**Markov Model Replication in Excel**

### AIMS OF THE EXERCISE

The purpose of this extension to the exercise is to replicate the TreeAge Markov model in Excel.

**OVERVIEW OF STEPS TO COMPLETE EXERCISE**

1. Preparing parameters and naming cells
2. Building a Markov model for no drug therapy
3. Adapting the model for the new drug therapy
4. Estimating cost-effectiveness (deterministically)
5. Preparing parameters for probabilistic sensitivity analysis
6. Estimating cost-effectiveness (probabilistic)
7. Half cycle correction

**ORIENTATION**

Open the Excel file **Exercise.xlsm** and click on *‘Yes’* to enable the macros. To check if you have macros enabled in Excel 2007, click on the ‘*Start*’ button, then *Excel Options>Trust Centre>Trust Centre Settings>Macro Settings* and select ‘*Enable all macros*’. In Excel 2010, select *File>Options>Trust Center>Trust Center Settings…>Macro settings* and select *‘Enable all macros*’. If your macro settings are usually set to a lower security level, please ensure this setting is changed back after you have completed this exercise.

The <**Diagram**> worksheet contains the structure of the Markov model that you will build. The Excel version of the model is built across a number of worksheets with separate sheets for the model parameters <**Parameters**>, the Markov model arms of the model <**No drug**>/<**Drug**>, the results <**Analysis**>, and the probabilistic analysis <**PSA**>.

The model consists of four states that characterise a chronic disease:

* Asymptomatic disease (a patient has acquired the disease, but is not suffering any ill consequences of the disease)
* Progressive disease (a patient is showing symptoms of disease)
* Dead from the disease
* Dead from other causes.

Disease states are represented by ovals. The model will be run for several cycles, and at the end of each cycle patients can move between disease states. Possible transitions between states are shown by the arrows. The circular arrows indicate that patients can also remain in the same disease state from one cycle to the next. State costs are shown in red while transition costs are shown in pink. Transition probabilities between states are shown in blue and the utilities of each state are shown in green. All the cells for you to complete in the exercise are coloured green.

A drug has become available that may reduce the likelihood of disease progression for patients in the asymptomatic state. You want to calculate the cost-effectiveness of this drug by using a Markov model to compare the costs and outcomes of two options: i) no drug therapy and ii) drug therapy. **Table 1** reports the parameters that you will use to build your model.

**Table 1. Parameter values for the Markov model**

|  |  |  |
| --- | --- | --- |
| **Name** | **Value** | **Description** |
| ***Costs*** |  |  |
| cAsymp | 500 | Cost of one cycle in the asymptomatic disease state |
| cProg | 3000 | Cost of one cycle in the progressive disease state |
| cDrug | 1000 | Cost of drug therapy for one cycle |
| cDeath | 1000 | Transition cost associated with transition to the “dead from disease” state following the progressive disease state |
|  |  |  |
| ***Quality of life adjustments*** | | |
| uAsymp | 0.95 | Quality of life weight for one cycle in the asymptomatic disease state |
| uProg | 0.75 | Quality of life weight for one cycle in the progressive disease state |
|  |  |  |
| ***Transition probabilities*** | | |
| tpProg | 0.01 | Coefficient of increase for the probability of entering the progressive disease state i.e. the transition probability will increase by 0.01 with each additional year |
| tpDcm | 0.15 | Probability of dying from disease during the progressive disease state |
| natDeath | 0.0138 | Other cause mortality for age 55-64 |
|  | 0.0379 | Other cause mortality for age 65-74 |
|  | 0.0912 | Other cause mortality for age 75-84 |
|  | 0.1958 | Other cause mortality for age 85 and over |
|  |  |  |
| ***Other parameters*** | | |
| eff | 50% | Relative risk of disease progression as a result of drug therapy |
| Ini\_age | 55 | The initial age at which patients start the model |
| oDR | 3.5% | Discount rate for outcomes |
| cDR | 3.5% | Discount rate for costs |

### Please follow the step by step guide to this exercise, which begins on the next page.

### EXERCISE

### 1. Preparing parameters and naming cells

First you have to prepare the <**Parameters**> worksheet to allow the use of parameter names throughout the rest of this exercise. The discount rates and initial age at which patients are deemed to start the model are already included in the worksheet. In this section, you will enter the transition probabilities, state and transition costs, and state utilities, and name the parameters of your model. Take some time to read the descriptions of each variable.

To begin with, focus on just four columns: ‘*Name’*, ‘*Live value’*, ‘*Deterministic’* and ‘*Description’*. The *‘Live value’* column contains an **IF(…)** statement in cell D5 that enables an easy switch between *‘deterministic’* (column C) and *‘probabilistic’* (column D) values. At the moment, you should have a value of *‘0’* in cell D5. It may be useful at this stage to look up the IF statement syntax in the Excel help file. Using **Table 1**:

1. Enter the cost rewards in cells C12:C15 e.g. for ‘*cAsymp’* you should enter 500 in cell C12.
2. Enter the utility rewards for the asymptomatic and progressive states in cells C19:C20.
3. Enter the transition probabilities for progression and death from disease in cells C25-C26.
4. Enter the time dependent probabilities for death from other causes in cells B27:B30.
5. Enter the relative risk reduction of the drug (eff) in cell C33.

Using names instead of cell references will help you to build the model and prevent mistakes.

1. Highlight the *‘Name’* and *‘Live value’* of the asymptomatic cost parameter (i.e. highlight cells A12:B12), select *Formulas>Create From Selection*, ensure that *‘Left Column’* is selected, and click *‘Ok’*.

Check whether the name has been set up successfully by selecting cell B12 - if everything is correct you should read *‘cAsymp’* in the *‘Name Box’* at the top left of the screen above the workbook.

1. Name all of the remaining parameters down to row 38.

### 2. Building a Markov model for no drug therapy

Open the <**No Drug**> worksheet. You should first model death from other causes (natDeath - column C) using the time dependent variables you named in the <**Parameters**> worksheet. You can do this by using the **IF(…)** statement in Excel.

1. Use an **IF(…)** statement to choose between natdeath55, natdeath65, natdeath75, and natdeath85 based on the age of the patient at each point in time ***(ini\_age+cycle)***:

=IF(ini\_age+A6<55,"error",IF(ini\_age+A6>=85,natDeath85,IF(ini\_age+A6>=75,natDeath75,IF(ini\_age+A6>=65,natDeath65,IF(ini\_age+A6>=55,natDeath55,"error")))))

Next, you will build a Markov trace that will show the number of patients that are in any one state at any one time. Columns E to H represent the four main states of the model and column I provides a check (the sum across E to H must equal the size of the original cohort).

1. Familiarise yourself with the possible transitions between model states - make sure you understand **Table 2**:

**Table 2: Transition matrix (no drug arm)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **State**  **(From)** | **(To)**  **Asymptomatic** | **Progressive disease** | **Dead (from disease)** | **Dead**  **(other causes)** |
| Asymptomatic | 1-tpProg\*cycle-natDeath | tpProg\*cycle |  | natDeath | |
| Progressive disease |  | 1-tpDcm-natDeath | tpDcm | natDeath | |
| Dead (from disease) |  |  | 1 |  | |
| Dead (other causes) |  |  |  | 1 | |

1. Use the transition matrix to populate the Markov model. Start with row 7 and once you have this row correct copy it down to the other rows. It is good practice to complete all states as in the transition matrix and then sum across the states (in column I) to make sure the total is equal to the size of the original cohort (i.e. 1000). Do not set up one state as the residual of the remaining states. Make sure your formulas are referring to the right cell in the ‘natDeath’ column, e.g. the transition of individuals from Asymptomatic to Dead other causes in cycle 1 should be informed by the value in cell C6.
2. In column K, calculate the life years for each cycle of the model. You just need to add the number of patients in the asymptomatic and progressive state per cycle.
3. In column L, calculate the quality adjusted life years by cycle. You should multiply the state utilities by the number of patients in each respective state per cycle.
4. In columns M and N, calculate the life years and QALYs appropriately discounted. The discount factor is ***1/((1+discount rate)^cycle))***.
5. In column P, calculate the state costs for each cycle of the model. Again multiply the number of patients in each state by the state costs.
6. In column Q, calculate the transition cost from the progressive state to dead from disease. It is assumed that these costs occur in the beginning of the cycle.
7. In column R, sum the state and transition costs. This will help you estimating the total discounted costs in column S. The discount factor is ***1/((1+discount rate)^cycle))***.
8. In row 54, sum the columns to get the total effects and costs of the entire cohort. Finally, in row 55, divide the totals by 1000 to get per patient predictions of life expectancy, QALYs and cost for this arm of the model.

### 3. Adapting the model for the new drug therapy

1. Select the left hand highest corner cell (i.e. grey cell above 1 and to the left of A) right-click on the selected cells and chose *‘Copy’*. Open the <**Drug**> worksheet, right-click on cell A1 and select *‘Paste’*. You have now have an exact duplicate model of the ‘No drug’ arm. This is to avoid repeating building the model arm again from scratch.
2. Change cells A1 and G3 to ‘Drug therapy’.

There are only two differences between the ‘No Drug’ and ‘Drug’ arms of the model. First, you have to introduce the drug treatment effect on preventing disease progression. Second, you have to incorporate the costs of the drug therapy.

1. Use **Table 2** to make your changes to this arm of the model. Start with row 7 and once you have this row correct copy it down to the other rows.

**Table 2: Transition matrix (drug arm)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **State**  **(From)** | **(To)**  **Asymptomatic** | **Progressive disease** | **Dead (from disease)** | **Dead (other causes)** |
| Asymptomatic | 1-tpProg\*cycle\*eff-natDeath | tpProg\*cycle\*eff |  | natDeath |
| Progressive disease |  | 1-tpDcm-natDeath | tpDcm | natDeath |
| Dead (from disease) |  |  | 1 |  |
| Dead (other causes) |  |  |  | 1 |

### 4. Estimating cost-effectiveness (deterministically)

You have successfully replicated the TreeAge model in Excel. Now, select the <**Analysis**> worksheet to estimate the cost-effectiveness of drug therapy compared to no drug therapy.

1. Link the cells for costs, life-years and QALYs with the total discounted values per patient of the two different treatment arms (<**Drug**> and <**No drug**>).
2. In cells D11:F11, calculate the difference in cost, life years and QALYs between both arms.
3. In cells D14:E14, calculate ICERs using life-years and then QALYs as the measure of health benefit.

* The results should be (without half-cycle correction):
* the cost for the no drug intervention is £11,832 and the cost for the drug intervention is £20,115 (cost difference of £8,283)
* the QALYs for the no drug intervention are 9.22 and the QALYs for the drug intervention are 10.48 (QALY difference of 1.26)
* The LYs for the no drug intervention are 10.24 and the LYs for the drug intervention are 11.41 (LY difference of 1.18)
* the incremental cost effectiveness ratio is therefore £6,573 per QALY (£8,283/1.26) or £7,047 per LY (£8,283/1.18)..

*(see Markov\_Excel\_solution.xls)*

### 5. Preparing parameters for probabilistic sensitivity analysis

1. Open the <**Parameters**> worksheet and enter the value *‘1’* in cell D5. You will see that most ‘live value’ cells now become blank. The model parameters of initial age, discount rates and death from other causes were left deterministic.

Do section below or copy values from ‘parameters for PSA\_solution’ file.

**Table 3** has all the statistics that you need to make the model probabilistic. You will find them set out for you in the worksheet.

**Table 3: Model parameters - statistics for PSA**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Parameter** | **Mean** | **Standard Error** | **95% CI** | **Distribution** |
| ***Costs*** | | | | |
| cAsymp | 500 | 127.6 | - | Gamma |
| cProg | 3000 | 510.2 | - | Gamma |
| cDrug | 1000 | 102.0 | - | Gamma |
| cDeath | 1000 | 255.1 | - | Gamma |
|  |  |  |  |  |
| ***Quality of life adjustments*** | | | | |
| uAsymp | 0.95 | 0.026 | - | Beta |
| uProg | 0.75 | 0.077 | - | Beta |
|  |  |  |  |  |
| ***Transition probabilities*** | | | | |
| tpProg | 0.01 | 0.001 | 0.008-0.012 | Log Normal |
| tpDcm | 0.15 | 0.026 | - | Beta |
|  |  |  |  |  |
| ***Other parameters*** | | | | |
| eff | 0.5 | 0.051 | 0.4-0.6 | Log Normal |

1. Uncertainty surrounding the cost parameters is modelled using a Gamma distribution which is constrained on the interval 0 to positive infinity. You should use the method of moments approach to estimate the hyperparameters of the gamma distribution (alpha and beta), where:

***Alpha = (mean^2)/(se^2)***

***Beta = (se^2)/(mean)***

and ***se = standard error of the mean***

Use the ***GAMMAINV(RAND(), alpha, beta)*** function in column D to generate a random draw from each distribution. It may be useful at this stage to look up the function in the Excel help file.

1. State utilities and the transition probability ‘*tpDcm*’ are modeled using a Beta distribution, which is constrained on the interval 0 to 1. Again you should use the method of moments approach to estimate the hyperparameters of the beta distribution (alpha and beta), where,

***(alpha + beta) = [mean\*(1-mean)/(se^2)] – 1***

***alpha = mean\*(alpha + beta)***

and ***se = standard error of the mean***

Use ***BETAINV(RAND(), alpha, beta)*** function in column D to generate a random draw from each distribution.

1. The effectiveness of the drug, *‘Eff’*, and the coefficient of increase ‘*tpProg’* are each modelled using a Log Normal distribution, which assumes that the parameters are normally distributed on a log scale and is useful for modelling uncertainty in ratios. The mean and 95%CI of the effectiveness parameter can be log-transformed to give:

***-0.69 (-0.91 to -0.51) on log scale***

And calculate the SE on the log scale:

***(LN(High CI) – LN(Low CI))/(1.96\*2)   
(-(0.51) – (-0.91))/(1.96\*2) = 0.10***

Enter the mean of logs in cell G33 and the standard error on log scale in cell H33.

1. In cell D33, generate a normally distributed random variable with mean –0.69 and SE 0.10 ***(NORMINV(RAND(), -0.69, 0.10))*** and exponentiate the resulting variable.
2. Now, use the same approach to model uncertainty around the parameter ‘tpProg’.

### 6. Estimating cost-effectiveness (probabilistic)

Open the <**PSA**> worksheet and take some time to look at the contents. The purpose of probabilistic sensitivity analysis (PSA) is to model the uncertainty of the model parameters using probability distributions of their values. For each run of the model a value of each parameter is taken at random from its probability distribution. Then, costs and effects are estimated for each run using the randomly collected values of all parameter distributions. If we perform a large number of runs or trials, such as 1000, we will obtain a distribution of costs and effects that represents the uncertainty in the results of the model conditional on the distributions of the parameter values and the structure and assumptions of the model.

1. Cells B3:J3 are labelled with the parameter names. The aim is to record the parameter values as well as the overall cost and effect results (cells B4:Q4) for each round of simulations. Link the cell under the parameter name with the relevant probabilistic parameter, either using the name of the parameter or the *‘live value’* in the <**Parameters**> worksheet. Link the cells under the cost and effect labels with the respective cells in the <**Analysis**> worksheet.
2. Cells B4:Q4 should update to new parameters when the <*F9*> key is pressed, if you have correctly entered ‘1’ in cell D5.

We have built a Macro for you that will repeatedly copy cells B4:Q4 into the 1000 rows below. This will enable you to estimate the mean costs and effects of the 1000 probabilistic trials for each arm of the model (see cells L1007:Q1007) and the respective ICERs.

1. Open <**Analysis**> and click on the *‘Run probabilistic sensitivity analysis’* button to perform the 1000 simulations. Once the simulations are finished, link cells K9:M10 in the <**Analysis**> worksheet with cells L1007:Q1007 in the <**PSA**> worksheet.
2. Estimate the ICERs using both life-years and QALYs as the measure of health benefit.

Click on the ‘Run CEAC’ button to perform the calculations required for the cost-effectiveness acceptability curve. See cells S6:T6 in <**PSA**> worksheet for code to estimate net benefits from the probabilistic pairs of costs and effects and cells V6:W6 for code to identify which of the interventions has the highest net benefit.

1. Open **<CEA curve>** to see the cost-effectiveness acceptability curve.

**7. Half cycle correction**

In both the arms of the model, we assumed that transitions between model states occur at the beginning of the cycle with the number of patients in each state being counted at the end of the cycle. However, the transitions could occur at any point in time during each cycle. One possibility is that on average transitions happen halfway through the cycle. Hence, our 1000 patients would transit from the asymptomatic to the progressive/dead from other causes states halfway through cycle 1 rather than at the beginning. Each of these patients would then contribute an additional half-cycle to the estimate life expectancy of 13.7 years. We should therefore add this 0.5 to total life expectancy as well as adjusting the quality adjusted life expectancy (QALE) and lifetime costs. QALE and costs should be adjusted by adding half the expected utility and costs of the individuals at the beginning of cycle 1, i.e. 1000 individuals in asymptomatic state and 0 individuals in the remaining states, to the estimated lifetime QALYs and costs.

Remember that we are adjusting our estimates in this way because we assume that the transitions occur at the beginning of the cycle while we count the number of people in each state at the end of the cycle.

In columns V to AD, in <**Drug**> and <**No Drug**> worksheets, you will find the half cycle correction made for you. Take some time to look at row 6 of these columns.

Now, select the <**Analysis**> worksheet to estimate the cost-effectiveness of drug therapy compared to no drug therapy with half-cycle correction.

Link the cells for costs, life-years and QALYs with the total discounted values per patient of the two different treatment arms (Row 55 of columns V to AD in <**Drug**> and <**No drug**>).

In cells D20:F20, calculate the difference in cost, life years and QALYs between both arms. In cells D23:E23, calculate ICERs using life-years and then QALYs as the measure of health benefit.

* the cost for the no drug intervention is £12,082 and the cost for the drug intervention are £20,865 (cost difference of £8,783)
* the QALYs for the no drug intervention are 9.69 and the QALYs for the drug intervention are 10.95 (QALY difference of 1.26)
* The LYs for the no drug intervention are 10.74 and the LYs for the drug intervention are 11.91 (LY difference of 1.18)
* the incremental cost effectiveness ratio is therefore £6,970 per QALY (£8,783/1.26) or £7,472 per LY (£8,783/1.18).

*(see Markov\_Excel\_solution.xls)*