Charles Sturt University

Quantitative Consulting Unit

# An Introduction Presentation for Learning Bayesian Networks by hands-on practice with examples

Dr. John Xie

Statistics Support Officer, Quantitative Consulting Unit,

Research Office, Charles Sturt University, NSW, Australia

Website: http://www.csu.edu.au/qcu

CHARLES STURT UNIVERSITY

# Presentation Outline

- Concepts, definitions, and examples about what is a Bayesian Network (BN) and what can BN models do for us

- Principles of data analysis and the theoretical basics of BN.

- Demonstration of various applications with a number of BN models.

- An overview of a 3-day workshop on **Learning BNs by Hands-on Practice with Examples**.



# A Contingency Table and Probabilities

**Computer brand and hard-disc memory**

| | BrandA | BrandB | BrandC | row sum |
|---|---|---|---|---|
| **≤ 500GB** | 55 | 8 | 10 | 73 |
| **>500GB** | 70 | 32 | 25 | 127 |
| Column sum | 125 | 40 | 35 | ( 200 ) |

( grand sum)

**Probability ≈ Relative Frequency** $= \dfrac{Number\ of\ times\ an\ event\ occurred}{Number\ of\ trials\ performed}$

# A Contingency Table and Probabilities

Joint probability distribution

**Computer brand and hard-disc memory**

| | **BrandA** | **BrandB** | **BrandC** | **row variable** |
|---|---|---|---|---|
| **≤ 500GB** | 27.5% | 4% | 5% | 36.5% |
| **>500GB** | 35% | 16% | 12.5% | 63.5% |
| **Col. variable** | 62.5% | 20% | 17.5% | ( 100% ) |

Marginal probability distribution for **Brand**

Marginal probability distribution for **Hard-Disc Memory**

# Marginal probability distribution, joint probability distribution, and conditional probability distribution

Conditional probability distributions:

$Pr(\text{memory} \leq 500GB \mid \text{BrandA}) = 27.5/62.5 = 0.44$

$Pr(\text{memory} > 500GB \mid \text{BrandA}) = 35/62.5 = 0.56$

$Pr(\text{BrandA} \mid \text{memory} \leq 500GB) = 27.5/36.5 = 0.753$

$Pr(\text{BrandB} \mid \text{memory} \leq 500GB) = 4/36.5 = 0.110$

$Pr(\text{BrandC} \mid \text{memory} \leq 500GB) = 5/36.5 = 0.137$

# Conditional Probability

$$\Pr(A|B) = \frac{\Pr(A \text{ and } B)}{\Pr(B)} \qquad \Pr(B|A) = \frac{\Pr(A \text{ and } B)}{\Pr(A)}$$

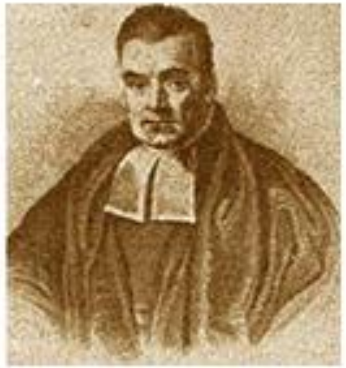
$$\Pr(\text{A and B}) = \boxed{\Pr(B|A)\Pr(A) = \Pr(A|B)\Pr(B)}$$

$$\Downarrow$$

$$\Pr(B|A) = \frac{\Pr(A|B)\Pr(B)}{\Pr(A)}$$

# Bayes Theorem



**Bayes Theorem / Bayes Rule** is named after Reverend Thomas Bayes (1701-1761) which is stated mathematically as:

$$\Pr(B|A) = \frac{\Pr(A|B)\Pr(B)}{\Pr(A)} = \frac{\Pr(A\ and\ B)}{\Pr(A)}$$

Define: event T = test outcome (+ or -); event D = cancerX (true or false).

Known: $\Pr(\mathrm{D} = true) = 0.01$

$\Pr(\mathrm{T} = +|\mathrm{D} = \mathrm{true}) = 0.8; \Pr(\mathrm{T} = -|\mathrm{D} = false) = 0.9.$

therefore, $\Pr(\mathrm{T} = +) = 0.107 (= 0.8 * 0.01 + 0.1 * 0.99)$

We have: $\Pr(\mathrm{D} = \mathrm{true}|\mathrm{T} = +) = \frac{\Pr(\mathrm{D}\ and\ \mathrm{T})}{\Pr(\mathrm{T})} = \frac{\Pr(\mathrm{T}|\mathrm{D})\Pr(\mathrm{D})}{\Pr(\mathrm{T})} = \frac{0.8*0.01}{0.107} = 0.0748$

Thomas Bayes' picture is downloaded from internet website (accessed on 2 August 2016):
https://www.bing.com/images/search?q=Thomas+Bayes&view=detailv2&id=C100E7A4DA7B74569545884EC8CF873EA3CE078D&selectedindex=22&ccid=D7975i%2B4&simid=608043103954862875&thid=OIP.M0fbf7be62fb83f4e45cbd958a9e4dd86o2&mode=overlay&first=1

# The mini-history of the development of Bayesian Networks:

**Bayesian networks** (BNs) have established themselves as the basis for a new generation of **probabilistic expert systems**, which allow for effective modelling of physical, biological and social systems operating under uncertainty. Originally developed as a modelling tool from artificial intelligence since late 1980s, today BNs have found their applications range across the sciences, industries and government organizations. Although the theoretical foundation and computational algorithms underlying BNs are highly involved in subjects such as computer science, mathematics and statistics, the applications of BN models are very intuitive and relatively straightforward because of the availability of many well tested BN application software packages.

# Demonstration of Example BN Models using Netica

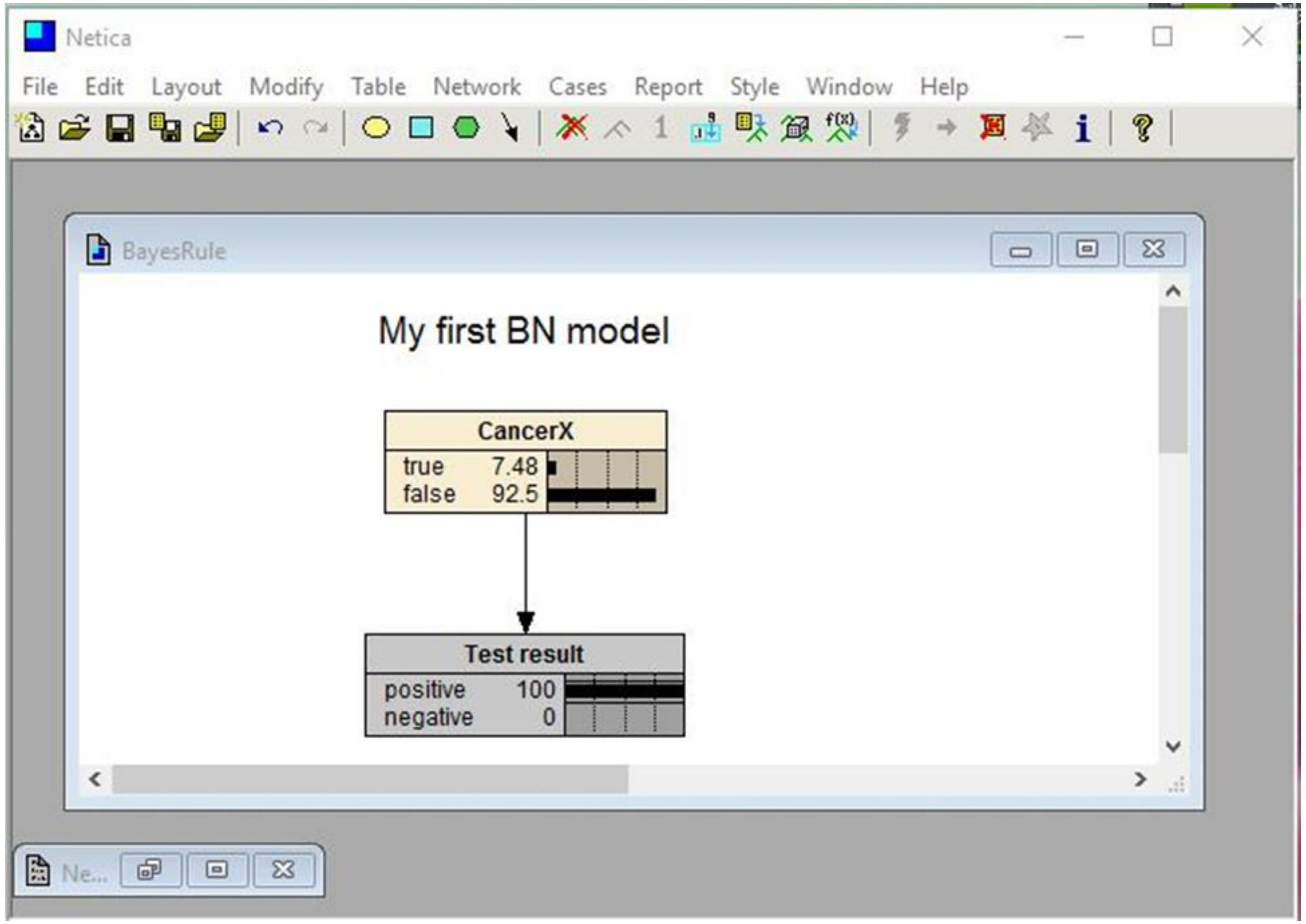
- Total 10 example BN models will be demonstrated using Netica.
- **Netica** is believed to be the most widely used Bayesian Network application/development software which is the product of Norsys software corp. (https://www.norsys.com/).

# Example 1:
# the Bayes Rule model

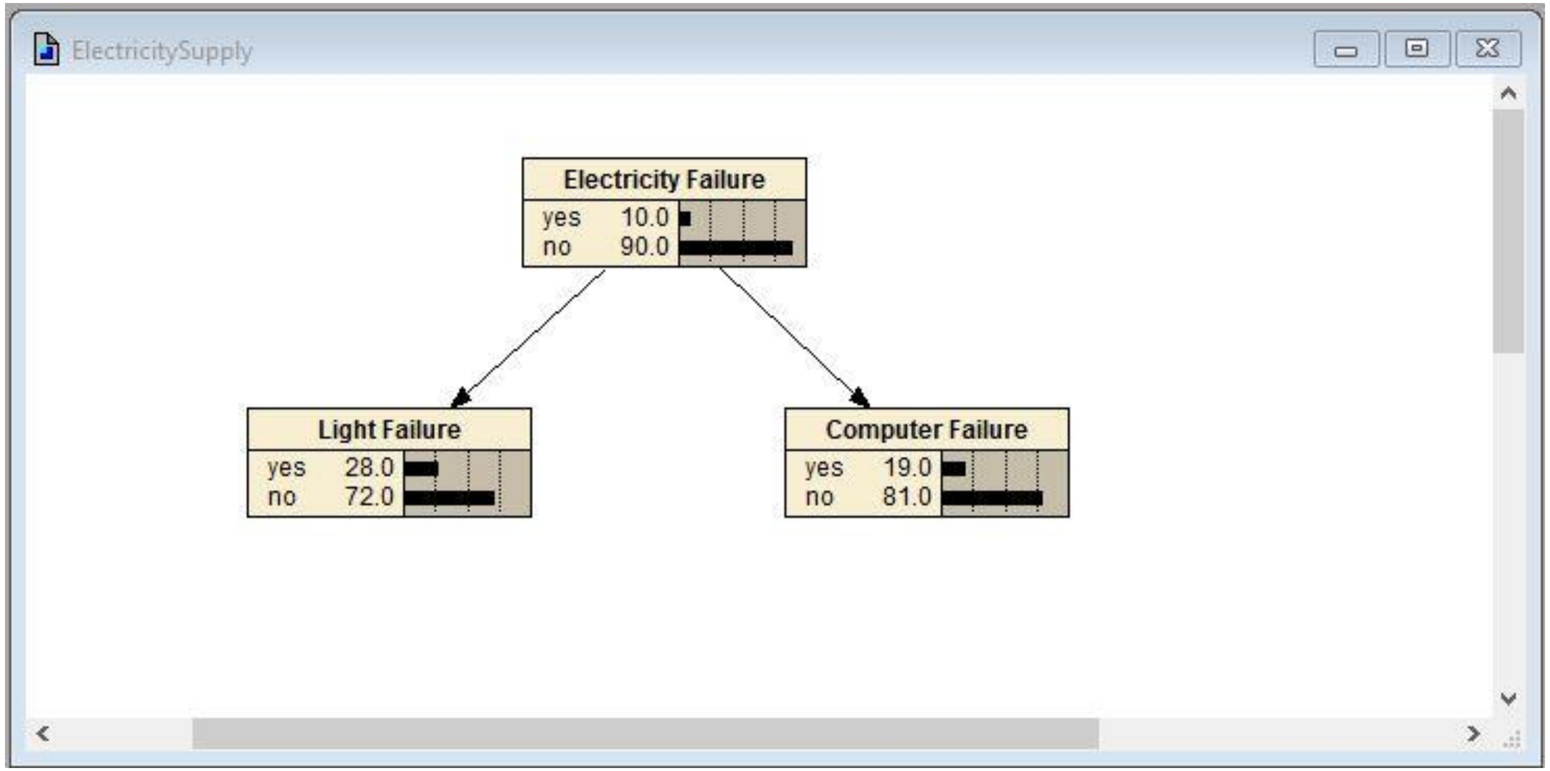
**The conditional probability**

P(CancerX = true|Test result = positive)  can be obtained by simply selecting the ‘positive’ state in ‘Test result’ (i.e., clicking on ‘positive’) and the answer is 7.48 % as immediately appearing in the ‘CancerX’ node.




This example is adapted from Online material: Bayesian AI Bayesian Artificial Intelligence: Introduction. IEEE Computational Intelligence Society, IEEE Computer Society. Kevin Korb, Clayton School of IT, Monash University; kbkorb@gmail.com. Downloaded from http://abnms.org/resources.php on 03 August 2016.

# Example 2: A three-node Bayes Net model




This example is adapted from Figure 4 in Technical report: Bayesian Networks. Michal Horný, 2014. Technical report No. 5, Department of Health Policy & Management, School of Public Health, Boston University.

# Example 3: A four-node Naïve Bayes model

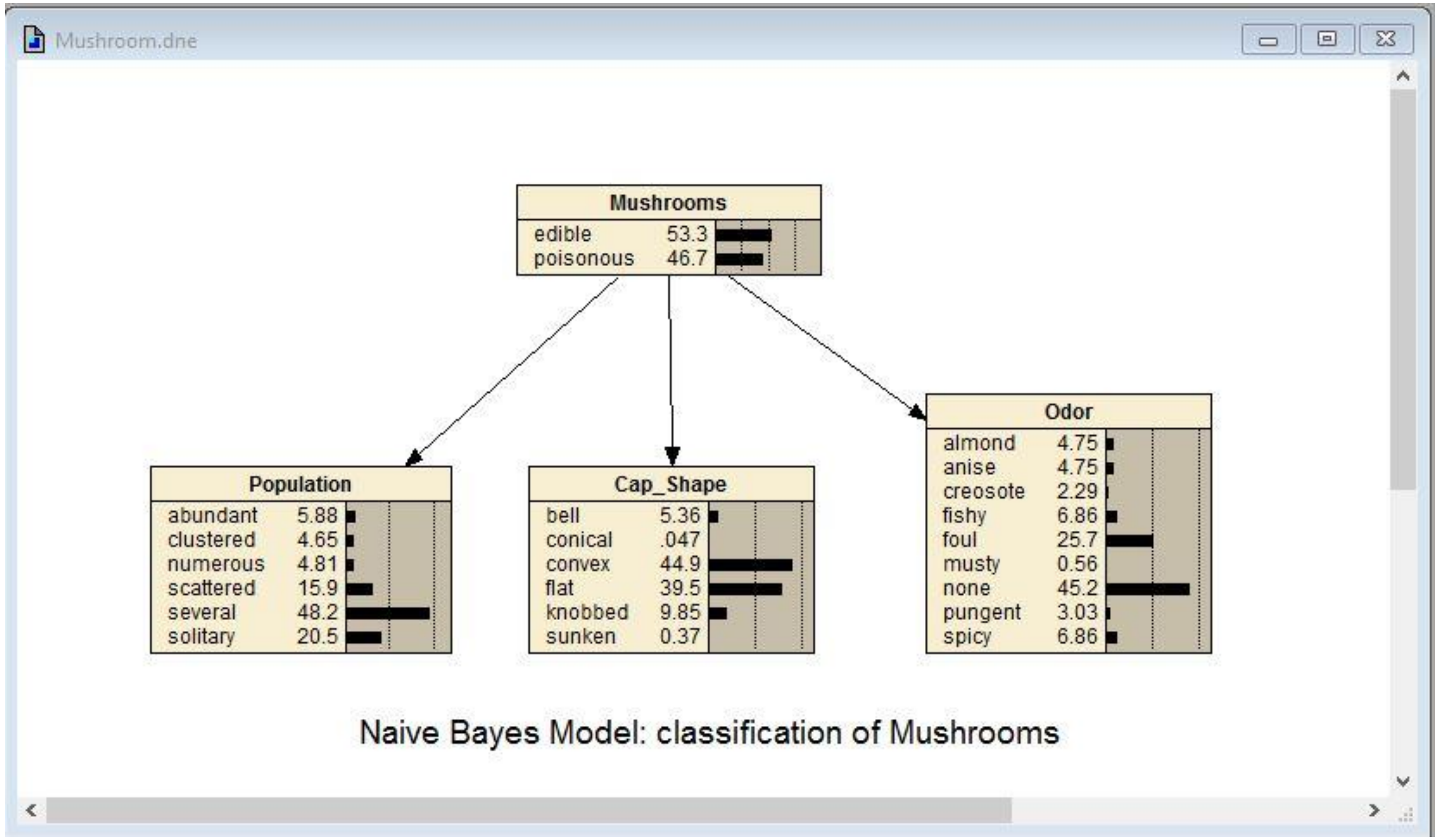
The **Naïve Bayes model** is particularly suitable for modelling classification related problems.

**Background information of the problem:**
Assume we plan to pick up mushrooms for some purposes (e.g., to prepare for a nice dinner or for research). We want to construct a BN model for classifying each mushroom based on a database of mushrooms.

| Class | n |
| --- | --- |
| edible | 5333 |
| poisonous | 4667 |

# Example 3: A four-node Naïve Bayes model




This example is adapted from Exercise 8.4 in eBook: Bayesian Networks and Influence Diagrams: a guide to construction and analysis. Uffe B. Kjaerulff & Anders L. Madsen, 2008. Springer.

**Bayesian Networks are so called because**

(1) it is a network type of graphical models;

(2) the Bayes Theorem / Bayes Rule is the core theoretical foundation underlies the Bayesian Networks.

# Types of nodes available and the corresponding acceptable data types in Netica

| Types of nodes | data types allowed |
|---|---|
| Nature | discrete / continuous |
| Decision | discrete / continuous |
| Utility | continuous |
| Constant | discrete / continuous |

# Bayesian Networks / Bayesian Belief Networks maybe categorised into three groups:

1. **Bayes nets**: BN models consist only the Nature Nodes for reasoning under uncertainty.
2. **Decision nets / influence diagrams**: BN models consist Nature Nodes, and at least one Decision Node and one Utility Node for decision making under uncertainty.
3. **Dynamic Bayesian Networks (DBNs)**: (a.k.a. time-sliced Bayesian Network) BN models (Bayes nets or Decision nets) with at least one "time delay" link for modelling feedback and/or time series processes.

# What can Bayesian Networks do for us ?

- Diagnosis
- Prediction
- Financial risk management, portfolio allocation, insurance
- Modelling ecosystems
- Sensor fusion
- Monitoring and alerting

cited from the Netica online tutorial webpage from the Netica software website:
http://www.norsys.com/tutorials/netica/nt_toc_A.htm

# Basic steps for construction of a BN model

1. Define the network variables (nodes) and their values / states. (e.g., target variables for queries, variables representing observables or inputs, and intermediary variables)

2. Define the **network structure**. (i.e., we need to specify the links/edges of a Bayesian Network. Often links indicate a cause and effect relationship among the nodes being connected.)

3. Define the network **conditional probability tables** (CPTs). (based on expert's opinion, subjective judgement, or derived directly from empirical data)

# Conditional Probability, Chain Rule, and Joint Probability

The joint probability (distribution) of A and B:

$$\Pr(A \text{ and } B) \equiv \Pr(A, B) = \Pr(A|B) * \Pr(B) = \Pr(B|A) * \Pr(A)$$

Determining the joint probability of A, B, C, D using **the Chain Rule**:

$$\Pr(A, B, C, D) = \Pr(A|B, C, D) * \Pr(B|C, D) * \Pr(D|C) * \Pr(C)$$

or $= \Pr(C|A, B) * \Pr(B|A) * \Pr(A) * \Pr(D|A, B, C)$

or ... (the order and combinations of the terms are flexible!)

For A and B to be independent of each other, the if and only-if conditions are: $\Pr(A, B) = \Pr(A) * \Pr(B) = \Pr(B) * \Pr(A)$

**The joint probability** of A, B, C, D will be reduced to:

$$\Pr(A, B, C, D) = \Pr(A) * \Pr(B) * \Pr(D) * \Pr(C)$$

# What is a Bayesian Network ?

- A Bayesian Network is a type of probabilistic network which represents and process probabilistic knowledge.
- The **qualitative component** of a Bayesian Network encodes a set of (conditional) dependence and independence statements among a set of random variables, informational precedence, and preference relations.
- The **quantitative component** of a Bayesian Network specifies the strengths of dependence relations using probability theory and preference relations using utility theory.

**cited and adapted from** Chapter 4 of eBook: Bayesian Networks and Influence Diagrams: a guide to construction and analysis. Uffe B. Kjaerulff & Anders L. Madsen, 2008. Springer.



# What is a Bayesian Network ?

In a Bayesian Network, we must quantify the **local relationships** between a variable (node) and its parents (nodes). In particular, we specify a conditional probability table (CPT) for each variable (node) in the network, which is a conditional probability distribution for a variable given its parent variables (parent nodes).

**These local conditional distributions induce a global probability distribution over all network variables** (the joint probability distribution).



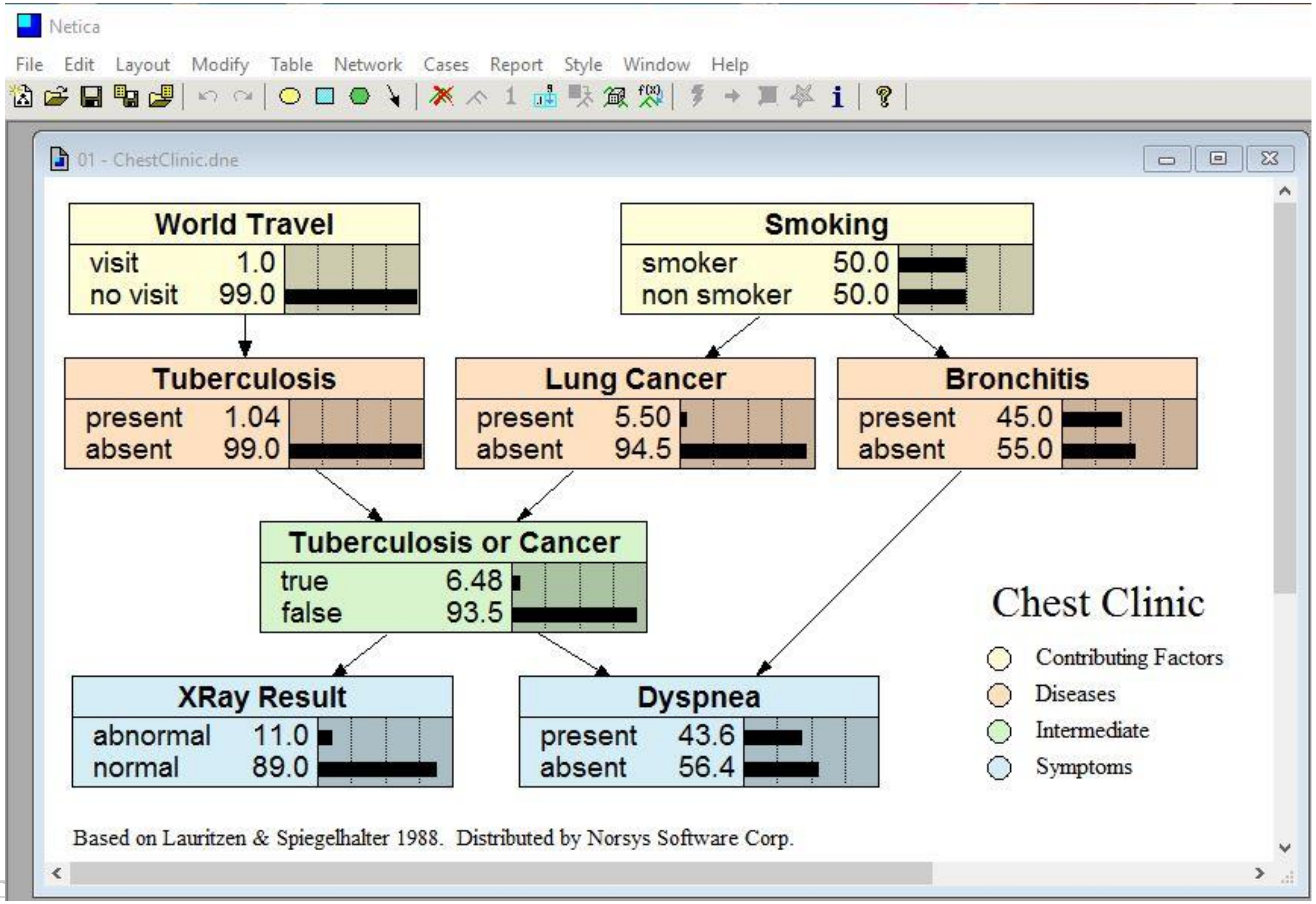
  

- A 15-minute break
- Demonstration of seven more Bayesian Networks models using Netica

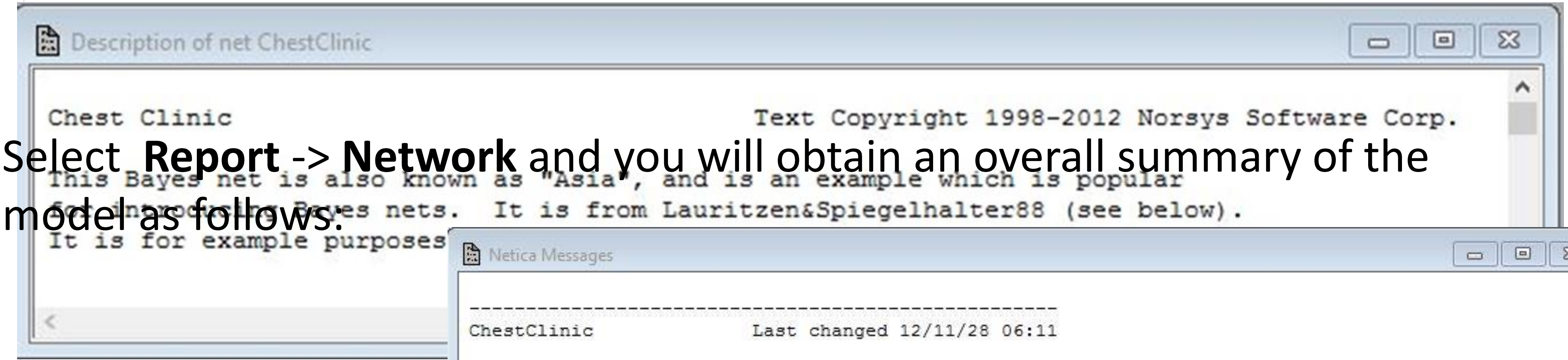
# Example 4: The chest clinic model

**Background information of the problem:** In the Chest Clinic model, a physician is diagnosing her patients with respect to lung cancer, tuberculosis, and bronchitis based on observations of symptoms and possible causes of the diseases.

# Example 4: The chest clinic model



**Example 4:** from Netica main menu, select **Window** -> **Description of Net**

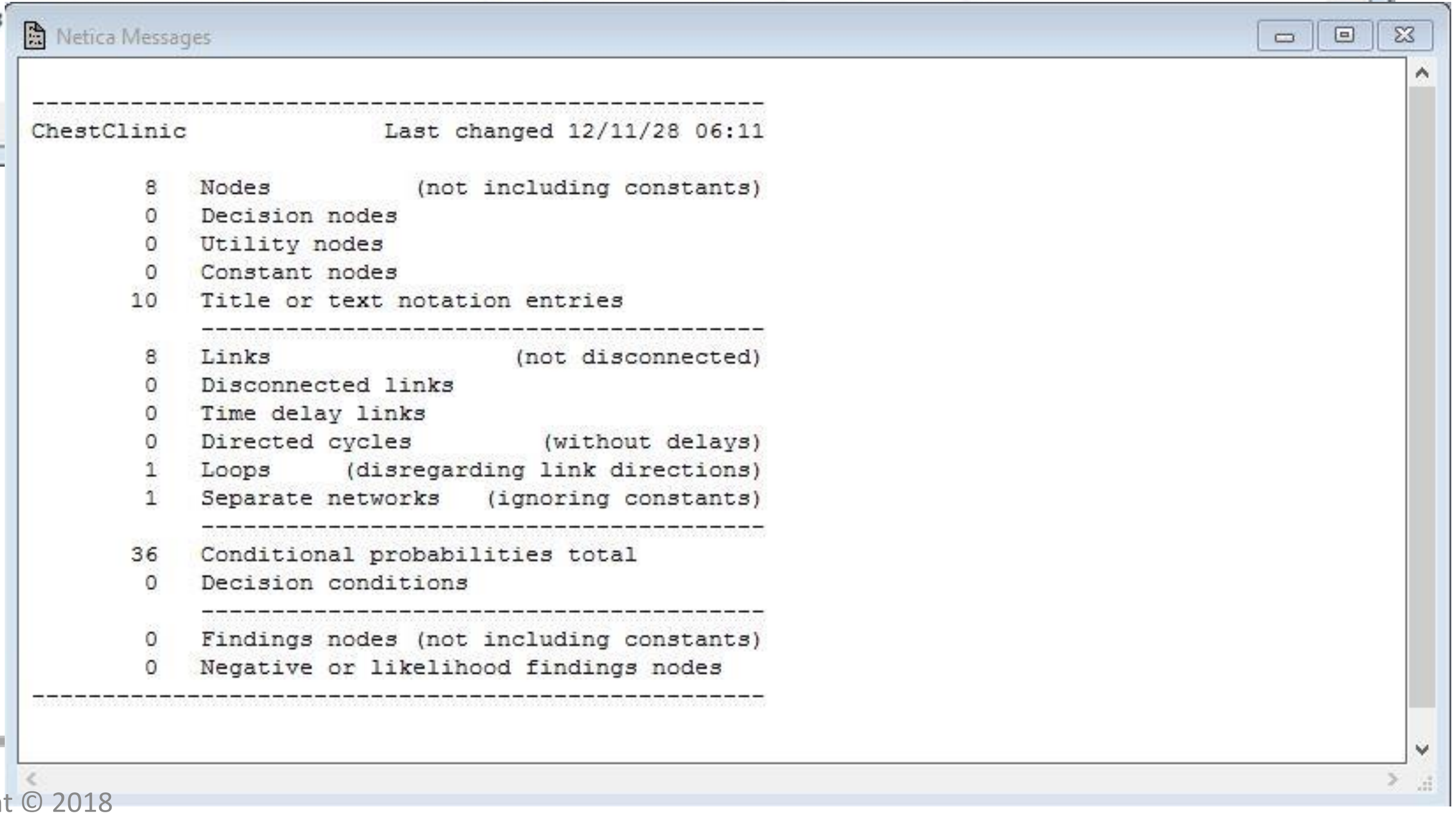


Description of net ChestClinic

Chest Clinic                    Text Copyright 1998-2012 Norsys Software Corp.

This Bayes net is also known as "Asia", and is an example which is popular
for introducing Bayes nets.  It is from Lauritzen&Spiegelhalter88 (see below).
It is for example purposes

Select **Report** -> **Network** and you will obtain an overall summary of the model as follows:

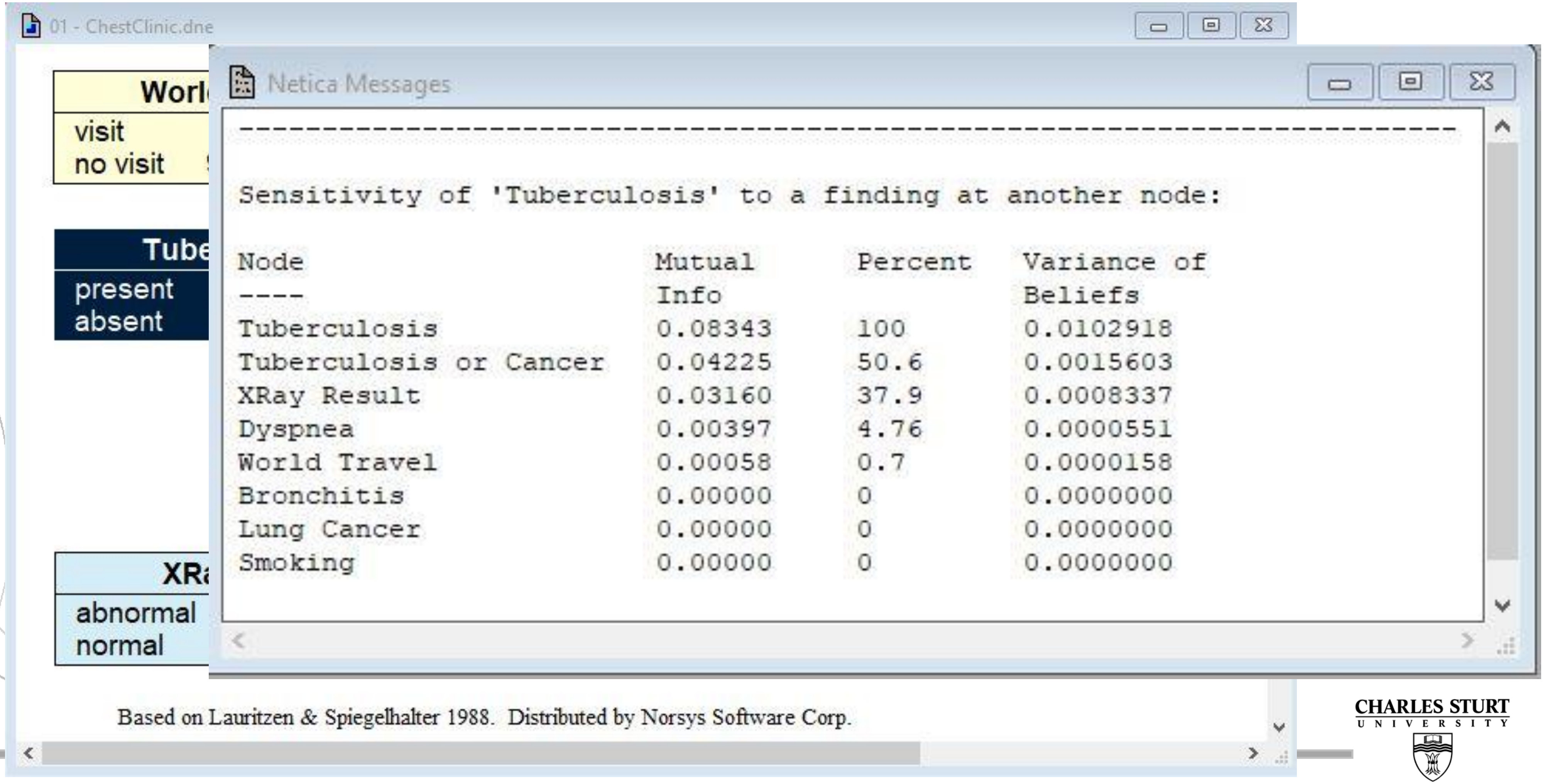


Netica Messages

```
------------------------------------------------------------
ChestClinic                Last changed 12/11/28 06:11

         8   Nodes            (not including constants)
         0   Decision nodes
         0   Utility nodes
         0   Constant nodes
        10   Title or text notation entries
             ----------------------------------------
         8   Links                    (not disconnected)
         0   Disconnected links
         0   Time delay links
         0   Directed cycles              (without delays)
         1   Loops         (disregarding link directions)
         1   Separate networks     (ignoring constants)
             ----------------------------------------
        36   Conditional probabilities total
         0   Decision conditions
             ----------------------------------------
         0   Findings nodes (not including constants)
         0   Negative or likelihood findings nodes
------------------------------------------------------------
```

Highlight a node and then select **Network -> Sensitivity to Findings** to perform a sensitivity analysis.



# Example 5: Oil Wildcatter decision net model

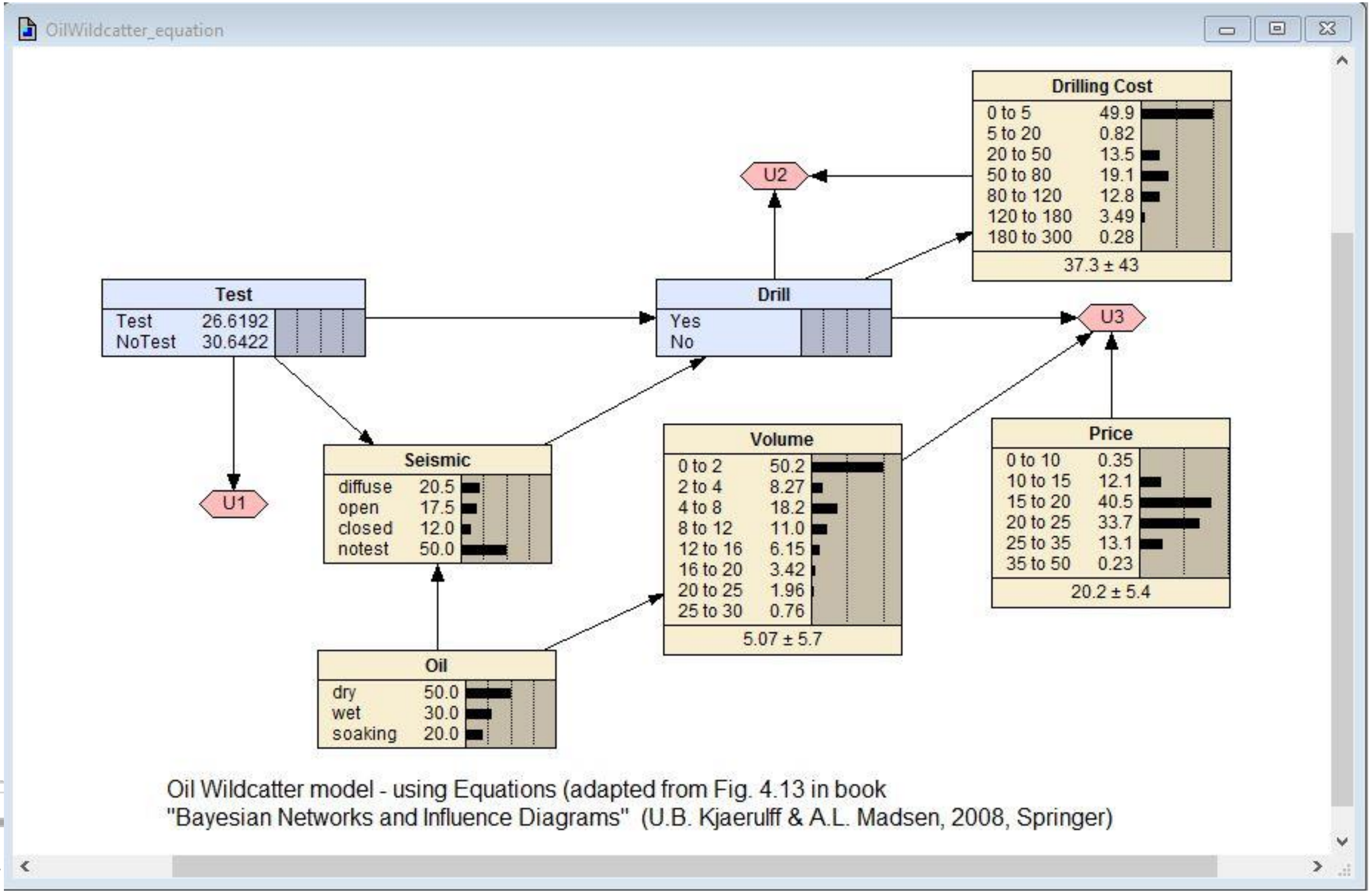
This is a multiple-decision-node (two decision nodes and three utility nodes) decision net model.

**Background information**: The decision maker would first make a decision on whether or not to perform a test – test of the geological structure of the site under consideration. When performed, this test will produce a test result (the Seismic node), Seismic depending on the amount of the actual state of oil deposit (the Oil node). Next, a decision (the Drill node) on whether or not to drill is made. There is a cost associated with drilling (Cost node), while the revenue is a function of oil volume (Volume node) and oil price (Price node).

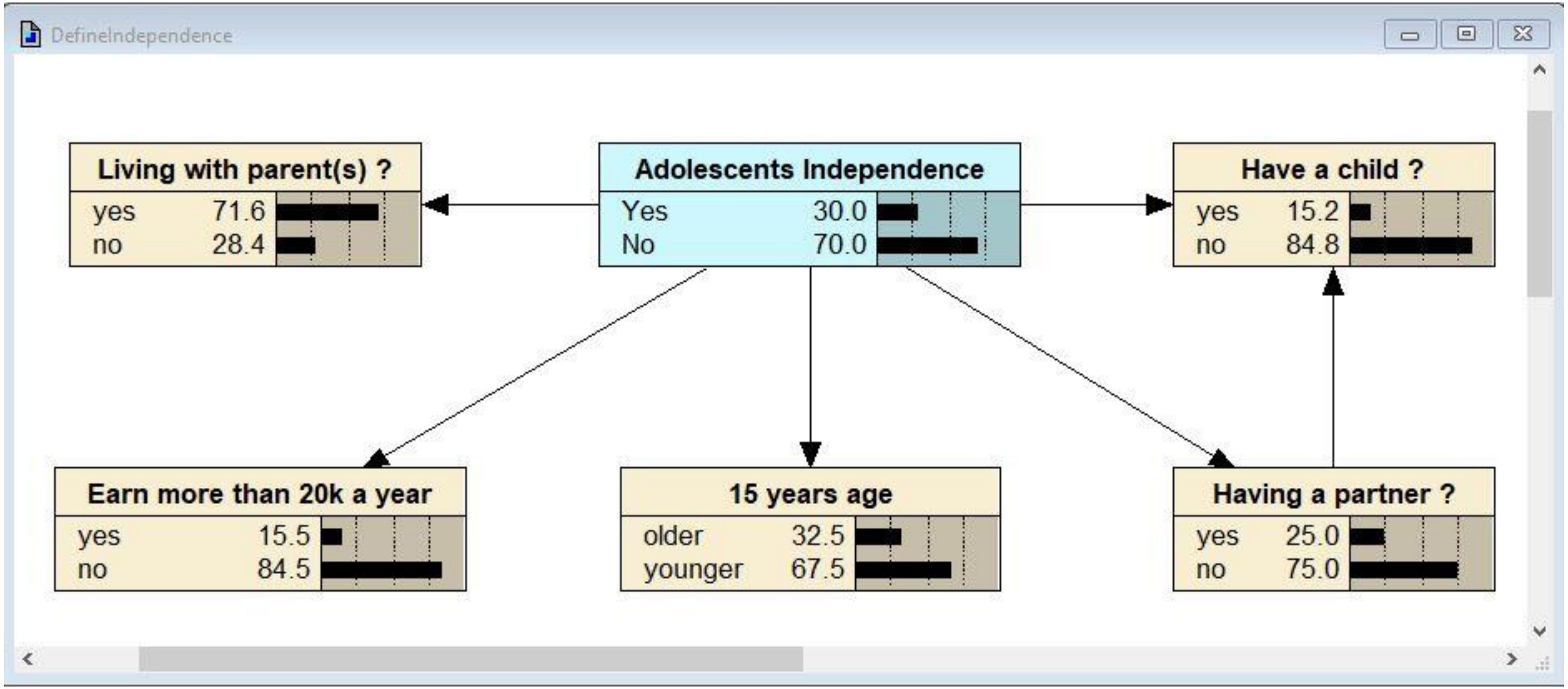


# Example 5: Oil Wildcatter decision net model

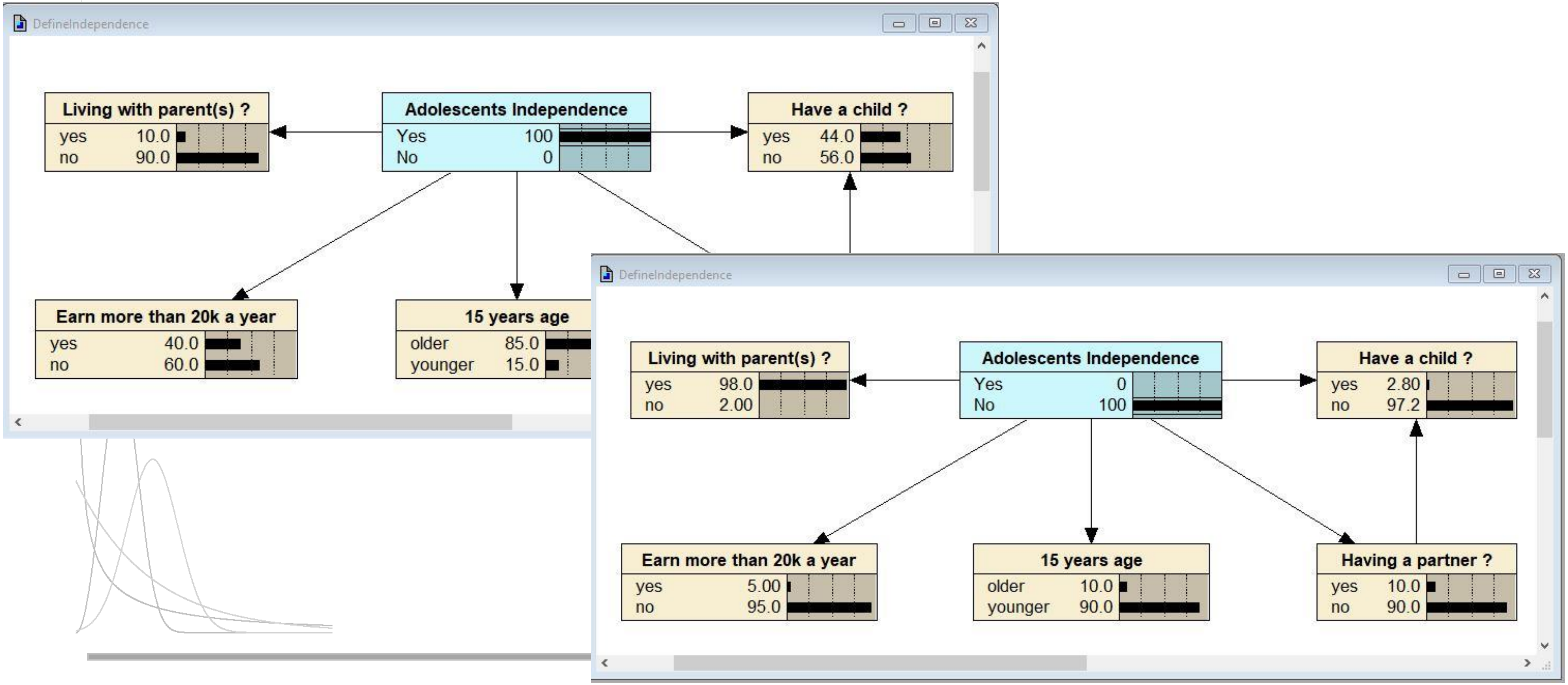


# Example 6: A probabilistic definition of Adolescents' Independence

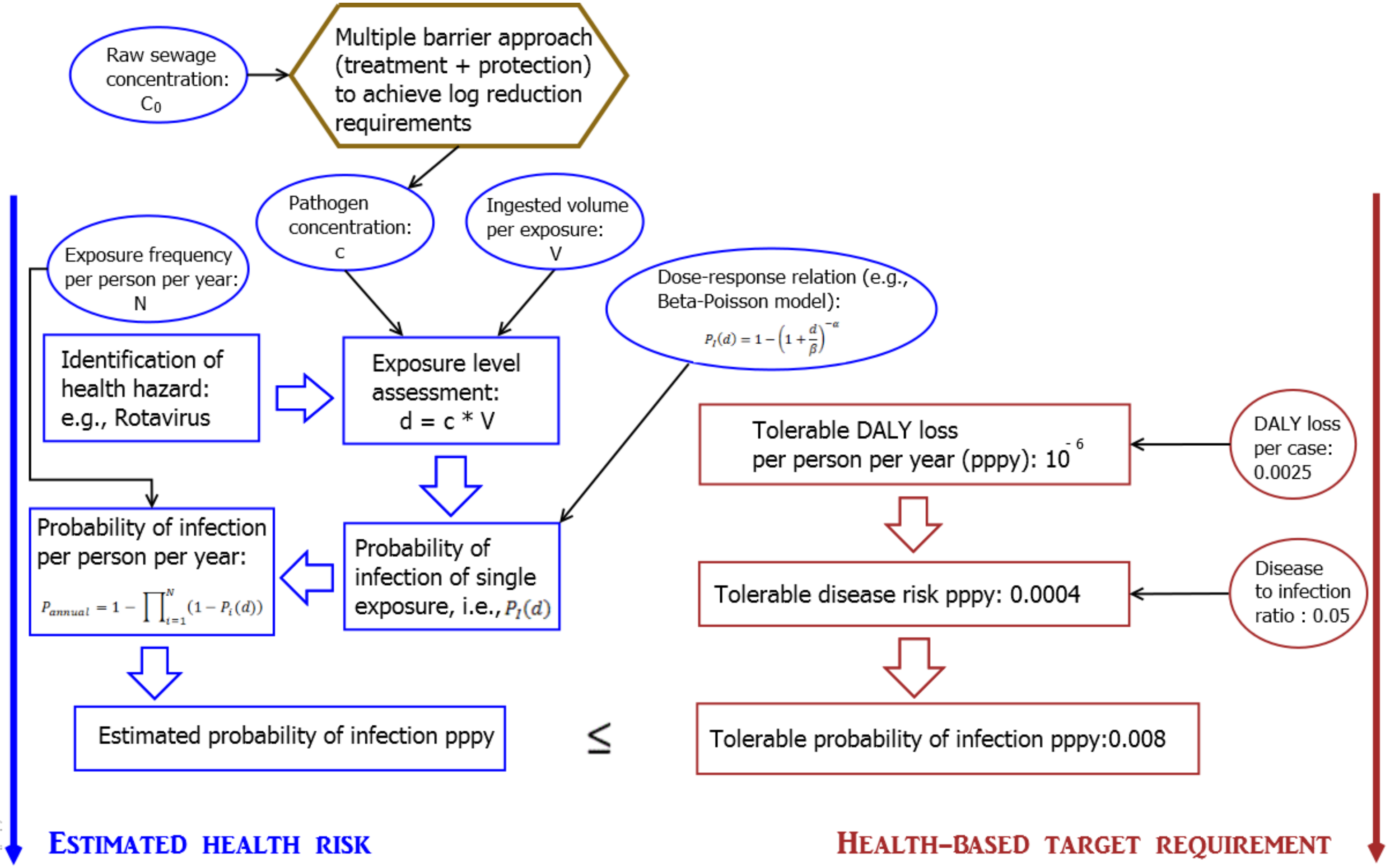


# Example 6: A probabilistic definition of Adolescents' Independence

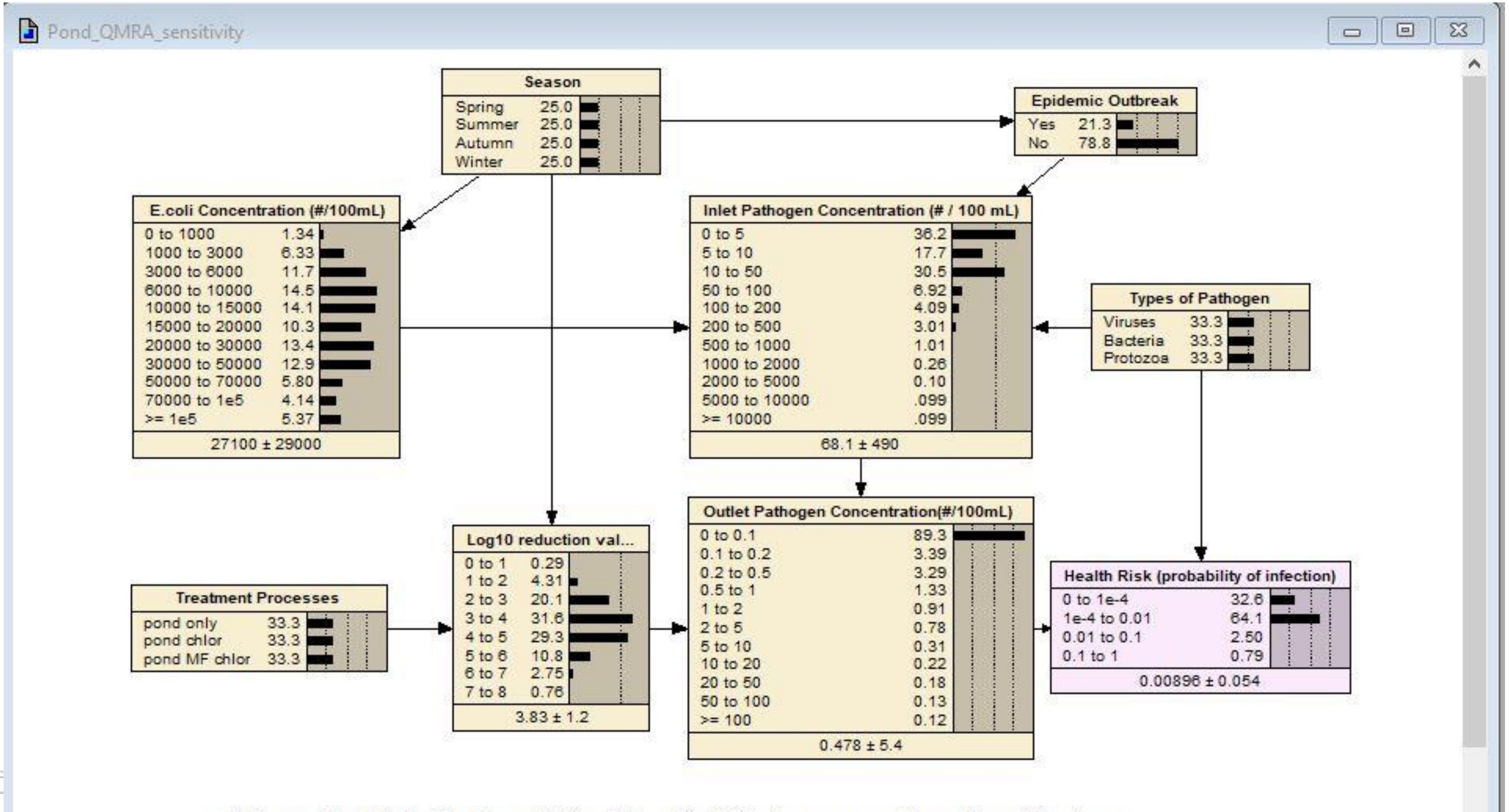


# Example 7: A BN model for Quantitative Microbial Risk Assessment (QMRA): the schematic QMRA flow-chart



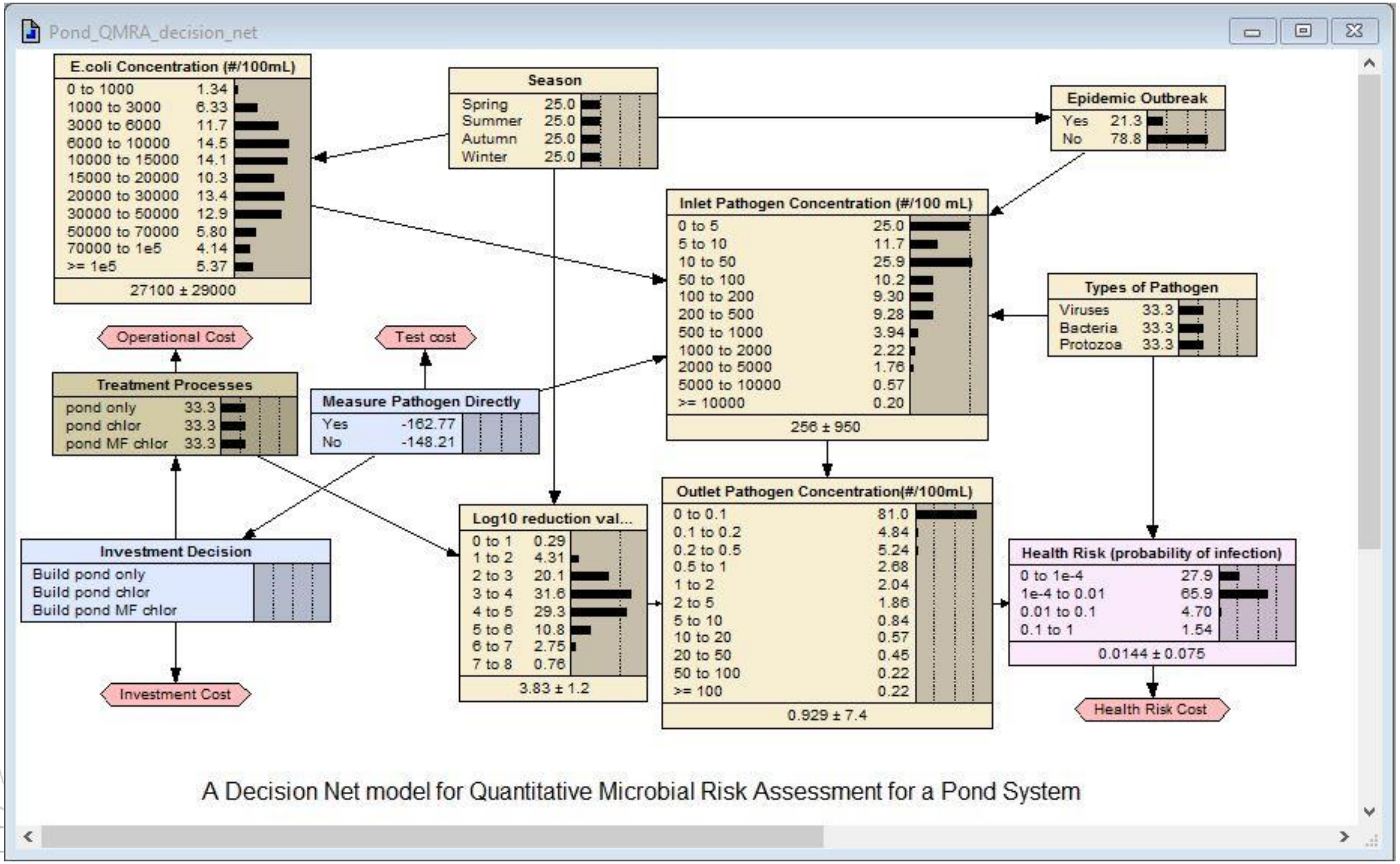

  

# Example 7: A Bayes net model on QMRA analysis



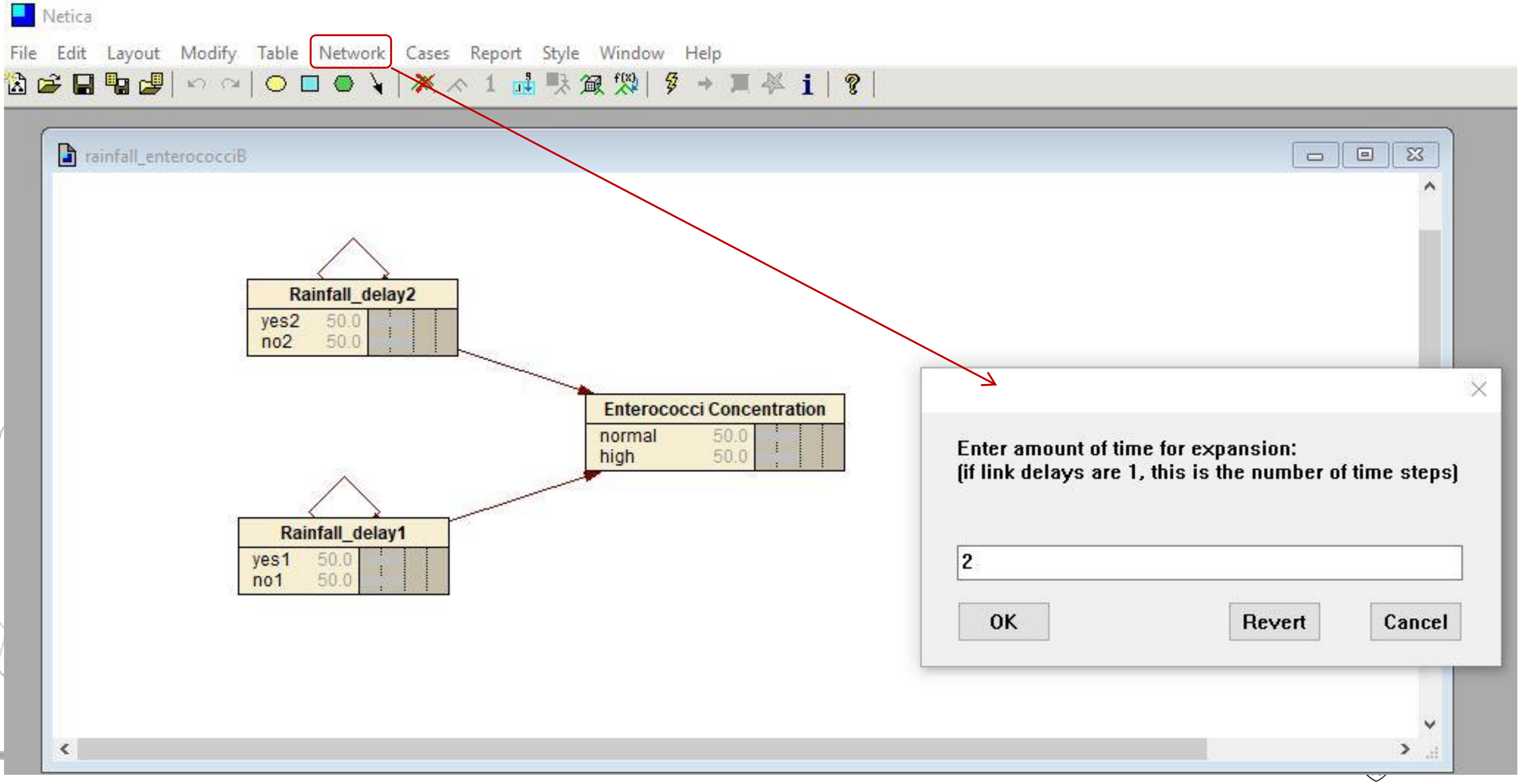

A Bayes Net Model for Quantitative Microbial Risk Assessment for a Pond System

# Example 8: A decision net model on QMRA analysis

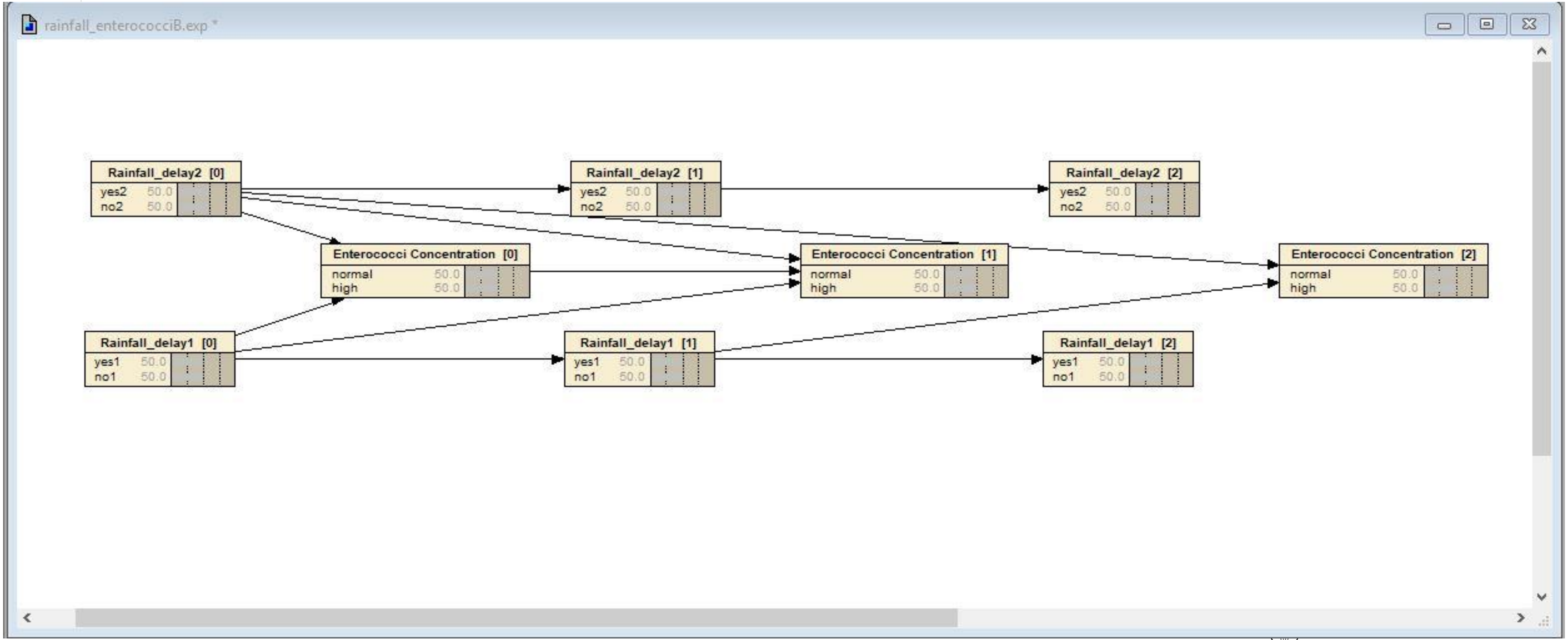



A Decision Net model for Quantitative Microbial Risk Assessment for a Pond System

# Example 9: A simple rainfall-enterococci DBN model



# Example 9: A simple rainfall-enterococci DBN model



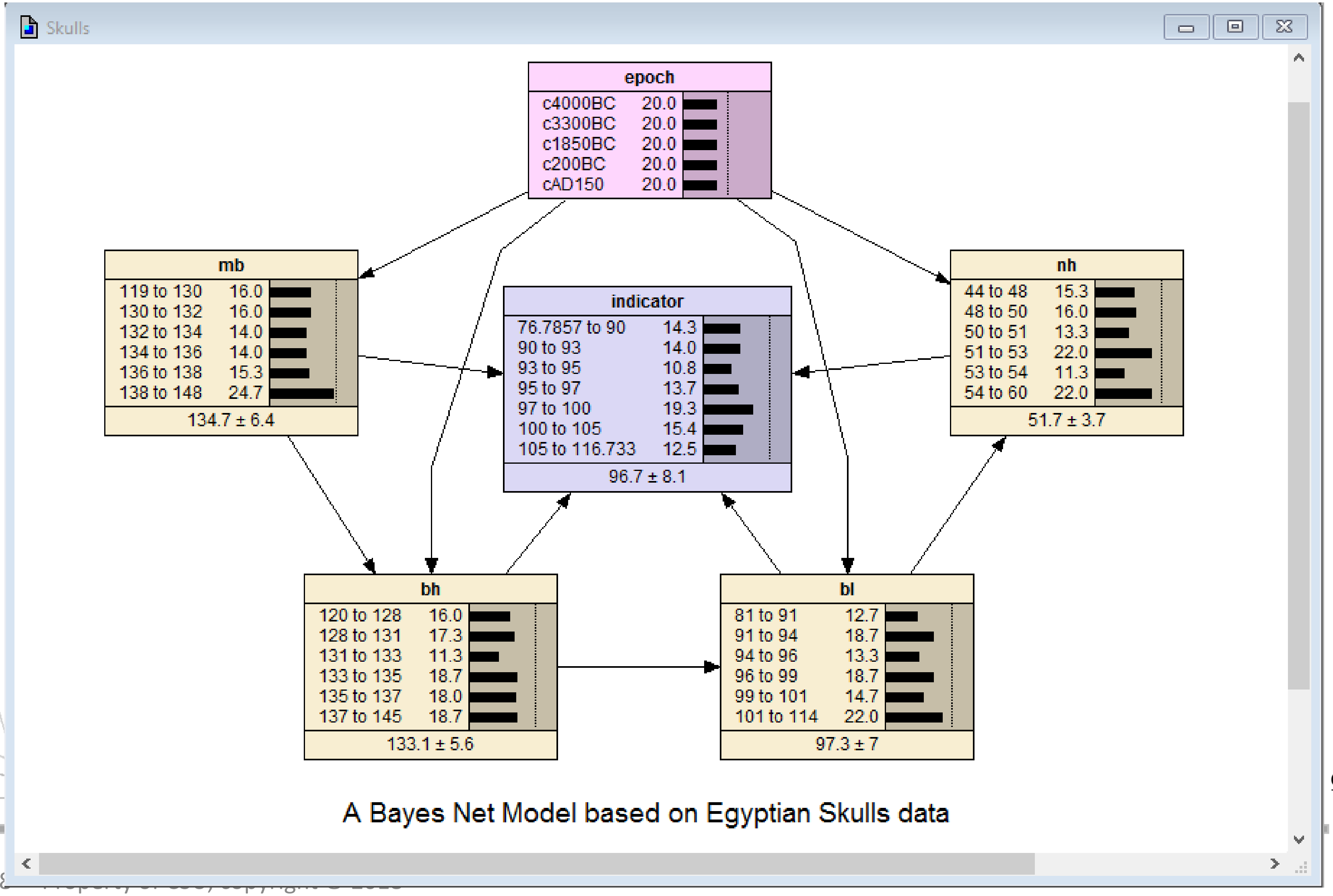
# Example 10: The Egyptian Skulls data model

- **Background information:** The data are measurements made on Egyptian skulls from five epochs. A data frame with 150 observations on the following 5 variables.
- **epoch**: the epoch the skull as assigned to, a factor with levels c4000BC c3300BC, c1850BC, c200BC, and cAD150, where the years are only given approximately, of course.
- **mb**: maximum breaths of the skull; **bh**: basibregmatic heights of the skull; **bl**: basialiveolar length of the skull; **nh**: nasal heights of the skull.

- The **question** is whether the measurements change over time. Non-constant measurements of the skulls over time would indicate interbreeding with immigrant populations. This has been treated as a typical MANOVA problem.

Source: D. J. Hand, F. Daly, A. D. Lunn, K. J. McConway and E. Ostrowski (1994). A Handbook of Small Datasets, Chapman and Hall/CRC, London.

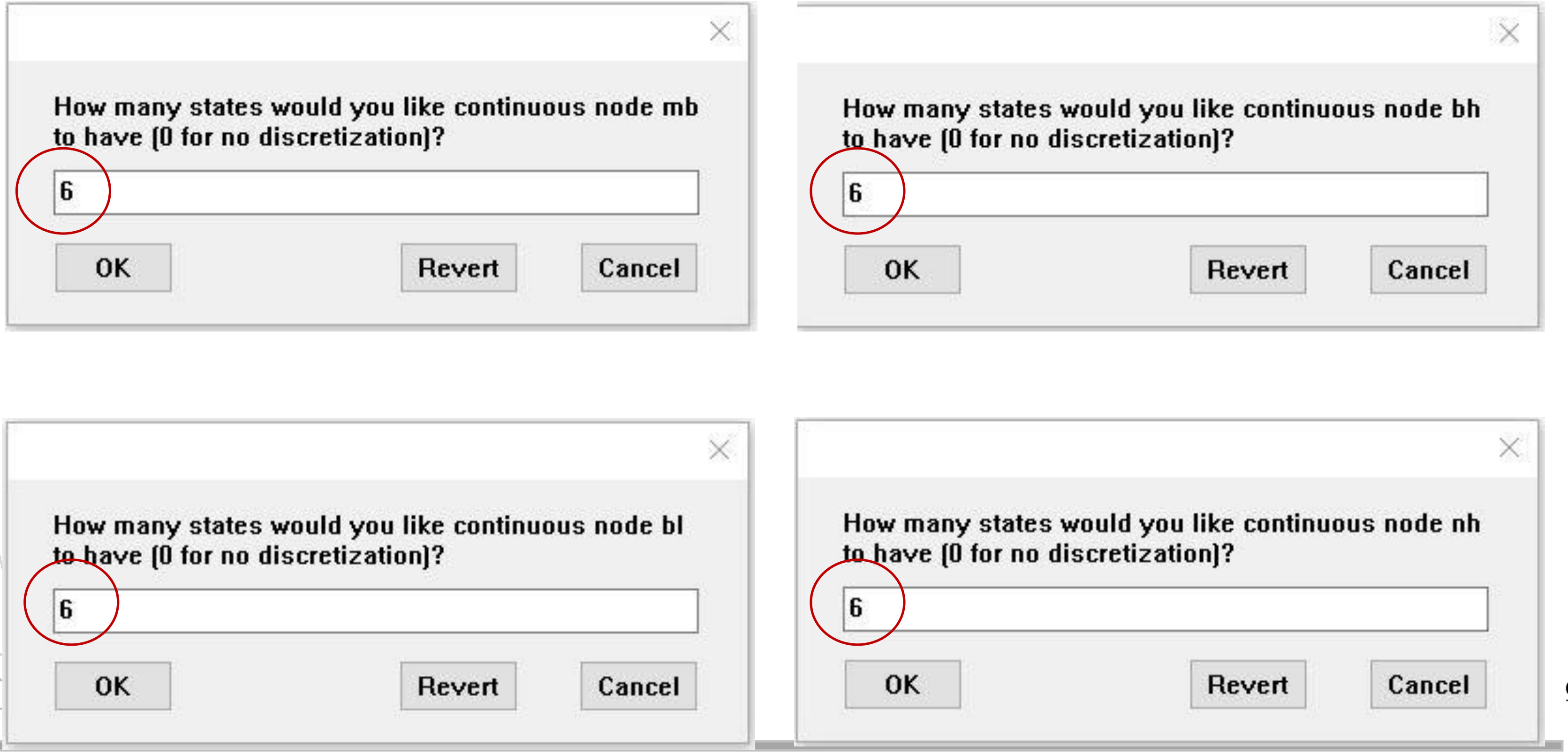
# Example 10: The Egyptian Skulls data model




A Bayes Net Model based on Egyptian Skulls data

# Example 10: The Egyptian Skulls data model

Learning the model structure: Open a new net window;
**Cases** -> **Learn** -> **Add Case File Nodes …** ; Select skulls.csv

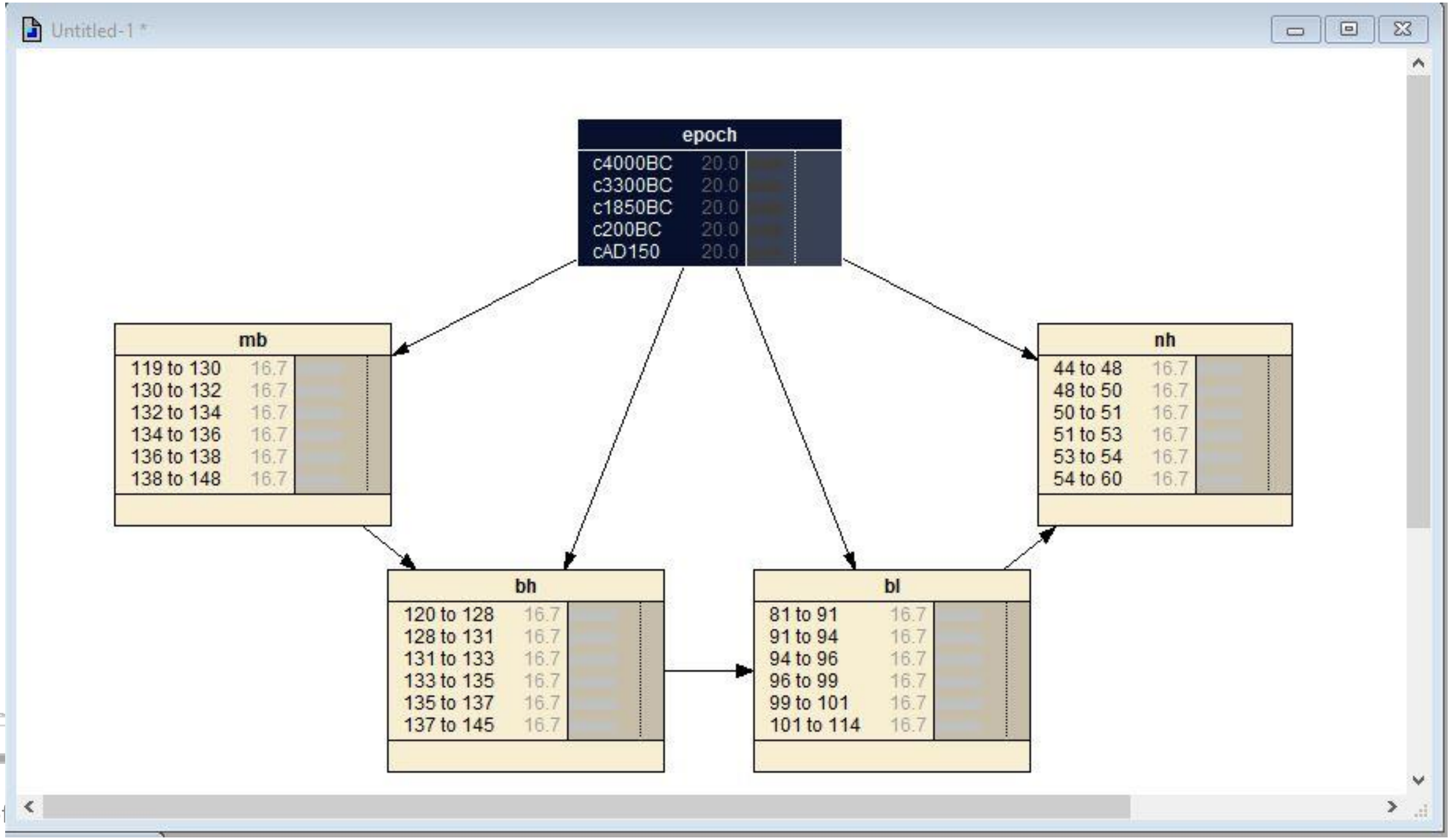




# Example 10: The Egyptian Skulls data model

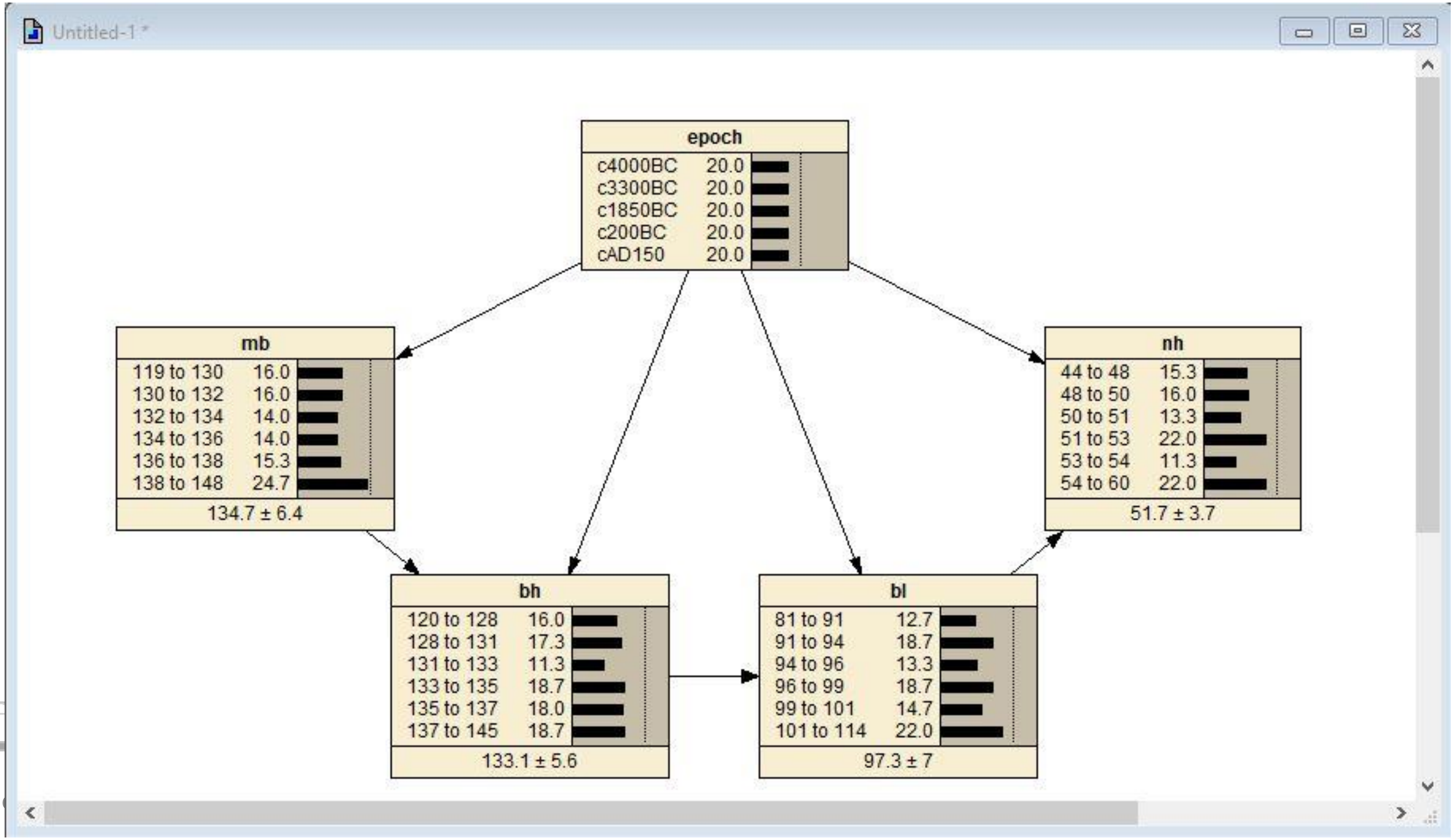
Learning the model structure: Open a new net window;
**Cases** -> **Learn** -> **Add Case File Nodes …** ; Select skulls.csv



# Example 10: The Egyptian Skulls data model

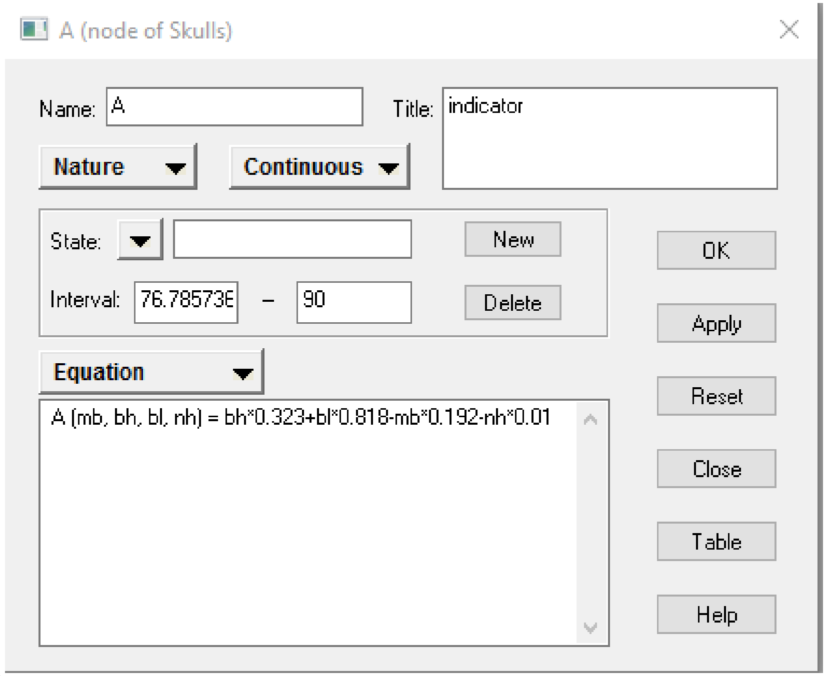
Learning the CPTs: **Cases** -> **Learn** -> **Learning Using EM ;** Select skull.csv .

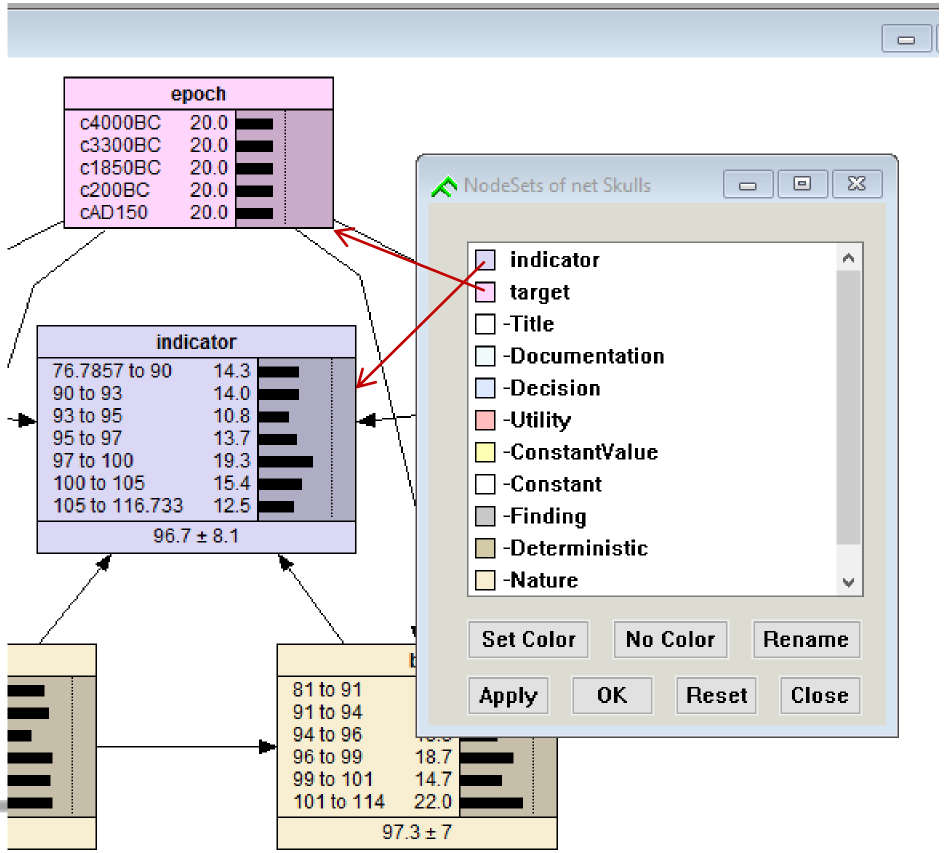
The finished model:



# Example 10: The Egyptian Skulls data model

A slightly modified version of the finished model: (1) add a new node ‘Indicator’ which is a function of the four multiple response variables. (2) refine the display style so that the research question is well highlighted.



# Where to go from here: An introduction of the Bayesian Network workshop

**“Learning Bayesian Networks by hands-on practice with examples”**

-- a 3-day workshop

# Who should come (the intended audience) for these Bayesian Network workshops

Any researchers and/or practitioners who are dealing with problems with reasoning and decision making under uncertainty, particularly with complex system with mixed types of data (e.g., numeric and categorical, measurements and expert knowledge), should come to this Bayesian Network workshop. No experience with BN modelling is required for the participants. Only basic data analysis skills are expected (e.g., cross tabulation analysis of two or more categorical variables).

**If you can use Microsoft Word confidently,**

**you are ready to learn Bayesian Networks using Netica!**

# Learning Bayesian Networks by hands-on practice with examples:

- The first two days of the workshop essentially covers the topics of construction and simple analysis of Bayes net and decision net (i.e., influence diagram) models; introduction to elicitation of model parameters from empirical data.
- The third day of the workshop establishes participants' ability to build non-trivial BN models (Bayes net) and simple decision net models or very simple Dynamic Bayesian Network (DBN) models for solving their research / real life problems. The model building methods include both manually specifying CPTs or learning model structure from a data file via Netica's built-in optimization algorithm.

# Bayesian Network Software Tools

- AgenaRisk
- Analytica
- BayesiaLab
- BNT
- CaMML
- Genie
- Hugin
- JavaBayes
- MSBNx
- Netica
- Tetrad
- Uninet
- WinBugs

**Netica,** the world's most widely used Bayesian network development software, was designed to be simple, reliable, and high performing. For managing uncertainty in business, engineering, medicine, or ecology, it is the tool of choice for many of the world's leading companies and government agencies.

Source: Australasian Bayesian Network Modelling Society website: [http://abnms.org/resources.php](http://abnms.org/resources.php)

- The free version of the BN software package **Netica Application** (https://norsys.com/netica.html ) is the primary tool for teaching and learning construction and analysis of BN models for the workshops.
- All BN models used/built in these workshops are available as resource materials and the selected literature references are recommended for participants for further study or future reference.

# Important Message

**Reminder:** All examples used in this presentation should be considered  for demonstration purpose only for learning Bayesian Networks!

While the methodology is theoretically sound, the data and analysis outcome shall not necessarily be taken as real or a valid solution to the real problem.



# Thank You !

And

Look forward to meeting you in
the 3-day Bayesian Network workshop.

**John Xie** contact details: Email: gxie@csu.edu.au
Phone: +61-2-69332229

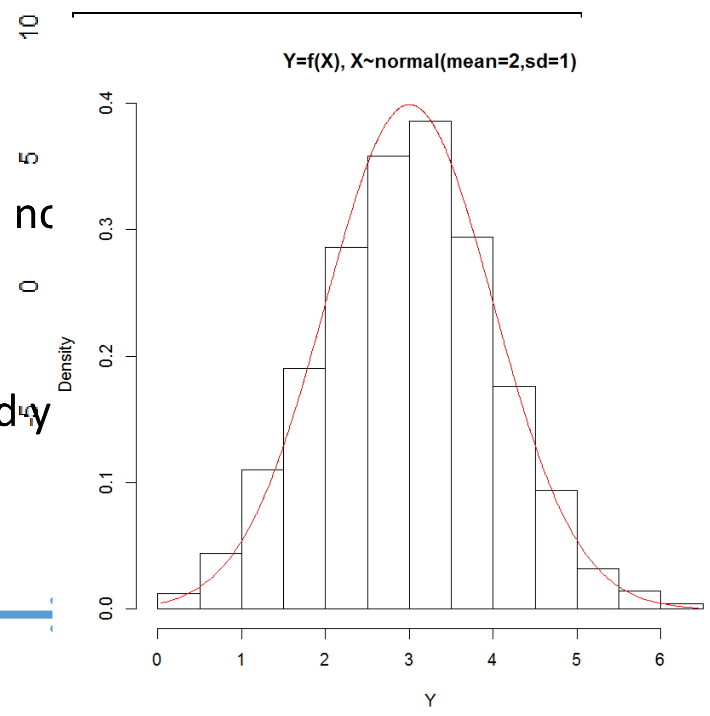
# Why (and How) things get complicated

A number question: **3 + 2 = 5**

An equation with one unknown: **x + 2 = 5** x can nc

An equation with two unknowns: **x + y = 5** x and y

A mathematical function: **y = 5 – x**




**X** is a **random variable** and **Y = f(X)**: **Y = 5 – X , X ~ normal(mean=2, sd = 1)**