2012), much as is done when analyzing clinical trials.

In DES that represent the operation of physical facilities, such as an outpatient clinic or emergency department, relevant outputs might include estimates of mean waiting times and measures of the distribution of waiting times to inform the likelihood of patients incurring very long delays (Stahl et al. 2004). The runs for such models are generally defined with respect to the time frame of interest (e.g., one day). In order to remain realistic, each replication is then restricted to the number of individuals presenting at the facility within that time frame. Thus, generating stable estimates of the outputs requires undertaking multiple replications within each run.

By contrast, HTA models that are not addressing specific physical facilities are not required to reflect realistic numbers of patients, nor are they restricted to finite time periods over which individuals can enter the model, and thus, a model run can include as many individuals as are required to achieve stable outputs. Instead of multiple replications, the modeler can increase the number of entities until the results stabilize sufficiently. In the context of clinical trials, a similar decision looms: carry out a single very large trial, or engage in a series of smaller trials and aggregate the results across the series afterward. In the real-life situation of clinical trials, the decision is often driven by regulatory (e.g., a minimum of two separate trials required) and practical considerations. Given there are no limits to the number of entities that can be included in a replication (other than whatever is imposed by the hardware available), a DES model run can comprise a single large model replication.

As, due to the stochastic nature of DES, multiple model runs using the same input values will produce different model outputs, it is desirable to ensure that multiple model runs using the same input values generate reasonably stable output values for the parameters of interest to the decision maker (e.g., mean total cost by intervention and QALYs obtained with each intervention). There is no consensus on the definition of stable model outputs. Stability is observed when the difference in results produced by an additional model replication using the same input values but alternative random number sequences is considered insignificant (Figure 5.2). What constitutes an insignificant difference may vary according to the decision problem being addressed and the analyst making the assessment. A rule of thumb used by many is that differences of less than 1% across model runs are insignificant. Since HTAs are carried out to inform decisions, the threshold for decision making may be used as a guide to the degree of stability required. If the variation does not alter the decision, then the results can be considered sufficiently precise. Although the base case analyses may be

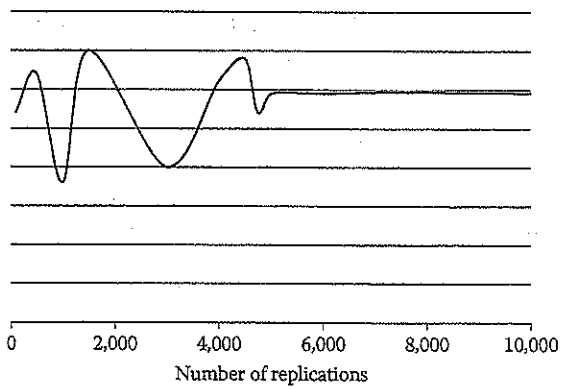


FIGURE 5.2
Control chart used to select the number of replications. This simulation was run for a varying number of replications, and it was found that variation in the value of an outcome of interest became minimal after about 5000 replications.

driven to a high stability (i.e., a low threshold of change), this tolerance may be relaxed when conducting the full uncertainty analysis in order to stay within feasible running times while still reflecting uncertainty.

An alternative approach to attaining sufficiently stable results is to determine the standard error of the model outputs that is considered tolerable and use this to estimate the number of entities to include in a model run. For example, in a published analysis of osteoporosis treatments using a DES (O'Hagan et al. 2007), a mean incremental net benefit of £1,308 was estimated in one model run using the base case input values, with an individual-level variance of about 2.4×10^9. A run size of 15,000 individuals across model runs produced a standard error of 400 (the square root of $2.4 \times 10^9/15{,}000$), and this provided sufficient confidence that the true mean incremental net benefit was positive.

5.1.2 Reflecting Uncertainty in Parameter Inputs

In addition to controlling for stochastic variation that arises from the random sampling of the pathways taken through the model for each entity, it is generally important to represent the effects of uncertainty around the parameter values used as inputs to the model. This is known as parameter uncertainty (parameter here is used as a synonym for estimated input). This kind of uncertainty arises because the true value of each model input parameter is unknown and the value used as an input is only an estimate. For example, the effects of a new technology may be estimated from a clinical trial. Those estimates, regardless of how large and well designed the trial may be, are uncertain, and this is typically reflected in a 95% confidence interval or

other such statistical measure. In a model using that clinical trial result as an input, it would be important to incorporate the parameter uncertainty, generally based on the reported 95% confidence interval. In some cases, the input value may be known to be uncertain, but there may not be sufficient empirical data to quantify that uncertainty adequately. This poses a quandary for the analyst as ignoring variability in this parameter estimate is tantamount to assuming it is known with certainty. Applying an arbitrary range (e.g., ±50%)—although often done—is not an appropriate way to address this uncertainty, but such analyses can be useful to determine if the model outputs are sensitive to variation in these input parameters. If the outputs are judged to be sensitive to those changes, then additional effort should be expended to identify or collect relevant data to better inform the uncertainty. If there is insufficient time or funding to do this, then it may be necessary to elicit ranges for such parameters from experts in the clinical area.

Not all input values are uncertain, however. For example, the price of the technology, the recommended dose of a medication, and the time horizon are known with certitude, and, thus, there is no parameter uncertainty to consider (though there may be interest in how *sensitive* the model results are to alternative values for these inputs—see Section 5.2). Thus, for these kinds of inputs, it is inappropriate to consider variation in their values as part of parameter uncertainty.

The representation of parameter uncertainty is usually an important input to decision making, and so it is not appropriate to report only the results of the model run defined as the base case. Reporting the uncertainty around the base case results (mistakenly, but commonly, known as *sensitivity analysis*) is an essential component of all HTA analyses. Analyses of parameter uncertainty describe the variability among model runs, where the input values are allowed to change to take into account their uncertainty. These analyses can be either deterministic or probabilistic.

5.1.2.1 Deterministic Uncertainty Analyses

Deterministic uncertainty analysis involves the use of particular sets of input values other than those implemented in the base case analysis. These substitute values are chosen so that they are consistent with the estimated variance around the base case (mean) input parameter values. This type of analysis is called *deterministic* because the analyst determines which other specific sets of values to use in exploring the uncertainty. These alternative sets may be assembled by either changing selected values one at a time (i.e., *univariate* analyses) or creating new combinations of values (i.e., *multivariate* analyses). Typically, the alternate values are chosen so that they reflect the underlying range of uncertainty. For example, if a 95% confidence interval has been calculated for an input value, then the DES may be run using the lower border of the interval and again using the upper border. This provides a range of results that addresses uncertainty (but, important to note, the outputs do not

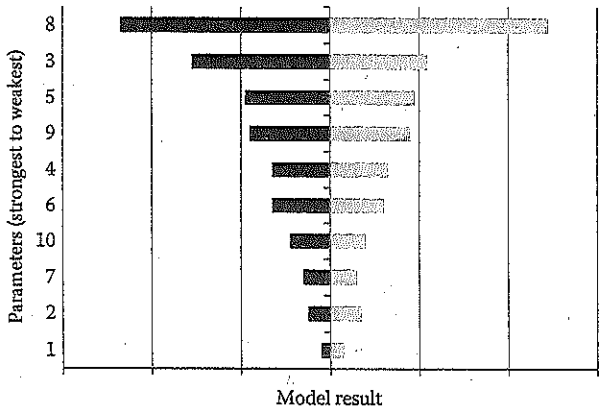


FIGURE 5.3
Display of a deterministic uncertainty analysis. This is a typical *tornado* diagram displaying the variation in an important model result with changes in the values of 10 input parameters. The base case result is at the center, and the bars give the extent of the change in results across the range of the input parameter values. Uncertainty in parameter #8 has the strongest effect on the results, with parameter #3 closely behind. Uncertainty in other parameters has much less impact.

constitute a 95% confidence interval). Results of the deterministic uncertainty analyses are often displayed using a tornado diagram (Figure 5.3).

Implementing this in a DES is very straightforward. The substitute inputs sets are entered in the simulation one at a time, and model runs are undertaken for each one. Depending on the number of inputs to be assessed, around 10 to 20 separate deterministic analyses are typically carried out for an HTA. As model running times tend not to be a major concern for such a limited set of analyses, each run should include the same number of individuals as in the base case. Of course, stochastic uncertainty is still an issue, but reporting of the results of the deterministic analyses focuses on the parameter uncertainty.

5.1.2.2 Probabilistic Uncertainty Analyses

Deterministic uncertainty analysis provides an indication of the parameters to which the base case results are most sensitive but does not provide an estimate of the overall uncertainty in the estimates due to all the components that have variance. To generate confidence intervals around ICERs or other measures (such as net monetary benefit), a probabilistic approach to sampling the sets of alternative input values is often used. This type of analysis, although about uncertainty, is most commonly known as a *probabilistic sensitivity analysis*, or by its acronym PSA (Baio and Dawid 2011; Claxton et al. 2005; Doubilet et al. 1984). In a PSA, the sets of alternate input values are

chosen by the model rather than the analyst. This involves sampling input values from distributions that describe the uncertainty around the true value of each input. Each sampled set of input values is run through the model, just as in the deterministic analyses. The results of each run are stored, all the outputs are reset to zero, and another set of input values is sampled and run. This produces a series of results, which the analyst hopes adequately reflects the underlying parameter uncertainty (Figure 5.4).

Many (often 1,000 or more) sets of inputs are generated and replications are run for each one. To facilitate the process, it is usually automated. Once this is completed, interpretation of the large number of results can be assisted by ordering the values according to a decision-relevant outcome (typically the ICER) and then deriving the cumulative distribution of these results (Figure 5.5). This is commonly known as a *cost-effectiveness acceptability curve* (CEAC) and is interpreted as providing the probability that the technology meets any given threshold criterion for cost-effectiveness (Fenwick et al. 2001; Groot Koerkamp et al. 2007).

A probabilistic uncertainty analysis is readily done in a DES. The analyst begins by specifying the distribution (see Chapter 3) for each input that has parameter uncertainty. Inputs whose values are known with certitude, such as

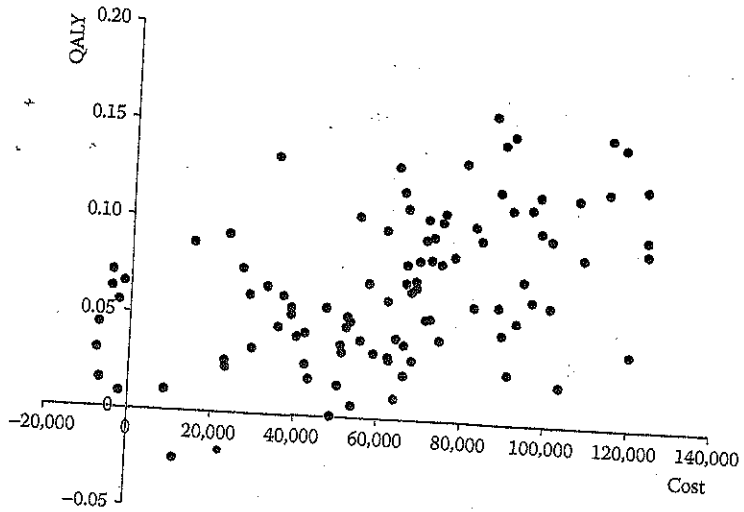


FIGURE 5.4
Scatter plot of results from a probabilistic uncertainty analysis. In this analysis, 100 replications were run, comparing two technologies, allowing the parameter inputs to change according to their uncertainty. The results of each replication, in terms of the net costs and QALYs gained, are plotted, showing the scatter produced by the uncertainty. Some results even reverse direction in the sense that they provide either cost savings or QALY losses. QALY, quality-adjusted life year.

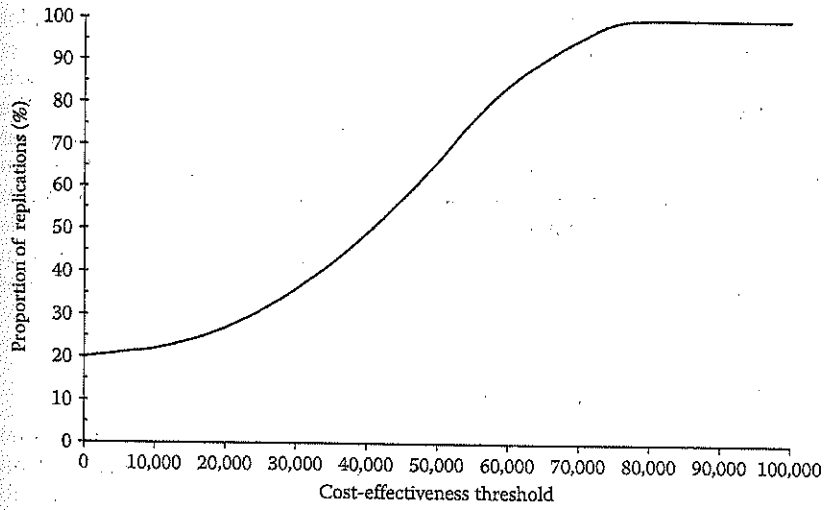


FIGURE 5.5
Cumulative distribution of cost-effectiveness ratios. A probabilistic uncertainty analysis was run and the results of the replications were ordered according to the ICER, from lowest to highest. The cumulative frequency was then computed and is displayed here. It indicates, for example, that two-thirds of the replications yielded a ratio below 50,000 and 75% were below 55,000. ICER, incremental cost-effectiveness ratio.

the discount rate, are not varied in this analysis, even if the analyst is interested in how sensitive the results are to variations in their values (see Section 5.2). There are two ways to implement sampling of the distributions that reflect parameter uncertainty. One is to create an event to do the sampling and store the selected input values in the appropriate global information places. This event has to happen immediately at the start of the simulation so that input values are available as needed. When using specialized DES software, this approach works well. It can also be implemented as an initial subroutine in a general programming language. In a spreadsheet, the values can be sampled directly into the cells that store them for the simulation. This process can be facilitated by using an add-in that has built-in distributions and tools for handling them. See, for example, an excellent recent review by Vose (2014). An alternative approach is to conduct the sampling externally to the simulation and supply a new set of inputs for each model run. Depending on the software used, special programming may be required to automate the process. Either way, the outputs are collected across all the replications and then they are analyzed to generate the required format for the presentation of the uncertainty results. A related form of analysis concerns the value of collecting additional data to reduce uncertainty (value of information analyses) (Claxton and Sculpher 2006), and this type of analysis is implemented in a DES in the same way as is PSA.

Given that many additional model runs are required for a PSA, model running times can increase significantly. This can be eased somewhat, as noted above, by reducing the threshold for achieving stability in the outputs. An alternative approach is to use analysis of variance to inform the combination of model runs and the number of entities per replication that are required to achieve defined levels of accuracy with respect to the mean of the estimate of interest and the variance estimates around that mean (O'Hagan et al. 2005). Alternatively, if a fixed time period for the analysis of a model is available (e.g., 100 hours), the analysis of variance method can inform the most appropriate combination of model runs and the number of entities per replication.

5.1.3 Structural Uncertainty Analyses

Structural uncertainty arises from the many decisions and assumptions that are made during the design and implementation of the model (Haji Ali Afzali and Karnon 2015). If alternative assumptions are made, then the results will likely change, and it is important for the decision maker to understand the degree and direction of those deviations. This structural uncertainty is taken into account by restructuring the model varying the assumptions made for the base case. Each alternative structure reflects one set of assumptions. As there are numerous aspects about which structural assumptions are made, and there may be many reasonable variations for each aspect, the volume of structural alternatives can be considerable. Needless to say, each restructured version of the model is still subject to both parameter and stochastic uncertainty, and the need to address those factors still applies. Thus, full accounting for structural uncertainty is quite challenging. It is not too surprising, then, that uncertainty around the structure of an HTA model is often overlooked by analysts! This should be of considerable concern to the HTA community as it has been shown that structural uncertainty tends to be very significant (Jackson et al. 2011).

DES facilitates the implementation of more complex model structures, which can represent disease processes and interventions more accurately, but also generates more opportunity for structural uncertainty. The more components a model has, the greater the number of alternative structures that may be considered reasonable possibilities. Fortunately, the DES framework also makes it easier to address structural uncertainty, particularly when using some of the specialized software packages.

The process for assessing structural uncertainty begins with careful documentation of the design, the assumptions made and implementation decisions taken, their basis, how they relate to each other, and any credible alternatives. During the implementation, provision can be made in the model structure for the more likely of these alternatives. For example, there may be some doubt about the form of one of the controlling equations—perhaps both a Gamma and a Weibull distribution fit the observed data equally well,

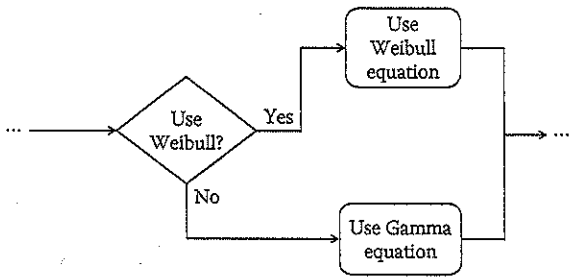


FIGURE 5.6
Implementation of a branching structure to enable structural uncertainty analyses. Here, the modeler sets a flag to indicate whether the Weibull equation should be used instead of a Gamma. Results of runs with the flag set one way can be compared to those with it set the other way to determine if the choice of distribution makes a difference in the information provided to the HTA.

and there is no *a priori* knowledge to help the selection. Having noted this, the modeler can then implement both forms of the equation and define a flag that tells the simulation which equation to use in any given model run (Figure 5.6). During execution, the flag is first set one way for one run and then reset the other way for another run. The results are then compared to determine the significance of the uncertainty around that particular structural factor. If it turns out that it matters (e.g., it changes the direction of the HTA decision), then at a minimum, the decision maker must be given this information. Ideally, further work is done to determine which structure makes more sense, perhaps by finding another dataset against which to validate the predictions, or even by collecting additional data.

Sometimes, alternative design decisions may involve different sets of events and connections. For example, one possibility in modeling the handling of an acute myocardial infarction may be to assume that everyone presenting to the emergency department is sent to the cardiac catheterization lab for angiography. An alternative might consider that some individuals are sent to the coronary care unit first. Ideally, this would be investigated further, but if that is impossible, then a structural uncertainty analysis should be carried out. There could easily be many more components affected by these choices. Consideration of the resulting structural uncertainty is facilitated by incorporating all the reasonable alternatives into the model framework and controlling which one is operative in any one model run by opening or closing one or more *gates*. Thus, in one run the gate allowing the *transfer to coronary care unit* logic is closed and in another it is opened (Figure 5.7).

The results from each model structure tested may be presented to the decision maker as alternative cases, analogous to the reporting of deterministic uncertainty analyses. This does not provide any information on the likelihood that each case reflects reality and may simply overwhelm the decision

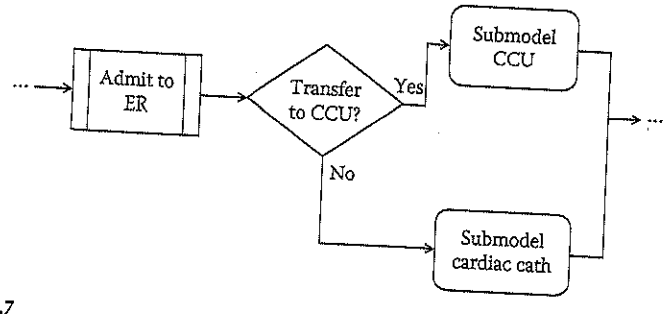


FIGURE 5.7
Use of a gate to implement very different structures reflecting diverse practices. Depending on how the gate is set, patients presenting to the emergency room with chest pain may be sent to the coronary care unit or to the cardiac catheterization suite for an angiogram. cath, catheterization; CCU, coronary care unit; ER, emergency room.

maker with figures that are difficult to interpret. Another possibility is to parameterize the structural uncertainty by specifying probability distributions for the flags and gate openings. For example, a structural alternative that is judged possible but very unlikely might have a controlling gate with an opening probability of 1%. Thus, 99% of the runs would not consider that structure but 1% would. For structural options where this can be done, the structural uncertainty is thus converted into parameter uncertainty and can now be analyzed using the deterministic, and even probabilistic, techniques described in Section 5.1.2. Although this approach of parameterizing structural uncertainty is conceptually simple, there are substantial practical hurdles. The likelihood of each option must be assessed quantitatively and the number of possible combinations may grow rapidly, with the likelihoods possibly correlated. These problems are not specific to DES, but DES makes it easier to address them by facilitating the incorporation of structural options, their controlling gates, and presenting the consequent alternative results.

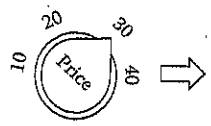
In some cases, it may be too difficult to parameterize alternative structural assumptions, but it may be possible to compare the validity of alternative model structures by comparing the outputs of model to observed data for those outputs (see Chapter 8 for more details on validation). The simplest approach is to select the model that achieves the best validation results, though such an approach does not represent the effects of structural uncertainty. A model averaging approach generates probability weights that represent the accuracy with which each model predicts the observed data. The probability weights comprise an empirical distribution of the probabilities that each of the tested model structures is the most valid model structure. Combined with the probability distributions specified for each input parameter, parameter and structural uncertainty can be represented jointly (Claeskens and Hjort 2008).

To implement the joint modeling of parameter and structural uncertainty, a model structure may be sampled at the start of each run, and the relevant gates are opened and closed to implement the sampled model structure. Next, a set of input parameter values is sampled, and the model run completed and results stored, and the process repeated for the necessary number of model runs. In some cases, the analyst may decide to implement two or more separate models to represent the alternatively specified model structures. Here, values are sampled from the structural probability weights distribution to match the aggregate number of model runs to be generated. Each of the separate models is run according to the number of times each model was sampled, with alternative sets of input parameter values sampled for each model run. The outputs from each model run across the multiple models are then combined and analyzed to jointly represent parameter and structural uncertainty.

5.2 Exploring Sensitivity to Input Values

While uncertainty refers to imprecision in the outputs resulting from imprecision in the input values and doubts about the design and implementation decisions, *sensitivity* refers to the model's responsiveness to changes in the input values. For example, the price of each of the technologies modeled may be known exactly (i.e., there is no uncertainty), but it may still be of interest to assess the degree of change in the outputs if there were to be a particular adjustment in price (Figure 5.8). In this case, the model runs will show how sensitive the results are to changes in the prices, but price variation does not introduce additional uncertainty.

HTAs are fundamentally concerned with the expected performance of a technology in the HTA's jurisdiction. In that local context, additional specificity is required to address the characteristics of the people for whom the intervention is deemed appropriate. Indeed, the model may be used to help



Price	Result
10	23,250
20	26,500
30	28,750
40	31,000

FIGURE 5.8
Sensitivity analysis looking at the impact of changing prices. In this analysis, the modeler successively changed the price of some input and obtained the output for a result of interest. The amount of variation in the result reflects how sensitive this is to variation in this price.

determine in which people the intervention is most cost-effective. Thus, one major sensitivity concern has to do with the population modeled: do the results change substantially if a different *subgroup* is considered?

5.2.1 Subgroups

For many HTAs, it is feasible to make separate decisions regarding the provision of the evaluated technology for different subgroups. In such cases, decision makers may be interested not only in the value of the intervention for the whole population but also in the degree to which an intervention is cost-effective in different subgroups of the population.

Analyses to address subgroup differences can be undertaken as if they were separate problems. Thus, a full set of base case, parameter, and structural uncertainty analyses is undertaken for each subgroup. Fortunately, when using a DES, it is not necessary to actually run the analyses separately. Instead, the analyst identifies the subgroups of interest (e.g., by specifying ranges of age, gender composition, disease severity, and extent of prior treatment). An additional attribute is defined to reflect which subgroup an entity belongs to, or if it is possible to be in several subgroups, then an attribute is specified for each one. During the simulation, each entity's membership in one or more subgroups is stored in the respective attributes set up for that purpose. As each replication proceeds, the outputs are stored in indexed global variables corresponding to the subgroups. At the end of the model run, the outputs are processed by subgroup. This provides the decision maker with the required information more efficiently than running each subgroup separately. An increase in the number of entities simulated is still required, however, to allow for stochastic uncertainty within each subgroup. If parameter uncertainty differs by subgroup, then it may get too complicated to try to run all subgroups in the same analysis. Similar issues arise if structural uncertainty is not common across subgroups.

5.2.2 Other Sensitivities of Interest

This idea of simultaneously considering various subgroups can be extended to exploring the sensitivity of results to changing the values of other model characteristics, even if they bear no uncertainty. These other inputs may not be specific to entities, but instead broadly affect the entire model. One characteristic that is of particular interest in many HTA contexts is the *perspective* of the analyses. Perspective refers to the point of view taken in the analyses, usually having to do with who pays for what. For example, intervention costs will be different from the perspective of the insurer to that of an individual covering co-pays and yet again to that of a hospital provider concerned only about admissions. To consider the impact of these different perspectives, analysts typically run separate analyses for each perspective. Using a DES, however, these various perspectives can be accommodated in a single run

by defining at the start what costs are pertinent to which perspectives and storing the outputs accordingly during each replication. Since much of the work of the simulation is processing the events, and there is much less effort involved in applying costs to any consequences, this simultaneous running of multiple perspectives is much more efficient than rerunning the simulation separately for each perspective.

The sensitivity of a DES to changes in the values of many types of inputs can be incorporated in a simultaneous analysis. For example, the analyst may be interested in the impact of different discount rates or in the degree to which the length of the time horizon matters. Just as with the perspective, these scenarios can be set up by specifying appropriate global variables to store the outputs. In the case of the discount rate, values discounted at different rates would be stored in separate variables. Since the discounting takes place when the related event occurs, the timing is known precisely, and any number of rates can be applied with no need to rerun the simulation. For different time horizons, it is simply a matter of reporting the outputs at the desired model times, but not terminating the simulation until the longest one has elapsed.

5.2.3 Threshold Values

While subgroup analyses address the variation in results in defined portions of the population, another sensitivity question can be about the value of a particular input that alters the results in such a way that the decision maker might reach a different conclusion. For example, if a technology is judged to be unattractive because its benefits do not justify its costs, the decision maker may be interested to identify the price at which the intervention becomes cost effective. This is known as a *threshold* analysis because it is trying to determine the value of an input that leads to the results crossing from one side of an HTA decision to the other.

If the relationship between the results and the input of interest is linear and easily derived, a threshold analysis can be addressed analytically outside the DES. In most cases, however, this is not possible because the connection is not linear or it cannot be readily deduced. In this situation, a threshold analysis requires running the model multiple times with different values of the input and using the results across runs to find the threshold value. This can proceed in the same way as a deterministic uncertainty analysis, varying the input value until the threshold is found. A more efficient alternative is to automate the process by letting the model select the value from a range and narrowing that range progressively until the threshold is found. Many specialized software packages now include functions or modules that facilitate this process.