

Introduction to Decision Trees in R

Three-strategy decision tree in R - HVE Example

The DARTH workgroup

Wednesday, September 13, 2023

This work is developed by the Decision Analysis in R for Technologies in Health (DARTH) workgroup:

- Fernando Alarid-Escudero, PhD
- Eva A. Enns, MS, PhD
- M.G. Myriam Hunink, MD, PhD
- Hawre J. Jalal, MD, PhD
- Eline Krijkamp, PhD
- Petros Pechlivanoglou, PhD
- Alan Yang, MSc

Please acknowledge our work. See details to cite below.

Overview

This guide has been developed by the Decision Analysis in R for Technologies in Health (DARTH) workgroup to guide you through the process of setting up, analyzing, and interpreting results of a decision tree model on a hypothetical disease setting in R.

Table of Contents

1. Introduction
2. Prerequisites
3. Setup Environment
4. Load Required Libraries
5. Define Parameters
6. Create Decision Tree Model
7. Cost-Effectiveness Analysis
8. Visualizing the Results
9. Acknowledgements

1. Introduction

This R Markdown document provides a foundation for constructing a decision tree model for HVE:

- Viral encephalitis can be caused by herpes virus (HVE) or other viruses (OVE)
 - $\Pr(\text{HVE})=52$
- Untreated HVE has 71% risk of complications; for OVE the figure is 1%
- For HVE, the drug vidarabine decreases risk of complications from 71% to 36%
- For OVE, vidarabine side effects increase risk of complications from 1% to 20%

The three strategies being compared are:

1. Do not treat
2. Treat
3. Biopsy

Health outcomes measured as remaining quality-adjusted life years (QALYs):

- Without complications: 20 QALYs
- With complications: 19 QALYs
- Death (due to biopsy): 0 QALYs
- One-time loss of 0.01 QALYs from biopsy

Costs are in Canadian Dollars:

- Cost of VE medical care, treatment, and diagnostics
- Medical care without complications: \$1,200
- Medical care with complications: \$9,000
- Vidarabine: \$9,500
- Biopsy: \$25,000
- Remaining lifetime costs
- Without complications: \$50,000
- With complications: \$60,000

Throughout this guide, you'll gain insight into how these strategies affect patient outcomes in terms of quality-adjusted life years (QALYs) and associated costs.

2. Prerequisites

Ensure you have R and RStudio installed. Familiarize yourself with R Markdown and R's decision tree modeling.

R Software

The R Project for Statistical Computing: <https://www.r-project.org/>

Installing R and RStudio

Install R

Download R version 4.3.1 from <https://cran.r-project.org/bin/windows/base/old/4.3.1/>

Download the file `R-4.3.1-win.exe` and follow the installation procedure.

Install RStudio

Download and install the free version of RStudio Desktop from: <https://posit.co/download/rstudio-desktop/#download>

3. Setup Environment

To ensure reproducibility, it's crucial to clear the current environment and set appropriate chunk options:

- `echo = TRUE` shows R code in the document.
- `warning = FALSE` hides warnings.
- `message = FALSE` hides messages from R packages when they are loaded.
- `eval = TRUE` runs all R code and display outputs in the document.

```
rm(list = ls()) # clears the current environment by removing all variables in it
```

4. Load Required Libraries

Before modeling, we need to ensure required libraries are available:

```
# use this package to conveniently install other packages
if (!require('pacman')) install.packages('pacman'); library(pacman)

# load (install if required) packages from CRAN
p_load("dampack")
```

5. Define Parameters

Parameters are crucial for our decision model. Here's a breakdown:

- Probabilities: Various probabilities like the prevalence of HVE, complications with untreated/treated HVE, and death due to biopsy.
- QALYs: Represents the quality-adjusted life years for various outcomes.
- Costs: Financial costs associated with different outcomes.

Below table contains all parameters of this model with values and suggested variable names in R. We will code them into R.

Parameter	Variable Name	Value
Prevalence of HVE	p_HVE	0.52
Probability of complications without treatment		
HVE	p_HVE_comp	0.71
OVE	p_OVE_comp	0.01
Probability of complications with vidarabine treatment		
HVE	p_HVE_comp_tx	0.36
OVE	p_OVE_comp_tx	0.20
Probability of complications due to brain biopsy	p_biopsy_comp	0.05
QALYs		
Remaining QALYs without complications	q_VE	20
Remaining QALYs with complications	q_VE_comp	19
Remaining QALYs if died during biopsy	q_death_biopsy	0
One-time QALY loss due to biopsy	q_loss_biopsy	0.01
Costs		
Medical care without complications	c_VE	1,200
Medical care with complications	c_VE_comp	9,000
Vidarabine	c_tx	9,500
Biopsy	c_biopsy	25,000
Lifetime without complications	c_lifetime	50,000
Lifetime with complications	c_lifetime_comp	60,000

Pay attention to the structure of the variable names. Make sure they are simple, intuitive and informative.

```
v_names_str    <- c('Do not treat', "Treat", "Biopsy")  # names of strategies
n_str          <- length(v_names_str)                  # number of strategies
wtp            <- 100000                                # willingness to pay threshold

# Probabilities
p_HVE          <- 0.52  # prevalence of HVE
p_HVE_comp     <- 0.71  # complications with untreated HVE
p_OVE_comp     <- 0.01  # complications with untreated OVE
p_HVE_comp_tx  <- 0.36  # complications with treated HVE
p_OVE_comp_tx  <- 0.20  # complications with treated OVE
p_biopsy_death <- 0.005 # probability of death due to biopsy

# QALYs
q_VE          <- 20    # remaining QALYs for those without VE-related complications
q_VE_comp     <- 19    # remaining QALYs for those with VE-related complications
q_death_biopsy <- 0    # remaining QALYs for those who died during biopsy
q_loss_biopsy <- 0.01  # one-time QALY loss due to brain biopsy

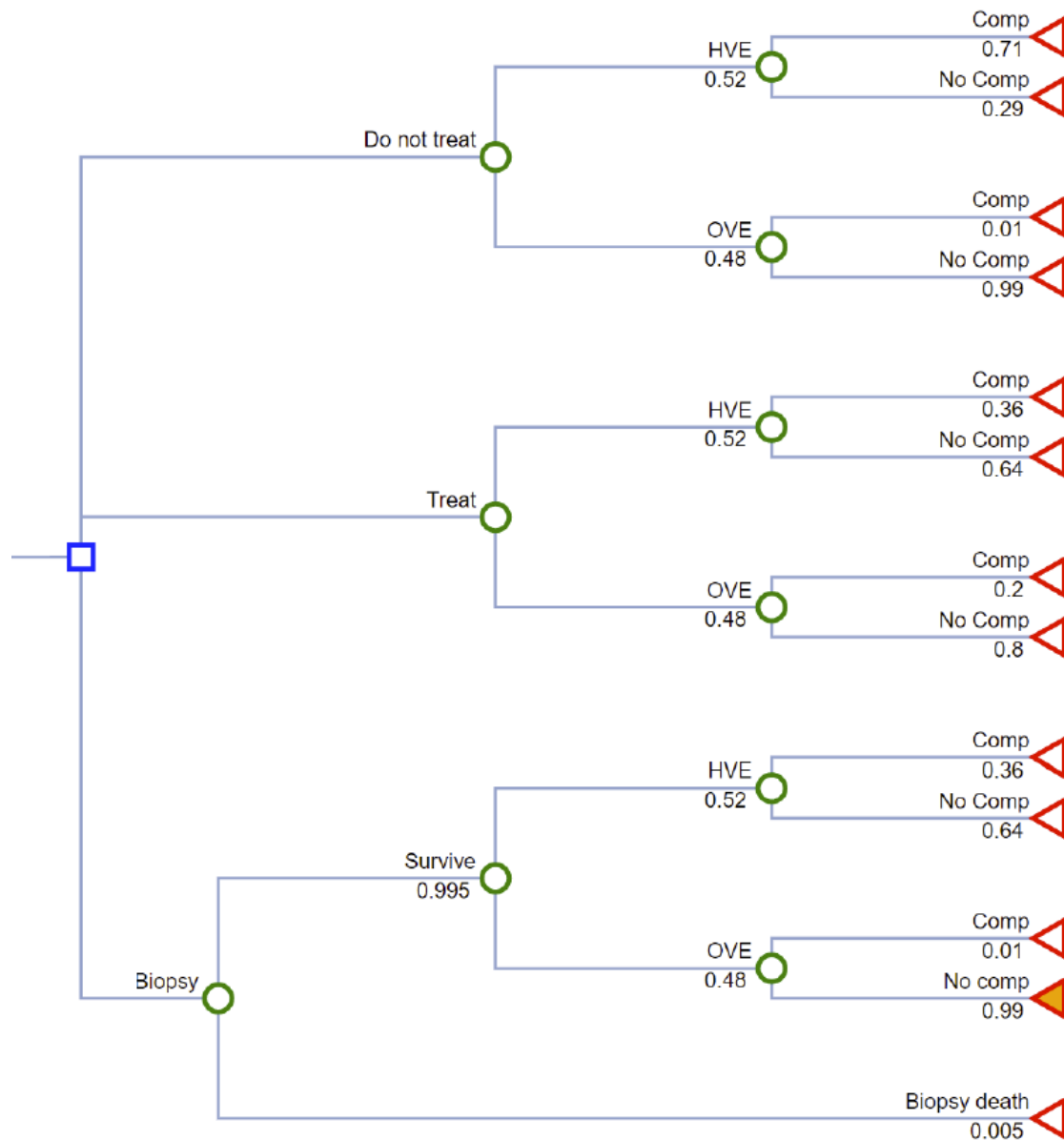
# Costs
c_VE          <- 1200  # cost of viral encephalitis care without complications
c_VE_comp     <- 9000  # cost of viral encephalitis care with complications
c_tx          <- 9500  # cost of treatment
c_biopsy      <- 25000 # cost of brain biopsy
c_lifetime    <- 50000 # expected remaining lifetime costs without complications
c_lifetime_comp <- 60000 # expected remaining lifetime costs with complications
```

6. Create Decision Tree Model

Decision trees evaluate all possible strategies. We will:

1. Define the probabilities associated with each path for every strategy.
2. Set the terminal node values for QALYs and Costs.
3. Calculate expected values.
4. Store the results.

Below is a figure of the decision tree:



1. Define the probabilities associated with each path for every strategy

```
# For each strategy, define vector of path probabilities for each path

# No treatment
v_prob_notx <- c(  p_HVE *      p_HVE_comp , # HVE, complications
                  p_HVE * (1 - p_HVE_comp) , # HVE, no complications
                  (1 - p_HVE) *      p_OVE_comp , # OVE, complications
                  (1 - p_HVE) * (1 - p_OVE_comp)) # OVE, no complications

# Treat all
v_prob_tx <- c(      p_HVE *      p_HVE_comp_tx , # HVE, complications
                 p_HVE * (1 - p_HVE_comp_tx) , # HVE, no complications
                 (1 - p_HVE) *      p_OVE_comp_tx , # OVE, complications
                 (1 - p_HVE) * (1 - p_OVE_comp_tx)) # OVE, no complications

# Biopsy
v_prob_biopsy <- c((1-p_biopsy_death) *      p_HVE *      p_HVE_comp_tx, # survive, HVE, comp
                  (1-p_biopsy_death) *      p_HVE * (1 - p_HVE_comp_tx), # survive, HVE, no comp
                  (1-p_biopsy_death) * (1 - p_HVE) *      p_OVE_comp , # survive, OVE, comp
                  (1-p_biopsy_death) * (1 - p_HVE) * (1 - p_OVE_comp), # survive, OVE, no comp
                  p_biopsy_death) # die from biopsy
```

Below are explanations of some of the probabilities:

$p_{HVE} * p_{HVE_comp}$: This calculates the probability that a person has the HVE condition and subsequently experiences complications under “Do not treat” strategy.

$p_{HVE} * (1 - p_{HVE_comp_tx})$: This calculates the probability that a person has the HVE condition but does not experience complications under “Treat” strategy..

$(1 - p_{HVE}) * p_{OVE_comp}$: This calculates the probability that a person without HVE (thus having OVE) and experiences complications under “Do not treat” strategy.

$(1-p_{biopsy_death}) * p_{HVE} * p_{HVE_comp_tx}$: This calculates the probability that a person who survives the biopsy, has HVE and then experiences complications under “Treat” strategy.

p_{biopsy_death} : The probability of death from biopsy.

2. Set the terminal node values for QALYs and Costs

For each strategy, we create a vector of terminal node values at each path terminus, including all branches (see tree figure above). We have one set for QALYs and another for costs.

```
## QALYs
v_qaly_notx <- c(q_VE_comp, # HVE, complications
               q_VE      , # HVE, no complications
               q_VE_comp, # OVE, complications
               q_VE)      # OVE, no complications

v_qaly_tx <- c(q_VE_comp, # HVE, complications
              q_VE      , # HVE, no complications
              q_VE_comp, # OVE, complications
              q_VE)      # OVE, no complications

v_qaly_biopsy <- c(q_VE_comp - q_loss_biopsy, # survive, HVE, complications
                  q_VE      - q_loss_biopsy, # survive, HVE, no complications
                  q_VE_comp - q_loss_biopsy, # survive, OVE, complications
                  q_VE      - q_loss_biopsy, # survive, OVE, no complications
                  q_death_biopsy)           # die from biopsy

## Costs
v_cost_notx <- c(c_VE_comp + c_lifetime_comp, # HVE, complications
                c_VE      + c_lifetime,      # HVE, no complications
                c_VE_comp + c_lifetime_comp, # OVE, complications
                c_VE      + c_lifetime)       # OVE, no complications

v_cost_tx <- c(c_VE_comp + c_lifetime_comp + c_tx, # HVE, complications
              c_VE      + c_lifetime      + c_tx, # HVE, no complications
              c_VE_comp + c_lifetime_comp + c_tx, # OVE, complications
              c_VE      + c_lifetime      + c_tx) # OVE, no complications

v_cost_biopsy <- c(c_biopsy + c_VE_comp + c_lifetime_comp + c_tx, # survive, HVE, complications
                  c_biopsy + c_VE      + c_lifetime      + c_tx, # survive, HVE, no complications
                  c_biopsy + c_VE_comp + c_lifetime_comp,      # survive, OVE, complications
                  c_biopsy + c_VE      + c_lifetime,           # survive, OVE, no complications
                  c_biopsy)                                     # die from biopsy
```


3. Calculate expected values

```
# Calculate expected value for each outcome and strategy
## Expected QALYs
total_qaly_notx <- v_prob_notx %*% v_qaly_notx
total_qaly_tx <- v_prob_tx %*% v_qaly_tx
total_qaly_biopsy <- v_prob_biopsy %*% v_qaly_biopsy

## Expected costs
total_cost_notx <- v_prob_notx %*% v_cost_notx
total_cost_tx <- v_prob_tx %*% v_cost_tx
total_cost_biopsy <- v_prob_biopsy %*% v_cost_biopsy
```

We utilize matrix operations for the calculations of expected values. In this case, for each strategy, the probability and the outcome (cost/ QALYs) are multiplied within each branch. Then, such multiplication-products from all branches are summed to obtain the total outcome for that strategy. This is done through matrix operation (dot product).

4. Store the results

We store the total costs and QALYs for all strategies in a vector, and along with a vector of strategy names they are stored in a dataframe for cost-effectiveness analysis (CEA).

```
## Store expected values in a vector
v_total_qaly <- c(total_qaly_notx, total_qaly_tx, total_qaly_biopsy) # vector of total QALYs
names(v_total_qaly) <- v_names_str # assign strategy names

v_total_cost <- c(total_cost_notx, total_cost_tx, total_cost_biopsy) # vector of total costs
names(v_total_cost) <- v_names_str # assign strategy names

## Final model output in a dataframe for CEA
df_output <- data.frame(Strategy = v_names_str,
                        Cost = v_total_cost,
                        Effect = v_total_qaly)
```

7. Cost-Effectiveness Analysis

We use the `calculate_icers` function from the R package `dampack` to conduct a CEA. The function takes a vector of costs, a vector of QALYs (or other effect measures), and a vector of strategy names as inputs and returns a table of CEA results including total costs, total QALYs, incremental costs, incremental QALYs (or other effect measures), the ICER, and status of dominance.

```
decision_tree_HVE_cea <- calculate_icers(cost      = df_output$Cost,
                                         effect    = df_output$Effect,
                                         strategies = df_output$Strategy)

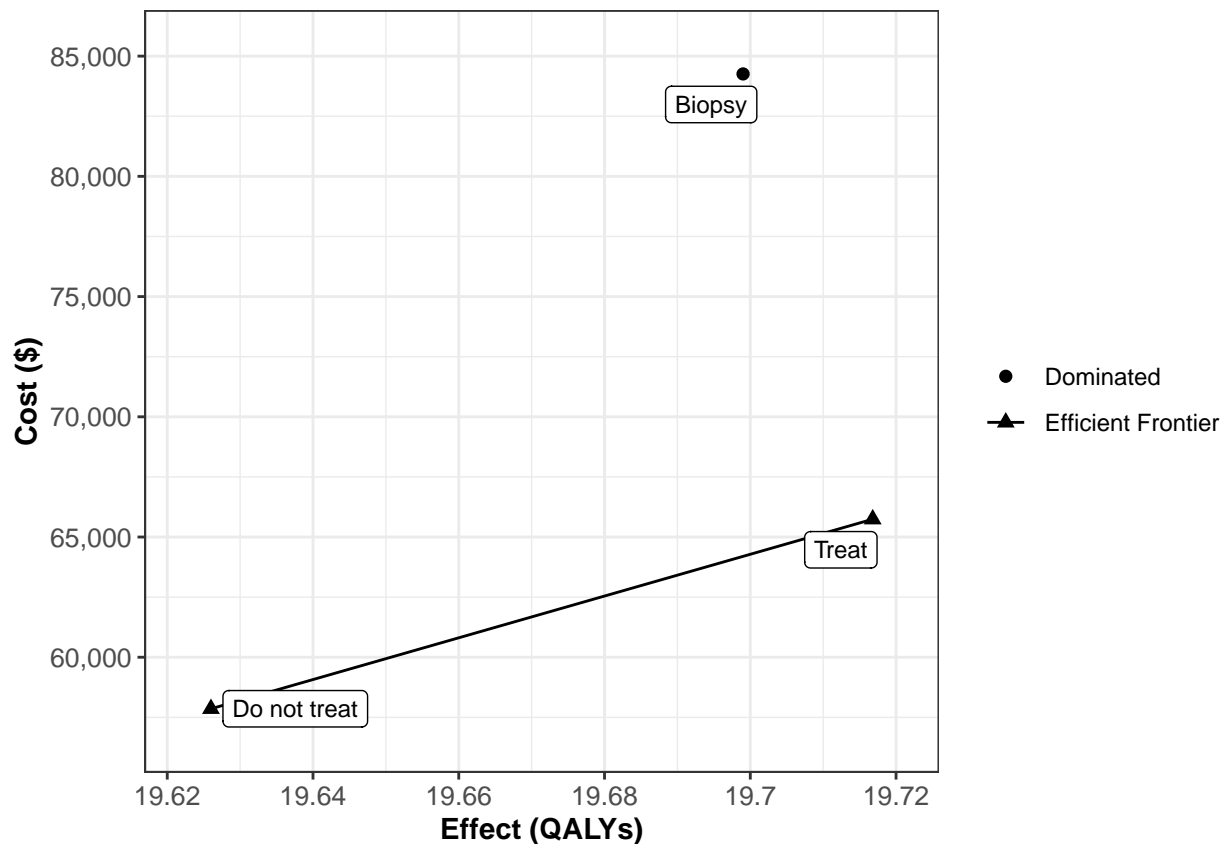
decision_tree_HVE_cea
```

##	Strategy	Cost	Effect	Inc_Cost	Inc_Effect	ICER	Status
## 1	Do not treat	57857.20	19.62600	NA	NA	NA	ND
## 2	Treat	65740.96	19.71680	7883.76	0.0908	86825.55	ND
## 3	Biopsy	84259.81	19.69901	NA	NA	NA	D

8. Visualizing the Results

The output of `calculate_icers` can be easily plotted using the `plot` function along with optional arguments like `effect_units` and `label`. The cost-effectiveness frontier is plotted.

```
plot(decision_tree_HVE_cea, effect_units = "QALYs", label = "all")
```



9. Acknowledgements

We kindly request you to add the following Acknowledgement paragraph to your further work where DARTH code formed the basis. We also like to remind you that you can add other sources of reference to this paragraph to acknowledge code you got from others.

Acknowledgement

For this work we made use of the template developed by the Decision Analysis in R for Technologies in Health (DARTH) workgroup: <http://darthworkgroup.com>.

The notation of our code is based on the following provided framework and coding convention: Alarid-Escudero, F., Krijkamp, E., Pechlivanoglou, P. et al. A Need for Change! A Coding Framework for Improving Transparency in Decision Modeling. *PharmacoEconomics* 37, 1329–1339 (2019). <https://doi.org/10.1007/s40273-019-00837-x>.

Other work from DARTH can be found on the website: <http://darthworkgroup.com/publications/>

Copyright for assignment work

Copyright 2017, THE HOSPITAL FOR SICK CHILDREN AND THE COLLABORATING INSTITUTIONS. All rights reserved in Canada, the United States and worldwide. Copyright, trademarks, trade names and any and all associated intellectual property are exclusively owned by THE HOSPITAL FOR Sick CHILDREN and the collaborating institutions. These materials may be used, reproduced, modified, distributed and adapted with proper attribution.