Stochastic trees and the *StoTree* modeling software

*StoTree* is a software tool for the formulation and solution of continuous-time Markov models of medical interventions. Markov chain models were introduced to the medical literature by Beck and Pauker [1] and provide a convenient means to account for medical treatment options and risks that occur not only in the present but also in the near and distant future. For a more recent introduction to Markov models in medicine, see [2] or [3]. *StoTree* represents Markov models using the graphic technique of *stochastic trees*, introduced by Hazen [4-6].

*StoTree* is implemented in the Excel spreadsheet environment. It is a Microsoft Excel add-in for Office 2007, 2010 and 2013 that occupies its own tab on the Office ribbon. It is written in VBA for Office and therefore requires an Office installation on a PC – it will not function on a Macintosh computer. Although I have not checked, I do not believe it will function in Office Online or Office 365.

This document contains three major sections. The first section *Stochastic trees* gives an introduction to stochastic trees. This is the place to start if you are a beginner. Those familiar with stochastic trees may skip to the second section *Using StoTree to construct a stochastic tree*. The third major section *Factored stochastic trees* discusses factoring, a powerful tool that is one of the unique features of *StoTree*. In its final subsection *The benefits of factoring*, we describe the advantages of using factoring in model construction.

**Installing StoTree**

You can download a copy of *StoTree* from this website. It should appear as the file StoTree7cX.xlam on your computer. Store it in any directory you like, but note where you have stored it. To activate it in Excel, follow the path File\Options\Addins, then click on Go… and Browse… to find the location at which you stored *StoTree*. Click on the appropriate buttons to complete the installation.

The default path for Office 2013 add-ins is c:\Users\[your identifier]\Appdata\Roaming\Microsoft\AddIns. Although it is not required, you may wish to store StoTree7cX.xlam at this location as it is the first place Excel looks for addins.

StoTree7cX.xlam is an *unlicensed* copy of *StoTree*. An unlicensed copy allows you to formulate Markov models, but not solve them. It also allows you to do *what-if* analyses on models that licensed users have formulated and solved. You can obtain a license at my vendor [FastSpring](#).
## Contents

Stochastic trees and the *StoTree* modeling software ................................................................. 1

- Installing *StoTree* ..................................................................................................................... 1
- *StoTree* v.7c ............................................................................................................................... 2

Stochastic trees ................................................................................................................................ 4

- Rates versus probabilities ............................................................................................................. 5
- Cohort analysis: Quality, cost and discounting ............................................................................ 6

Using *StoTree* to construct a stochastic tree ............................................................................. 8

Overview of *StoTree* capabilities ............................................................................................ 8

- Constructing a basic stochastic tree – in short ........................................................................... 8
- Constructing a basic stochastic tree - detail ................................................................................ 8
- Calculation settings ..................................................................................................................... 10
- The cohort sheet ....................................................................................................................... 10
- Tips and shortcuts for constructing a basic tree ........................................................................ 12
- QALYs and costs ....................................................................................................................... 13
- Sharing your constructed workbook ........................................................................................ 13
- Defined names for documenting and sensitivity ..................................................................... 14
- One-way sensitivity analysis ..................................................................................................... 15
- Time-dependent parameters ..................................................................................................... 15
  - Human mortality rates .............................................................................................................. 16
  - Constant intervals of clock time ............................................................................................ 17
  - Clock time versus sojourn time ............................................................................................. 17
- Decision nodes .......................................................................................................................... 18
- Terminal nodes .......................................................................................................................... 19
- Root nodes .................................................................................................................................. 19

Factored stochastic trees ........................................................................................................... 20

Using *StoTree* to construct a factored stochastic tree ............................................................ 20

- Dependencies ............................................................................................................................ 20
- Triggers ....................................................................................................................................... 22
- Priorities* .................................................................................................................................... 23

The cohort sheet for a factored tree .......................................................................................... 23
Active and Inactive factors .................................................................................................................. 24
Quality and cost for factored trees ..................................................................................................... 24
Referencing defined names instead of creating dependencies .......................................................... 24
Tips and tricks for constructing factored stochastic trees ................................................................. 25
  Use an additional factor instead of duplicate subtrees ................................................................. 25
  Dependencies may only reference stochastic nodes ................................................................. 27
  Use a negative rate to indicate an absent arrow ........................................................................ 27
The benefits of factoring .................................................................................................................. 28
Appendix ............................................................................................................................................. 29
  Built-in functions ......................................................................................................................... 29
    Gompertz hazard rate for human survival .............................................................................. 29
    Spike function ....................................................................................................................... 30
    Modulo function ................................................................................................................... 30
    Weibull hazard rate function ............................................................................................. 30
References ........................................................................................................................................... 32
Stochastic trees

For those familiar with stochastic or decision-analytic modeling, a stochastic tree can be characterized in several equivalent ways:

- as a continuous-time Markov chain transition diagram with chance and decision nodes added,
- as a decision tree with stochastic transitions added,
- as a continuous-time version of a Markov cycle tree.

A simple stochastic tree is shown in Figure 1, which depicts the risk of pulmonary embolism over time. It consists of nodes (No Embolism, Embolism, Dead) and arrows connecting the nodes. Nodes depict health states. The arrows are of two types, either wavy or straight. A node or health state can have incremental impact (impact that accrues over time) or instantaneous impact (an impact that is immediate), depending on the type of arrows emanating from it. Wavy arrows emanate from an incremental impact state (such as No Embolism) and indicate that the incremental impact state is occupied for a duration that is uncertain but dependent on the rate at which exit from that state occurs. In Figure 1, the exit rate from No Embolism is $r_{PE}$, and it labels the wavy arrow emanating from that health state. In StoTree, incremental impact states are known as stochastic nodes.

![Figure 1. A simple stochastic tree, depicting the risk of pulmonary embolism over time.](image)

Straight arrows emanate from an instantaneous impact state (such as Embolism), and point to the possible new states immediately following. Each straight arrow is labeled with a probability representing the chance that exit from that state occurs along that arrow. Exit from that state occurs immediately – the state is occupied for zero duration. (This is obviously an idealization – any health state must be occupied for a positive amount of time. However, zero duration can be a convenient modeling assumption for short-duration states.) Probabilities exiting instantaneous states must sum to 1. In StoTree, instantaneous impact states are known as chance nodes.

In Figure 1, exit from the instantaneous state Embolism leads to Dead with probability $p_{DEmb}$, and back to No Embolism with probability $1-p_{DEmb}$. The dashed border around the second No Embolism node means that it is a copy of the first node and that transition occurs back to the first node. Overall, this stochastic tree depicts a patient subject to repeated pulmonary embolisms with a chance of death after each embolism.

A second stochastic tree is depicted in Figure 2. Here the tree begins with a square node 0 depicting a decision that must be made initially. Choice must be made between Liver-Directed Therapy, Living Donor Transplant, and Resection (i.e. surgical removal of a tumor). Straight arrows emanate from node 0 with probabilities 1,0,0 summing to one like with chance nodes – but here each probability must be 1 or 0 to correspond to a decision chosen (or declined) with certainty. In Figure 2, the decision chosen is Liver-

---

1 All stochastic tree diagrams in this document are screen captures from StoTree.
Directed Therapy. Decision nodes are also instantaneous impact, are occupied for zero duration, and are exited immediately.

Figure 2. A stochastic tree with initial decision node.

Figure 2 includes non-death nodes without any emanating arrows – Cured, Not Cured, and Survive Rs. These are by convention stochastic nodes (incremental impact states), with the distinction that the state is occupied permanently and never departed.

Figure 3 shows a third stochastic tree, modeling HIV progression. Here all nodes are stochastic, and two stochastic transitions compete at both node A and node B.

Figure 3. A stochastic tree with competing stochastic transitions.

Top

Rates versus probabilities
The distinction between a rate and a probability may be made as follows. When an event (such as pulmonary embolism in Figure 1) can occur repeatedly over time, a rate satisfies

\[ \text{rate} = \text{event count per person per unit time}. \]
So if we observe \( k \) events among multiple subjects in a total of \( n\Delta t \) person-years, then the event rate \( r \) is estimated by

\[
r = \frac{k}{n\Delta t}.
\]

Rates have units of time\(^{-1}\). For example, we might observe \( k = 250 \) events in \( n\Delta t = 200 \) person-years. This would yield an estimated rate of \( 250/200 = 1.25 \) events per person per year. As you can see, a rate must be nonnegative, but can exceed one.

On the other hand, when time is not an issue, and we are only interested in whether or not a given event occurs (e.g., death given pulmonary embolism in Figure 1, or cure given liver-directed therapy in Figure 2), then the probability of that event satisfies

\[
\text{probability} = \frac{\text{event count}}{\text{person}}.
\]

So if we observe \( k \) events among \( n \) subjects, then the event probability \( p \) is estimated by

\[
p = \frac{k}{n}.
\]

The notions of rate and probability are related whenever we are counting events that occur over time but we are interested in a particular duration \( \Delta s \). If events occur at rate \( r \), and \( p \) is the probability of at least one event in duration \( \Delta s \), then

\[
p = 1 - \exp(-r\Delta s).
\]

(1)

For small \( \Delta s \), probability is approximately rate times duration:

\[
p \approx r\Delta s.
\]

For example, if the estimated rate is 1.25 events per person-year, as above, and we are interested in a one-month duration, then \( \Delta s = 1/12 \) yr, and the probability of at least one event in a one-month duration is

\[
p = 1 - \exp(-(1.25)(1/12)) = 0.0989.
\]

Note that the approximation \( p \approx r\Delta s \) gives \( p = (1.25)(1/12) = 0.1042 \), reasonably close since \( \Delta s \) is small, but not exactly correct.

Most software for Markov modeling is inherently discrete-time, and allows only probability parameters as input. Users are then required to convert rates to probabilities via (1). StoTree models continuous-time Markov chains, and allows both rate and probability inputs. StoTree converts rates to probabilities automatically as needed for calculations using (1), relieving the user of this task.

The relation (1) assumes that the event rate \( r \) is constant over time. Some rates – for example, mortality rates – are time-dependent. In this case the relation between rates and particular-duration probabilities is more complex. See, for example [3].

**Cohort analysis: Quality, cost and discounting**

The traditional method of in the medical literature of solving Markov models is to perform a cohort analysis. This method assumes that a cohort of identical individuals begins in some initial health state \( x_0 \) at time \( t = 0 \), and progresses forward in time through the health states in the model, according to the
assumed rates and probabilities in the model. The analyst selects a small time increment \( dt \), called the **cycle length**. The rates and probabilities in the model determine the **transition probabilities**

\[
\lambda_{xy}(dt,t) = \text{probability the state at time } t + dt \text{ is } y \text{ given that the state at time } t \text{ is } x.
\]

The analyst then sets the cohort size \( n_{x_0}(0) \) in state \( x_0 \) at time \( t = 0 \), and calculates the values of

\[
n_y(t) = \text{expected count of individuals in state } y \text{ at time } t
\]

for all time instants \( t = 0, dt, 2dt, 3dt, \ldots \) using the Kolmogorov forward equations

\[
n_y(t+dt) = n_y(t) + \sum_{z \neq y} n_z(t) \lambda_{yz}(dt,t) - n_y(t) \sum_{z \neq y} \lambda_{zy}(dt,t).
\]

(This is mathematically equivalent to solving the Kolmogorov system of ordinary differential equations using the Euler method of discretization. The Euler method is the least accurate of many numerical methods for solving ordinary differential equations, and there is no reason a more sophisticated method could not be employed. However, the Euler method has been the default for cohort analysis due to its simplicity, and is adequate for most medical analyses.)

The purpose of computing the expected counts \( n_y(t) \) is to use them to calculate expected **quality adjusted life years (QALYs)** and expected costs. The analyst uses the inputs

\[
Q_y = \text{expected quality accrued per unit time in state } y \text{ (the so-called utility of state } y)
\]

\[
C_y = \text{expected cost accrued per unit time in state } y
\]

to compute expected QALYs by forming the sum of products \( \sum_y Q_y n_y(t) \), possibly weighting the result by a time-discount factor, and then summing the all resulting quantities over all time instants \( t = 0, dt, 2dt, 3dt, \ldots \). The analogous operations are performed to calculate expected costs. It is also possible to include quality tolls \( Q_{xy} \) and cost tolls \( C_{xy} \) that are incurred whenever the cohort or a portion of it moves from state \( x \) to state \( y \). Note that if the utilities \( Q_y \) are all equal to 1 and there are no quality tolls, then expected QALYs is simply expected lifetime.

*Top*
Using *StoTree* to construct a stochastic tree

Overview of *StoTree* capabilities

*StoTree* is a Microsoft Excel add-in possessing its own tab on Excel’s ribbon (see Figure 4). You can use *StoTree* to formulate a continuous-time Markov model of health interventions of interest. *StoTree* allows you to graphically characterize the consequences of interventions in one or more compact diagrams. You can next specify rates for events that occur over time, and probabilities for one-time events. You can also input quality and cost parameters. You can then instruct *StoTree* to perform a cohort analysis in order to compute comparative effectiveness and/or cost-effectiveness. Finally, if you are concerned about the impact of particular parameter inputs, *StoTree* can perform a sensitivity analysis on such parameters.

![Figure 4. The StoTree tab on the Microsoft Excel ribbon. (Note: The identity of the remaining tabs will depend on what other software may be installed on your computer.)](image)

Constructing a basic stochastic tree – in short

Once you install *StoTree* as an add-in, you can use the commands on the *StoTree* tab to construct, populate and solve a stochastic tree. In overview, the steps are:

1. Insert a new factor.
2. Add one or more nodes.
3. Connect the nodes with arrows as desired.
4. Populate arrows with rate or probability parameters.
5. Instruct *StoTree* to setup a cohort analysis worksheet, and have *StoTree* calculate that sheet to compute expected QALYs and costs.

We discuss these steps in detail in the following sections.

*Top*

Constructing a basic stochastic tree - detail

To construct a basic stochastic tree using *StoTree*, you should open an empty Excel workbook, and proceed more or less from left to right on the *StoTree* tab. You should execute steps in roughly the following order:

1. *Insert a new factor.*
   
   Click on Insert Factor to insert a blank worksheet into your workbook, and set up your workbook to host stochastic tree construction. (You will see a dialog asking whether you wish to insert a stochastic tree into this workbook, and you should click Yes.) The resulting worksheet will have cells but no cell borders, and there will be an Active/Inactive toggle button in its upper left corner.

2. *Add one or more nodes.*
   
   a. Select a cell in the newly created worksheet where you would like *StoTree* to place a new node, and click on Add Node. A userform will appear that allows you to name the node and choose its type. *StoTree* will insert the name you provide as a caption in the node it draws.
b. Add other nodes to other locations in the worksheet by selecting other cells and clicking on Add Node again.
c. You can move nodes around by dragging them with your mouse, and you can alter their shape and captions as well.

3. Connect the nodes with arrows as desired.
   a. To draw an arrow from one node to another:
      i. Click first on an empty cell to deselect any node you may have currently selected.
      ii. Click on the node where you want your new arrow to start.
      iii. Click on the node where you want your new arrow to end. (Note: You do not have to press the Ctrl key – node selection in StoTree is sticky, and both nodes will remain selected when you click on the second one.)
      iv. Click on one of the arrow button on the StoTree tab. StoTree will draw an arrow of the type you chose from the first node to the second.
   b. You can drag to move nodes around your worksheet, but doing this will leave the arrows behind. Do not attempt to drag the arrows. Instead, click on the Refresh command on the StoTree tab, and arrows will re-attach to all nodes you have selected (or will reattach to all nodes if you have not selected any).

4. Revise your construction as needed.
   a. You may copy/paste to construct a tree having subtrees that are identical. You may do so as follows. (However, using factoring is potentially more simple and elegant – see the section below.)
      i. Select the root node of any subtree you wish to copy.
      ii. Click the Copy Subtree button on the StoTree tab. StoTree will copy the subtree to a new blank worksheet.
      iii. Click on a cell in your original worksheet where you would like the copied subtree to be pasted.
      iv. Click on the Paste Subtree node. StoTree will paste the copied subtree into your worksheet with its root node at the cell you selected.
      (Please note that you cannot use Excel’s Copy and Paste commands to duplicate existing nodes or arrows – or rather you can, but StoTree will not recognize the duplicates as part of your stochastic tree.)
   b. If needed, you can reroute arrows you have mistakenly drawn. Select any part of the arrow, and click on Reroute on the StoTree tab.

5. Populate arrows with rate or probability parameters.
   You will notice that the arrows that StoTree draws are initially labeled “Needs populating”. Click on any such arrow and then click the Parameters button on the StoTree tab. StoTree will open a user form where you may enter a desired rate (for a wavy arrow) or probability (for a straight arrow). You can ignore the Quality Toll and Cost Toll entries for now. You can enter a numeric value, or any numeric expression that will successfully evaluate in your Excel spreadsheet. You can include spreadsheet cell references in your expression if you provide defined names for the relevant cells (see below).

6. Instruct StoTree to setup a cohort analysis worksheet.
   a. Click on Setup Cohort Sheet. StoTree will open a new worksheet called the cohort sheet, and will insert formulas that will perform a cohort analysis when the sheet is repeatedly calculated.
b. Follow the instructions on the cohort sheet to manually perform a cohort analysis, or simply click on the Calculate sheet button on the StoTree tab to complete the cohort analysis in one step.

**Calculation settings**

StoTree provides default calculation settings for the cohort sheet it creates. However, you may modify these settings. Click on the Calculation Settings button on the StoTree tab to enter your desired settings. The settings you can choose are:

- **Initial Cohort.** The number of individuals in the cohort. StoTree calculates fractional values of the expected counts \( n_t \) as needed, so it is adequate to set the initial cohort size to 1.
- **Initial Time.** The clock time \( t \) at which the cohort begins. Usually this would be \( t = 0 \).
- **Cycle Time.** The increment \( dt \) by which time is incremented in each cycle of the cohort analysis. Ideally this should be a small value, but small \( dt \) increases calculation time. Larger values of \( dt \) will calculate quickly but results will be less accurate. Use large values of \( dt \) when debugging your model, but switch to small values after debugging is complete and you want accurate results.
- **Time Horizon.** The clock time at which you wish cohort analysis to stop.
- **Discount Rate.** If you choose discount rate to have value \( r \), then QALYs and costs incurred at time \( t \) will be weighted by the discount factor \( e^{-rt} \). Set the discount rate to zero if you want no discounting.
- **Calculate Expected QALYs (Yes/No).**
- **Calculate Expected Costs (Yes/No).**

**The cohort sheet**

We illustrate here the cohort sheet output that StoTree produces for the HIV-AIDS model of Figure 3. The user has supplied values for the parameters \( r_{AB}, r_{BC}, r_D, r_{D\_AIDS} \) (these are defined names in Excel – see below), and in addition has entered cost rates in each state. When the user clicks on the Setup cohort sheet button on the StoTree tab, StoTree produces the cohort sheet in Figure 5. In that figure, StoTree has performed one cycle of cohort analysis: The cycle time is 0.1, and in the **Time** column, only the time-instants \( t = 0, t = 0.1 \) appear. The initial cohort size is 1000 in state A, and during this cycle, the state counts have moved from \( (1000,0,0,0) \) to \( (948.749, 34.3946, 0, 16.8563) \) in states A,B,C and Terminal Node (death). Accrued QALYs for the cohort are 99.157 yr., and accrued costs are $225,880. The user can click the plus (+) signs on the left to inspect details in the calculations for counts, QALYs and costs that may be useful for debugging.

If the user presses the Calculate sheet button, or the same button on the StoTree tab, StoTree will perform all additional cycles of cohort analysis, using a cycle length of 0.1 yr., until the time horizon \( t = 20 \) is reached. The results are presented in Figure 6. Discounted cumulative QALYs for this cohort of 1000 is 3588.62, and discounted cumulative cost is $13 million.
Figure 5. The cohort sheet StoTree produces for the stochastic tree model of Figure 3 when the user clicks the Setup cohort sheet button on the StoTree tab.
As suggested by the other two buttons on the cohort worksheet, the user also has the option of resetting calculations to the start time, which will produce once again the worksheet displayed in Figure 5; and then stepping the calculations ahead one cycle at a time. This procedure could be useful for debugging.

Once you have instructed *StoTree* to create a cohort sheet of this type, you should save your workbook as a macro-enabled workbook, that is, with a "xlsm" suffix. The workbook now contains Visual Basic macros in a hidden code sheet, and must be saved with this suffix in order for the macros to function.

Although you need a licensed copy of *StoTree* to create a cohort sheet, the resulting workbook may be viewed and manipulated by anyone who has even an unlicensed version of *StoTree* installed. You may therefore share your workbook (see below) with colleagues possessing only unlicensed copies of the *StoTree* addin. Your colleagues may wish to perform what-if analyses on your model, and may do so without acquiring a license. See the section *Defined names for documenting and sensitivity* below for more information on how to do what-if analyses.

**Top**

**Tips and shortcuts for constructing a basic tree**

There are a number of shortcuts you can use to construct basic trees more quickly and elegantly.

1. You can instruct *StoTree* to draw multiple arrows in one click by first selecting node you want to serve as origin (call it node A), and next selecting all destination nodes (call them nodes B, C and D). Then a single click of an arrow button on the *StoTree* tab will draw arrows from A to B, A to C and A to D.
2. As already mentioned, node selection in *StoTree* is sticky – you can click several nodes in succession to select them all, without using the Ctrl key in Excel. In fact, if you wish to take advantage of the shortcut in point 1, you must select nodes without using the Ctrl key – otherwise *StoTree* does not recognize the multiple selection. You should also be careful, if you want to select only one node, that other nodes are not already selected. To achieve this, simply click on any cell in the worksheet and all nodes will be de-selected.

3. If you determine that an arrow or a node is no longer needed, you can simply click on the node or any part of the arrow and press the Delete key on your keyboard. *StoTree* will notice the deletion the next time you click one of the buttons on its tab. At this point, if you have deleted a node, then *StoTree* will also delete any arrows connected to that node. If you delete just part of an arrow, then *StoTree* will delete the entire arrow when it notices the partial deletion.

4. If you mistakenly draw an arrow of the wrong type or orientation, you can correct your mistake without deleting the arrow and redrawing it. Instead select the arrow and click the desired arrow button on the *StoTree* tab. *StoTree* will replace the mistaken arrow with the corrected arrow.

5. You can move labels on arrows from one side of the arrow to the other, or slide them along the arrow. Clicking the Refresh button will leave the label on the side of the arrow on which *StoTree* finds it.

6. You may wish to move or delete a group of nodes and arrows at once. You can do so by activating Excel’s Select Objects pointer (Home/Editing/Find & Select). You can use this pointer to select multiple graphics at once by dragging it around the objects. Once a group of nodes and arrows are selected, you can drag them to any desired location or delete them. If you wish to access Excel’s Select Objects pointer more quickly, you can add it to the Quick Access Toolbar (File/Options/Quick Access Toolbar).

7. Do not attempt to use Excel’s Copy/Paste tools to create new nodes or arrows. This will only result in new worksheet graphics that *StoTree* will not recognize as being part of your model. The reason is that *StoTree* maintains hidden databases of nodes and arcs, and these databases will not be updated by Excel’s Copy/Paste tools. You must use the *StoTree* tab on the Office ribbon to create new nodes or arrows. For similar reasons, do not attempt to create a new factor (see below for a discussion of factors) by having Excel copy a worksheet.

---

**QALYs and costs**

You can enter quality accrual rates (*quality coefficients* or so-called *utilities*) and cost accrual rates for any stochastic node in your tree. (Note chance and decision nodes have instantaneous impact, so quality and cost accrual does not apply to them.) Simply select the node and then click on the Parameters button on the *StoTree* tab. A userform appears into which you may enter these values for the selected node. You may enter quality tolls and cost tolls for any arrow in your model – again simply select part of the arrow and then click on the Parameters button.

---

**Sharing your constructed workbook**

Once you have used the *StoTree* addin to construct a Markov model in your workbook, and have saved your workbook as a macro-enabled workbook (.xlsm), you may wish to share your workbook with colleagues. Because it is an Excel workbook, anyone with Microsoft Office can view it. However the graphic objects and cohort-sheet formulas in your workbook contain references to macros in *your* copy of the *StoTree* addin. When a user on another computer recalculates the cohort sheet, or clicks on a graphic,
error messages will result – the macro references in her copy of your workbook will not be able to find your copy of the *StoTree* addin.

However, if the user has a licensed or unlicensed copy of the *StoTree* addin installed, she can convert her copy of your workbook to access her copy of *StoTree*. She does this by opening her copy of your workbook and using the key combination Ctrl+Shift+c. A successful conversion will result in a confirmation message similar to the one illustrated in Figure 7.

![Figure 7. To convert another user’s workbook to a form compatible with your copy of Microsoft Office, you should use the key combination Ctrl+Shift+c. This figure illustrates the confirmation message displayed when conversion is complete.](image)

After conversion, a user with a licensed copy of *StoTree* has complete access to the structural details of your workbook, and may add to or modify its contents in whatever way she desires, for instance by adding or deleting nodes or arcs, setting up a revised cohort sheet, and performing sensitivity analyses. A user with an unlicensed version of *StoTree* has only limited access to the computational features of *StoTree*. Such a user does not have access to the calculation features on the *StoTree* menu system. An unlicensed user *does*, however, have access to the calculation buttons on your workbook’s cohort sheet, and may use these to perform what-if analysis on your model – again, see the section Defined names for documenting and sensitivity below for more information on how to do what-if analyses.

**Top**

**Defined names for documenting and sensitivity**

You can enter any numeric value or expression (e.g., 22.3, 1/10, 12+3) into the Parameters userform. However, to properly document your model and make it more transparent and flexible, you are strongly advised to avoid numeric expressions, and instead use Excel’s defined names feature (on the Formulas tab) to provide names for cells containing your input data. Then use these names, or expressions containing these names, in *StoTree*’s Parameter userform when you populate parameters for your model. *StoTree* displays rate and probability expressions next to the appropriate arrows, so using defined names makes it easy to see to what arrows these inputs apply. See for instance, Figure 8. This is a method you can use to document your model.
A second advantage of using Excel’s defined names in the Parameter userform is that you may change the values in the named cells, and then Calculate the cohort sheet to obtain new expected QALYs and costs, without having to request StoTree to setup the cohort sheet again. As mentioned above, someone inspecting your workbook can do this even if his or her computer possesses only an unlicensed copy of the StoTree addin (e.g., perhaps you have sent others a copy of the workbook you have constructed). You or others can thereby quickly perform a manual sensitivity analysis by repeatedly changing values in named cells and clicking on Calculate Sheet. Note, however, that StoTree has automated this process via the One-Way Sensitivity button on the StoTree tab.

![Figure 8. The use of Excel’s defined names in StoTree. The user has given cell C5 the name rPE, and cell C6 the name pDEmb.](image)

**Top**

**One-way sensitivity analysis**

By clicking on the One-Way Sensitivity button on the StoTree tab, you can obtain instructions in how to set up a one-way sensitivity analysis on any input parameter that you have stored in a cell of your workbook. You proceed by setting up one of Excel’s column data tables, then selecting its corner cell, and finally pressing Ctrl+Shift+s. Details are provided by clicking on the One-Way Sensitivity button.

**Top**

**Time-dependent parameters**

As mentioned above, you may enter any valid Excel formula or expression containing defined names into StoTree’s Parameter userform. There is an additional defined name _t that StoTree automatically creates when you instruct it to setup a cohort sheet. The name _t represents the elapsed time in the cohort sheet, and you may use this name in any of the formulas or expressions you enter into the Parameter userform. You would most commonly use _t in a rate formula when you desire the rate to depend on elapsed time in your cohort analysis. Figure 9 shows an example. However, _t may be used in any parameter expression, for example for quality or cost coefficients. Figure 10 illustrates how you might want to use _t in a quality coefficient formula.
Figure 9. You can use the defined name _t to represent elapsed time in parameter expressions. Here, in a model of the outcome of hip replacement, the rate out of Success Primary is the Weibull hazard rate \( h(t) = \lambda t^{\gamma-1} \).

Figure 10. You can use the elapsed-time defined name _t in a quality coefficient. Here the node No Embolism has zero quality after 5 years, and quality 1 before that. The resulting cohort analysis (with initial cohort set to 1.0) will calculate the 5-year survival probability.

Top

Human mortality rates

StoTree includes a function GmpHR that returns time-dependent human mortality rates for any age, gender and race. GmpHR stands for “Gompertz hazard rate” and is based on a Gompertz approximation to human mortality rates. You can employ this function in conjunction with the current time variable _t to supply the correct human time-dependent mortality at any clock time in a cohort analysis – see for example Figure 11. The Gompertz approximation to human mortality is valid for ages 30 or greater. See the appendix Gompertz hazard rate for human survival for documentation on this function.
Constant intervals of clock time

Occasionally you may wish a state transition to occur repeatedly at time intervals of length \( t_{inc} \). In essence this means that the associated transition rate is infinite every \( t_{inc} \) units of clock time and zero for other clock times. The \textit{StoTree} function \texttt{Spike()} produces this effect (the name indicating that the transition rate “spikes” to infinity every \( t_{inc} \) units of time). In cohort analysis, \textit{StoTree} steps off time using the Excel defined name \_t in units of the Excel defined name \_dT, which \textit{StoTree} sets equal to the cycle time you specify in the calculation settings. The function call \texttt{Spike(\_tInc)} returns the rate \( 1/\_dT \) (effectively an infinite rate) when \_t is within \_dT/2 of \( k \cdot t_{inc} \) for any integer \( k \). Figure 12 illustrates the use of \texttt{Spike()} in a model of testing at multiples of time increment \( t_{inc} \).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure12.png}
\caption{The use of the \textit{StoTree} function \texttt{Spike()} in a model of testing at multiples of time increment \( t_{inc} \). Here the transition rate from \textit{Wait} to \textit{Test} is zero unless \_t is close to some multiple of \texttt{tInc}, in which case the rate is “infinite” (equal to the inverse cycle time \( 1/\_dT \)). That is, \texttt{Spike(\_tInc)} produces a transition every \texttt{tInc} units of time.}
\end{figure}

\texttt{Spike()} also has an optional parameter \( t_0 \) that specifies the start time for spikes. The function call \texttt{Spike(\_tInc,t0)} produces the first transition spike at time \( t_0 \), and successive spikes at \( t_0+k \cdot t_{inc} \) for every positive integer \( k \). If \( t_0 \) is omitted (as above), then its value is assumed to be zero.

See the appendix for detailed documentation on \texttt{Spike()}.

Clock time versus sojourn time

You may at times wish to have a rate out of some state depend not on the clock time \_t but on the so-called \textit{sojourn time} in a state – the time elapsed since the state was entered. Unfortunately, Markov
models as usually formulated do not explicitly include a sojourn-time variable, except in the special case in which sojourn time and clock time coincide, which happens only for states occupied initially at time 0.

The distinction can be seen in Figure 9, where the failure rate of the initially implanted artificial hip depends on its age, which is equal to the sojourn time in the Success Primary state. In that state, age is equal to clock time \( t \) because a new hip is implanted at time 0. Therefore the age-dependent failure rate can be made a Weibull function of \( t \). On the other hand, in the state Success Revision, age is never equal to clock time because the replacement hip is never implanted (if it is at all) at time 0. The modeler would like to make the failure rate of the replacement hip depend on its age, but cannot because there is no sojourn time variable available. As an approximation, the modeler has in this case entered a constant failure rate \( rrr \) for the replacement hip.

This approximation may be adequate in some cases, but if it is not, then the modeler must resort to using what are known as tunnel states, a sequence of states that mimics elapsed sojourn time. StoTree can handle tunnel states, but we do not discuss the method here.

**Top**

**Decision nodes**

In a stochastic tree, a decision node is meant to indicate that an individual has a choice among the branches leading from that node. However, in StoTree, decision nodes are treated the same as chance nodes. You the user would assign probability one to the branch corresponding to whatever decision is to be analyzed, and probability zero to other branches. StoTree does not perform any kind of optimization among decision branches – it merely evaluates QALYs and/or costs associated with the desired decision branch. To evaluate a different decision, you would have to change its branch probability to one. This can be automated using Excel’s defined names, as is illustrated in Figure 13.
Figure 13. You can use a defined name (here, Decn) to specify which decision branch is to be selected. Change the value of the Decn cell and press Calculate Sheet to evaluate a different decision. You can also use StoTree’s One-Way Sensitivity command to perform a sensitivity analysis on Decn, and thereby evaluate all decision alternatives in one step.

**Terminal nodes**

Death states are common in medical decision analyses. They can be represented in StoTree as stochastic nodes with zero quality and no emanating arrows. However, it is computationally more efficient to use the StoTree node type terminal to represent death states. When you create a terminal node, it appears shaded in your spreadsheet (see several of the preceding figures). You can change an existing stochastic node to terminal in the Parameters userform if the node has no successors.

**Root nodes**

StoTree begins its cohort analysis by assuming all cohort members occupy the root node in the stochastic tree. You can set which node in your tree is the root node using a checkbox in the Parameters userform for that node. If you neglect to do so, StoTree will make an educated guess about which node is the root node in your tree. However, for some types of trees, there may be more than one plausible root node, and you should make sure StoTree has selected the one you desire.
Factored stochastic trees

Often in constructing models of medical prognosis, one encounters processes that evolve in parallel. Consider for example, the pulmonary embolism model depicted in Figure 1. Tsevat et al.[7] consider the use of Warfarin to reduce the incidence of pulmonary embolism. However, Warfarin can increase the incidence of systemic hemorrhage, a process depicted in Figure 14. The pulmonary embolism and systemic hemorrhage processes are proceeding in parallel – both processes evolve simultaneously. It is possible, but cumbersome, to formulate a single Markov model or stochastic tree that depicts both processes. Fortunately, this is not necessary in StoTree. One of the most important features of StoTree is that it allows the modeler to formulate parallel processes as inputs. It also allows the modeler to specify links between the processes, so that what happens in one process can influence what happens in others, and can in turn be influenced by others. Parallel processes are known as factors, and the overall combination of processes is called a factored stochastic tree.

![Diagram of factored stochastic tree process](image)

*Figure 14. A model of systemic hemorrhage. This process evolves in parallel with the pulmonary embolism process depicted in Figure 1.*

**Using StoTree to construct a factored stochastic tree**

How do you construct a factored stochastic tree? The procedure in StoTree is quite simple: Click on the Insert Factor button on the StoTree tab for each factor you wish to create. StoTree will insert new worksheets into your workbook in which you can construct new stochastic factors in the same way as described in the sections above. Once you have created multiple worksheets, each containing its own factor, StoTree allows you to link what happens in one factor to what happens in others using dependencies and triggers. We discuss each of these in turn.

**Dependencies**

In StoTree, a dependency indicates how parameters in one factor depend on the state or states of other factors. For example, in the StoTree model with factors depicted in Figure 1 and Figure 14, both the rate of embolism and the rate of hemorrhage should depend on whether Warfarin is being administered. The decision on whether to do so is shown in Figure 15, which depicts a third factor with an initial decision node selecting between Warfarin and No Warfarin.
Once the user has created the *Warfarin* factor, s/he can make the hemorrhage rate in the *Systemic Hemorrhage* factor depend on whether Warfarin is being administered. The user activates the *Systemic Hemorrhage* factor, clicks on the arrow out of *No Hemorrhage*, and proceeds as in Figure 16, Figure 17, Figure 18, and Figure 19 below.

**Figure 15.** The *Warfarin* factor, depicting the decision of whether to administer Warfarin.

**Figure 16.** Creating a dependency in the *Systemic Hemorrhage* factor. The user has selected the arrow from No Hemorrhage and clicked on the Parameters button to bring up the userform pictured here. Next the user has clicked on the dropdown arrow in the Rate combobox to reveal the option (more).

**Figure 17.** The user has clicked on the (more) entry in the Rate combobox dropdown. As a result, the ===> button next to the combo box has been activated. The user will next click on this button to bring up the Dependency userform.
Figure 18. The Dependency userform. The user has selected Warfarin from the If Factor dropdown, and indicated that the chosen rate will equal $r_{HemW}$ if the Warfarin factor is in state Warfarin, and will equal $r_{HemNo}$ if the Warfarin factor is in state No Warfarin. The user will next click on Back to return to the Parameters userform.

Figure 19. The process is complete. The arrow out of No Hemorrhage is labeled with a partial description of the dependency just entered.

Top

Triggers

Sometimes it is desirable to have a transition in one factor produce a state change in another factor. The user can arrange this by clicking on the Trigger button on the StoTree tab. Consider once more the Tsevat model of Warfarin treatment. If a hemorrhage occurs while the patient is on Warfarin, it would seem natural to discontinue Warfarin for some period of time. In other words, the occurrence of a hemorrhage should trigger transition from Warfarin to No Warfarin in the treatment factor. To make this a temporary change, the user can add a Temporarily Discontinue node to the Treatment factor, as shown in Figure 20. The user has also added an arrow from Temporarily Discontinue back to Warfarin, with an associated rate of 12/yr (resume Warfarin in an average of one month). Finally, the user needs to add a trigger that sends the Warfarin factor from Warfarin to Temporarily Discontinue. This is shown in Figure 21.

Note that the dependencies that the user has added to the Pulmonary Embolism and Systemic Hemorrhage factors – see Figure 18 – do not account for the new state Temporarily Discontinue. The user must revisit these dependencies and add this new state to them.
**Top Priorities**

In a multifactor tree, there can be many combinations of health states. Consider a particular health-state combination. There may be multiple ways to exit this combination health state, corresponding to multiple exit arrows across factors. If a combination state contains decision nodes, *StoTree* executes arrows out of these first. If it contains chance nodes, *StoTree* executes arrows out of these next. Finally, *StoTree* executes arrows out of stochastic nodes. If there are ties – for example, decision nodes in two different factors – *StoTree* breaks the tie by executing the arrow from the alphanumerically first factor first.

A user may enter several triggers on an arrow. *StoTree* executes these in the order entered, which is the order listed in the trigger userform.

Triggers may also be attached to nodes. *StoTree* executes such a trigger *just after* the node is entered.

**The cohort sheet for a factored tree**

It is important to remember that in a factored stochastic tree, the factors depict processes that are proceeding in parallel. So *any combination of states*, one from each factor, may in principle occur. *StoTree* accounts for this in constructing the cohort sheet, and lists many such state combinations, as long as they can be reached from the combination root node. Figure 22 illustrates this for the example tree of this section.
You may note that the three factor trees in this model contain respectively 3, 5, and 5 states. Yet there are nowhere near $3 \times 5 \times 5 = 75$ combination states in the cohort sheet of Figure 22. *StoTree* exploits known properties of stochastic processes to bypass the portrayal of terminal (death) nodes, chance nodes, and decision nodes (unless they are root nodes). Only stochastic nodes and root chance/decision nodes need be portrayed on the cohort sheet. This is the primary reason you should use a terminal node rather than a stochastic node to model the death state—a stochastic-node death state will add to the number of combinations that need display, whereas a terminal node will not. The three factors in this model contain respectively 1, 2, and 4 stochastic or root nodes, producing $1 \times 2 \times 4 = 8$ combinations. The cohort sheet contains only 7 combination states because one combination state cannot be reached from the combination root node.

**Top**

**Active and Inactive factors**

*StoTree* will form the cohort sheet using only combinations of states from *active* factors. You can use the active/inactive toggle button in each factor worksheet to change a factor from active to inactive and vice versa. This toggle button lies in the upper left corner of each factor worksheet (for example, see Figure 8).

You might, for example, wish to deactivate a factor rather than deleting it if you think you might use it later and do not wish to have to reconstruct it.

**Top**

**Quality and cost for factored trees**

In order to perform cohort analysis for a factored stochastic tree, *StoTree* must assign quality coefficients and cost rates to combinations of stochastic nodes. The rule *StoTree* uses to accomplish this is that the cost rate for a combination is the sum of the cost rates in each factor, and the quality coefficient for a combination is the product of the quality coefficients in each factor.

**Top**

**Referencing defined names instead of creating dependencies**

As an alternate to creating a dependency, you may also take advantage of the fact that when constructing the cohort sheet, *StoTree* creates defined names that are the same as the factor worksheet names. Notice in Figure 22 that the factor names *Pul_Emb*, *Sys_Hem*, and *Warfarin* appear near the top of the cohort.
sheet. You may reference these names in any parameter formula. In particular, you may use Excel’s if() function to create a parameter formula that has the same effect as a StoTree dependency. See Figure 23.

Figure 23. Bypassing StoTree’s dependency feature by using Excel’s if() function and the defined name Warfarin. StoTree creates the defined name Warfarin because the user has created a factor worksheet with that name. The user has entered the formula in red as a rate into the Parameter userform for the arrow out of No Hemorrhage. Compare with Figure 19.

**Top**

**Tips and tricks for constructing factored stochastic trees**

**Use an additional factor instead of duplicate subtrees**

As mentioned above, StoTree has Copy Subtree/ Paste subtree features for easily duplicating subtrees in any particular factor. A disadvantage of doing so, however, is that copy/paste can produce large bushy trees. For example, in Figure 24, the user has duplicated the Test subtree, because periodic screening applies to both the healthy and diseased populations. A more elegant approach is to create a separate factor depicting the outcomes of testing, and use dependencies to link its parameters back to health and disease. This approach is illustrated in Figure 25.

```
rHemW = 0.45 /yr
rHemNo = 0
```
The advantage is a simpler and more elegant model formulation. However, the factored approach does require the use of dependencies at the Test branches and the Positive branches (see Figure 25). The resulting model details are thereby obscured. Whether to factor in this way is largely a matter of taste – computationally there are no advantages. Had the parameters on the Test and Positive branches been independent of health or disease, dependencies would not be necessary and the factored approach would be superior. The general rule is to consider factoring when you find yourself duplicating subtrees.
Figure 25. A factored approach to model periodic screening. Compare to Figure 24. The Test subtree has been split off into its own factor.

Dependencies may only reference stochastic nodes

You may notice that the dependency userform only allows you to access stochastic nodes in each factor. Recall that chance nodes are instantaneous impact, and have zero duration. They therefore cannot influence rates, probabilities, QALYs or costs in other factors.

**Top**

Use a negative rate to indicate an absent arrow

As we have seen, *StoTree* uses dependencies to allow the transition rate or probability along an arrow to depend on the state of another factor. However, at times you may want the *existence* or the *possibility* of a transition to depend on another factor’s state. It may be tempting to simply use a zero rate or probability in the dependency userform to indicate that transition is impossible for some states of the influencing factor. However, if you do this, *StoTree* will include these transitions in the cohort sheet with zero rate or probability attached. This is inefficient because the cohort sheet will then include impossible state combinations, and multiplications by zero. To get *StoTree* to build a cohort sheet that omits the impossible transitions for some states in the other factor, simply enter a *negative* rate or probability into the dependency userform for those states.

For instance, in Figure 25 above, consider the transitions out of the chance node *Positive* in the testing factor and its two possible successors *Detected* and *False*. Here a positive test has occurred. If the disease factor is *Disease*, then transition from *Positive* must be to *Detected*, and cannot be to *False*. Similarly, if the disease factor is *Healthy*, then transition must be to *False*, and cannot be to *Detected*. The user has identified these impossible transitions by entering $-1$ appropriately in the dependencies along the arrows out of *Positive*—see Figure 26. Had the user instead entered values of 0, *StoTree* would have had to include the impossible state combinations (*Healthy, Detected*) and (*Disease, False*) in the cohort sheet it generates.
Figure 26. The dependency along the arrow from Positive to Detected in the testing factor from Figure 25. The \(-1\) entry in this dependency indicates that if the disease factor is Healthy, then transition from Positive to Detected is impossible. It would be acceptable to enter the value 0 here as well, but this would have resulted in the impossible state combination (Healthy, Detected) in the generated cohort sheet, and an inefficient multiplication by zero there.

The benefits of factoring

The ability to handle multiple factors is one of the primary strengths of _StoTree_. Formulating factored stochastic trees has several advantages:

- It simplifies the model formulation process. The analyst can focus attention on one component (factor) of the model at a time, refining it as needed, and only later worry about other components and interactions between components.
- It allows the easy formulation of large Markov models. The models shown in this document are small examples. In our experience, models containing tens or hundreds of states may easily arise. These are readily formulated in factored form because most factors contain on the order of 2–5 states. Formulating large models such as these is impractical without factoring.
- It assists in the presentation of the model to others. Even those not versed in the theory of continuous-time Markov decision processes can gain an intuitive graphical understanding of simple 2- to 5-state factors of most models.
- It allows the easy swapping or adding of model components. The _StoTree_ software allows the analyst to add, remove, or substitute model components (factors) as needed. This opens up the possibility that model components formulated and populated by other analysts may be imported, or that repositories of model components may be created for the community of modelers.
Appendix
Built-in functions
Gompertz hazard rate for human survival
It is known that for ages 30 or more, human mortality rates can be closely approximated by the Gompertz mortality rate function
\[ \mu(t) = b_1 \cdot \exp(b_0 + b_1 t) \]
with values of \( b_0, b_1 \) depending on age, gender and race. StoTree provides a VBA function GmpHR that returns these mortality rates for different ages (\( \geq 30 \)), genders (Male or Female), and races (White or Black). We provide documentation of this function below.

'See my Excel workbook 'Deaths 2005 by Age,Race,Sex'.xls.
'
Function GmpHR(Age As Double, Gender As String, Optional Race As String = "White") As Double
Dim b0 As Double, b1 As Double
If Race = "White" Then
 If Gender = "Male" Then
   b0 = -6.707225204
   b1 = 0.079711236
 ElseIf Gender = "Female" Then
   b0 = -7.709089201
   b1 = 0.086873918
 End If
ElseIf Race = "Black" Then
 If Gender = "Male" Then
   b0 = -5.631988662
   b1 = 0.071532881
 ElseIf Gender = "Female" Then
   b0 = -6.586692612
   b1 = 0.077624643
 End If
End If
If Age < 0 Then
  GmpHR = 0
Else
  GmpHR = b1 * Exp(b0 + b1 * Age)
  If GmpHR > 1E+100 Then GmpHR = 1E+100 'To prevent overflow issues in Excel
End If
End Function

Top
Spike function

'Returns 1/dT if t-t0 is closest to some multiple of tInc, and zero otherwise. By 'closest is meant within dT/2. In other words, returns 1/dT when t is closest to 't0+k*tInc for any integer k (positive, negative or zero); and zero otherwise. 'Extracts t and dT from names on the cohort sheet. t0 is an optional parameter that 'takes value zero if it is omitted.

Function Spike_(tInc As Double, Optional t0 As Double = 0) As Double
    Dim M As Double, t As Double, dT As Double
    'Extract values of _t and _dT from cohort sheet
    t = Evaluate(ActiveWorkbook.Worksheets(CohortSheetName).Names("_t").Value)
    dT = Evaluate(ActiveWorkbook.Worksheets(CohortSheetName).Names("_dT").Value)
    'Calculate elapsed time module tInc, and spike if it is close to zero.
    M = ModC(t - t0, tInc)
    If M >= -dT / 2 And M < dT / 2 Then
        Spike_ = 1 / dT
    End If
End Function

Top

Modulo function

'Due to the binary representation of floating point numbers in Excel, Excel's Mod function can return erroneous results. For example, Mod(.03,.01) returns .01 in Excel, not 0. This function corrects the result to 0 if mod/Modulus is within dM of 1.

Function ModC(Argument As Double, Modulus As Double) As Double
    Dim dM As Double
    dM = 10 ^ (-12)
    ModC = Argument - Modulus * Int(Argument / Modulus)
    If ModC / Modulus > 1 - dM Then ModC = 0
End Function

Weibull hazard rate function

The Weibull is a commonly used distribution for modeling survival. With parameters \( \lambda > 0 \) and \( \gamma > 0 \), it’s hazard rate function is

\[
h(t) = \lambda \gamma t^{\gamma-1}, \quad t > 0.\]

The corresponding survival function is

\[
S(t) = P(T > t) = \exp(-\lambda t^\gamma), \quad t > 0.
\]

'Returns the Weibull hazard rate at time T.
Function WeiHR(t As Double, Lambda As Double, Gamma As Double) As Variant
If Gamma > 0 And Lambda >= 0 Then
    If t > 0 Then
        WeiHR = Lambda * Gamma * t ^ (Gamma - 1)
    ElseIf t = 0 Then
        If Gamma >= 1 Then
            WeiHR = 0
        Else
            WeiHR = "Error"
        End If
    Else
        WeiHR = "Error"
    End If
Else
    WeiHR = "Error"
End If
If IsNumeric(WeiHR) Then
    If WeiHR > 1E+100 Then WeiHR = 1E+100 'To prevent overflow issues in Excel
End If
End Function

'Returns the alpha-percentile of the Weibull distribution with parameters Lbda and Gamma.
'If impossible parameter Gamma, returns -1.
'
Function WeiPct(Alpha As Double, Lbda As Double, Gamma As Double) As Double
    If Gamma > 0 Then
        WeiPct = (-1 / Lbda * Log(1 - Alpha)) ^ (1 / Gamma)
    Else
        WeiPct = -1
    End If
End Function
References