Introduction to EBASS

2016-10-01

Introduction

The EBASS package provides functions to compute sample size calculation for cost-effectiveness studies.

Our sample size calculation method uses a frequentist approach based on the expected value of perfect information (EVPI). The EVPI is an estimate of the cost of current uncertainty by accounting for both the probability that a decision based on existing evidence regarding the expected mean Incremental Net Monetary Benefit (INMB) is wrong (i.e: $P(INMB \geq 0 \vee \mu_{inmb} < 0)$ or $P(INMB < 0 \vee \mu_{inmb} \geq 0)$) and for the magnitude of the consequences of making the wrong decision for the target population.

The EVPI can be defined as the expected opportunity loss from a bad decision avoided with perfect information.

$$EVPI = N \left[I(\mu_{inmb} > 0) \int_{-\infty}^{0} -bf(b)db + I(\mu_{inmb} \leq 0) \int_{0}^{+\infty} bf(b)db \right]$$

Where f is the density function of INMB which follows a normal distribution $\mathcal{N}(\mu_{inmb}, \sigma_{inmb}^2)$, N is the sample size of the target population and I is an indicator function.

The EVPI can then be used as a necessary requirement for determining the profitability of further research. Applying this decision rule, additional research should be considered only if the EVPI exceeds the cost of further research.

It is this decision rule, based both on the EVPI and the cost of research, that we use for determining the optimal sample size of a cost-effectiveness study. Our method is applicable to all cost-effectiveness studies comparing two independent groups.

Example

This example shows how was calculated the optimal sample size for a randomized trial-based cost-effectiveness analyses (CEA). The study objective was to compare telemedicine to face to face care in elderly patients with complicated chronic wounds in nursing homes. First, hypothesis are made for the incremental net monetary benefit (INMB), its variance, the size of the target population and the cost of inclusion of an additional participant. Then, sample size based on the EVPI is calculated. Each step for sample size calculation is explained in the sections below.

Lambda

Lambda is known as the willingness to pay. That is the ceiling cost-effectiveness ratio or the maximum acceptable cost of a unit of effectiveness. In this example, the lambda value is set to 20 000 euros per QALY.

```
library(EBASS)
object_lambda <- create_object_lambda (20000)
```

INMB estimate

The net monetary benefit (NMB) of an intervention is given by $E \times lambda - C$ where E and C are the effectiveness and cost of this intervention. When the NMB is positive, the value of the intervention's effectiveness overpasses its cost. When evaluating the cost-effectiveness of a new intervention in comparison with the reference, one can estimate the difference of NMB between the new (or experimental) intervention and the reference intervention. This difference is known as the incremental net monetary benefit (INMB), which is given by: $INMB = NMBexp - NMBref = de \times lambda - dc$. In this example, the difference of effectiveness (de) between telemedicine and face to face is expected to be 0.04 QALY and the difference of cost (dc) to be -168 euros.

```
object_inmb <- create_object_inmb(de = 0.04, dc=-168, object_lambda = object_lambda)
object_inmb$get_inmb()
```

```
## [1] 968
```

Given these elements, the expected INMB in the example is 968 euros. Without these elements, it is also possible to make an hypothesis on the value of the INMB with the 'create_object_inmb_direct' function :

```
object_inmb_direct <- create_object_inmb_direct(inmb = 968)
```

INMB variance

If data are available this variance can be calculated based of the common standard deviation of costs (σ_c), the common standard deviation of effectiveness (σ_e), the willingness to pay lambda and the coefficient of correlation (ρ) between the difference in costs (dc) and the difference in effectiveness (de) :

$$\sigma_{inmb} = 2 \times (\lambda^2 \sigma_e^2 + \sigma_c^2 - 2\lambda\rho\sigma_e^2\sigma_c^2)$$

In this example, the common standard deviation of costs is 2100 euros, the common standard deviation of effectiveness is 0.12 QALY and the coefficient of correlation is 0.1

```
object_var_inmb <- create_object_var_inmb(sde=0.12, sdc=2100, rho=0.1, object_lambda=object_lambda)
object_var_inmb$get_var_inmb()
```

```
## [1] 18324000
```

When absolutely no data regarding the variability of costs and effectiveness are available, it is possible to provide a value for the variance of the Incremental Net Monetary Benefit by using the "create_object_var_inmb_direct" function. If the standard deviation of costs and effectiveness differ in each group, one can use the "create_object_var_inmb_diff" function.

Note that both INMB and its variance use the lambda value in their formula. Modifying the lambda value will affect both. That's why the lambda value was set in its own object and passed by arguments to 'create_object_inmb' and 'create_object_var_inmb' functions.

```
# Modify the lambda value
object_lambda$set_lambda(10000)
# See that inmb and var_inmb were affected by this modification :
cat ("The new INMB value is :",object_inmb$get_inmb(), "and its variance :",
     object_var_inmb$get_var_inmb())
```

```
## The new INMB value is : 568 and its variance : 10692000
```

```
# reset the value to 20 000
object_lambda$set_lambda(20000)
```

Target population

The expected value of perfect information (EVPI) is estimated for the entire population targeted by the evaluated intervention. The size of the target population (POP) can be estimated through prevalence and incidence data from registries, large cohort studies, medico-administrative databases, or surveillance systems. It is usually easier to gather data on the annual number of individual susceptible to benefit for the new intervention. If the time horizon is longer than one year, POP has to be discounted. In this example, 52 000 persons can benefit from the new experimental intervention each year during 20 years. The annual discount rate in France is 4%.

```
object_pop <- create_object_pop(horizon = 20, discount=0.04, N_year = 52000)
```

EVPI decrease

Given three objects created above : object_pop, object_inmb and object_var_inmb, it's now possible to calculate the expected value of perfect information that will remain after the inclusion of n participants in the study. For example, for 300 and 302 participants :

```
object_evpi_decrease <- create_object_evpi_decrease(object_inmb, object_pop, object_var_inmb)
# As the allocation ratio is 1:1, setting a number of 150 participants
# in the experimental group means we will have the same number in the reference group
cat ("EVPI remaining after the inclusion of 300 participants : ",
    object_evpi_decrease$set_N_exp(150), "\n",
    "EVPI remaining after the inclusion of 302 participants : ",
    object_evpi_decrease$set_N_exp(151))
```

```
## EVPI remaining after the inclusion of 300 participants :   216457
##  EVPI remaining after the inclusion of 302 participants :   209209.2
```

The more individuals are included, the less EVPI will remain at the end of the study. $EVPI(n)$ is a decreasing function and the EVPI decrement can be calculated as follows : $EVPI_n - EVPI_{n+STEPn}$ where $STEP_n$ is the incremental step of n : $STEP_n = STEP_{ref} + STEP_{exp}$. The minimal numbers of patients to be included in the reference group ($STEP_{ref}$) and in the experimental group ($STEP_{exp}$) are arguments of the 'create_object_evpi_decrease' function and are set to 1 by default in ordrer to respect the allocation balance.

```
# EVPI_300 - EVPI_302
cat ("EVPI decrement between the inclusion of 302 and 300 participants :  ",
    object_evpi_decrease$set_N_exp(150) - object_evpi_decrease$set_N_exp(151))
```

```
## EVPI decrement between the inclusion of 302 and 300 participants :   7247.758
```

7247.76 euros is the value of the information gained throught the inclusion of 2 additional participants in the study (participants number 301 and 302). As the cost of inclusion in this study is constant and worths 2257.25 euros, the value of information gained through the inclusion of individuals 301 and 302 is higher than the cost of their inclusion.

```
cat((object_evpi_decrease$set_N_exp(150) - object_evpi_decrease$set_N_exp(151)) - 2 * 2257.25)
```

2733.258

Sample size

The optimal sample size of our planned cost-effectiveness study is the smallest n for which the information gained through the inclusion of additional individuals (the decrease in the EVPI) is lower than or equals the cost of their inclusion. Formally, no more participants should be included when $EVPI_n - EVPI_{n+STEPn} \leq STEP_n \times cost_{indiv}$ or $EVPI_n - EVPI_{n+STEPn} - STEP_n \times cost_{indiv} \leq 0$ where $cost_{indiv}$ is the cost of including an additional participant. The function 'sample_size' determines the optimal sample size based on this equation :

```
cost_indiv <- 2257.25
n_subject <- sample_size(object_evpi_decrease = object_evpi_decrease, cost_indiv =  cost_indiv)
```

optimal number of participants to include in the study :   328
optimal number of participants to include in the experimental group : 164
optimal number of participants to include in the reference group : 164

Indeed, the difference between information gained and costs of inclusion is still positive for participants 327 and 328 but turns negative from participant 329. The function 'graph_gain_n' produces a plot to explain this.
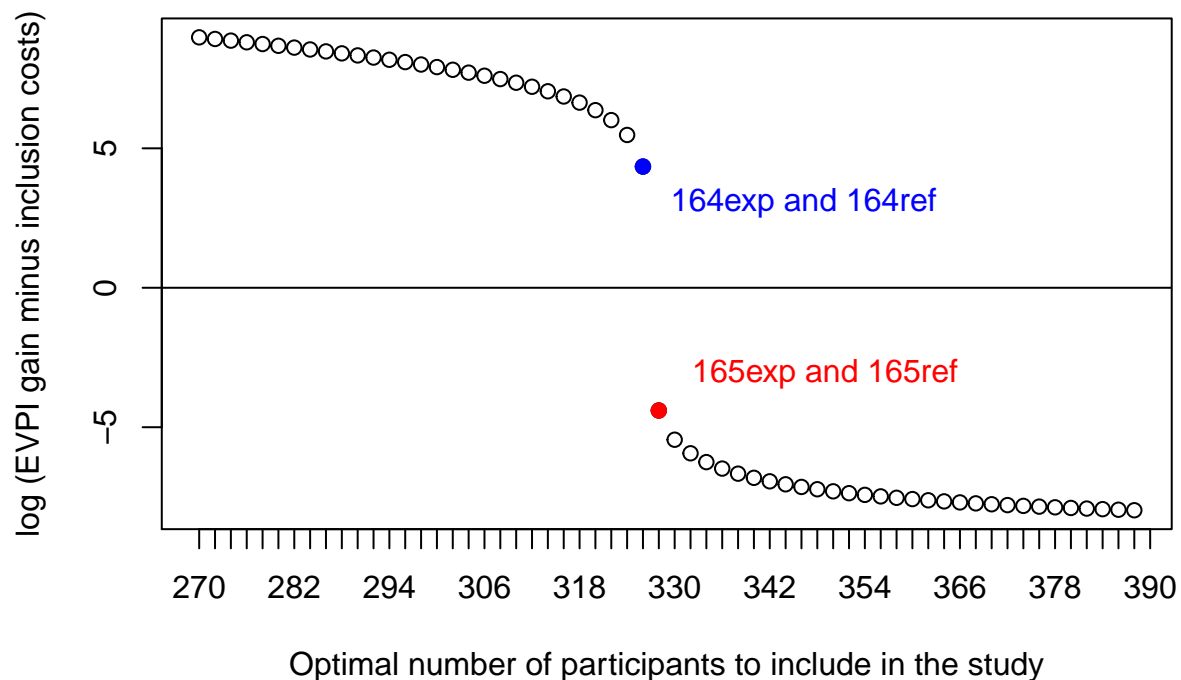
```
cat("difference between information gain and inclusion costs for participants 327 and 328 :",
(object_evpi_decrease$set_N_exp(163) - object_evpi_decrease$set_N_exp(164)) - 2 * cost_indiv)
```

difference between information gain and inclusion costs for participants 327 and 328 : 76.17842

```
cat("difference between information gain and inclusion costs for participants 329 and 330 :",
(object_evpi_decrease$set_N_exp(164) - object_evpi_decrease$set_N_exp(165)) - 2 * cost_indiv)
```

difference between information gain and inclusion costs for participants 329 and 330 : -80.45154

```
graph_gain_n(object_evpi_decrease, cost_indiv)
```

log (EVPI gain minus inclusion costs)

164exp and 164ref

165exp and 165ref

Optimal number of participants to include in the study

```
## Gain with 328 participants (164exp/164ref) : 76.17842
## Gain with 330 participants (165exp/165ref) : -80.45154
```

gamma risk

The gamma risk is the probability that a decision based on the expected mean of the Incremental Net Monetary Benefit (INMB) is wrong. After the inclusion of 328 participants :

```
object_evpi_decrease$set_N_exp(164)
gamma_risk(object_evpi_decrease = object_evpi_decrease)
```

```
## [1] 0.001890221
```

Simulation

What if we want to calculate the sample size for different hypothesis for time horizon for example ? Each object has setter functions to modify the hypothesis.

```
# calculate samples size for time horizon between 1 and 20 years
time_horizon <- seq (1:20)
Nindividuals <- sapply(time_horizon, function(x){
  object_pop$set_horizon(x)
  return (sample_size(object_evpi_decrease, cost_indiv)$N)
})
# plot the result
plot(time_horizon, Nindividuals, xlab="Time horizon in years",
     ylab="Number of individuals to include in the study",xaxt="n", type="b", pch=20)
axis(side=1, at=time_horizon, labels=as.character(time_horizon))
```
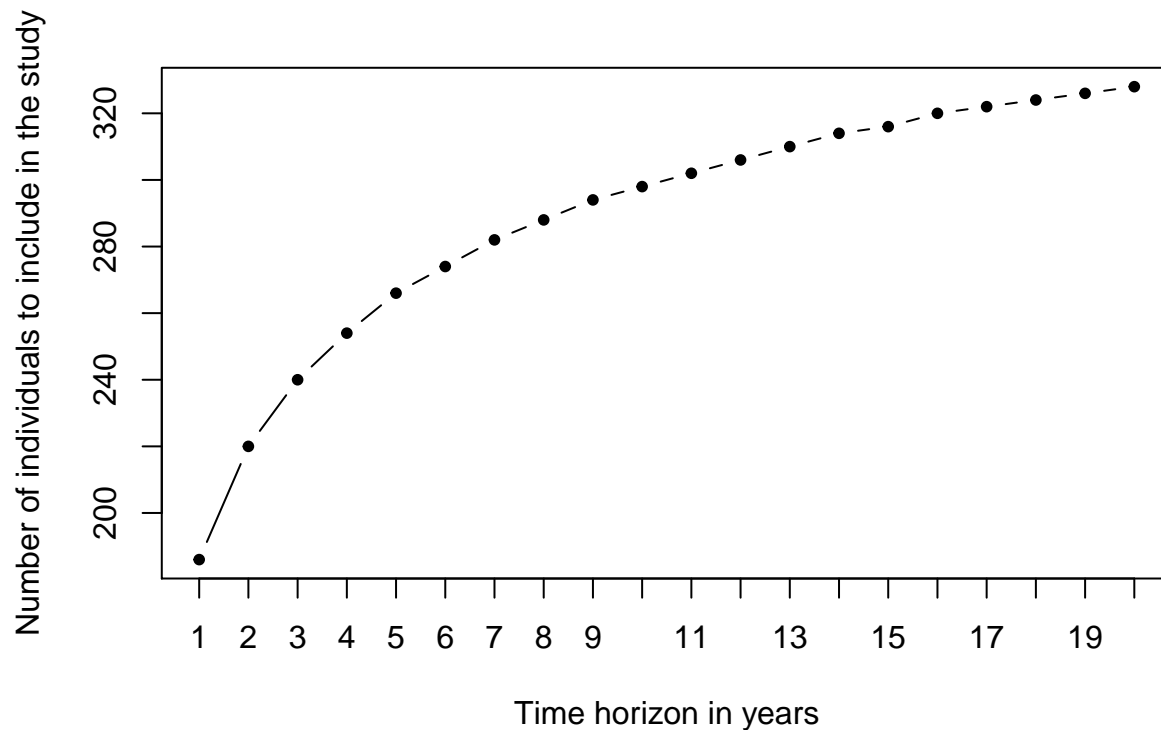
Code summary

```
# results hidden
library(EBASS)
object_lambda <- create_object_lambda (20000)
object_inmb <- create_object_inmb(de = 0.04, dc=-168, object_lambda = object_lambda)
object_var_inmb <- create_object_var_inmb(sde=0.12, sdc=2100, rho=0.1, object_lambda=object_lambda)
object_pop <- create_object_pop(horizon = 20, discount=0.04, N_year = 52000)
object_evpi_decrease <- create_object_evpi_decrease(object_inmb, object_pop, object_var_inmb)
cost_indiv <- 2257.25
n_subject <- sample_size (object_evpi_decrease, cost_indiv)
graph_gain_n (object_evpi_decrease, cost_indiv)
```

```
gamma_risk (object_evpi_decrease)
time_horizon <- seq (1:20)
Nindividuals <- sapply(time_horizon, function(x){
  object_pop$set_horizon(x)
  return (sample_size(object_evpi_decrease, cost_indiv)$N)
})
plot(time_horizon, Nindividuals, xlab="Time horizon in years",
     ylab="Number of individuals to include in the study",xaxt="n", type="b", pch=20)
axis(side=1, at=time_horizon, labels=as.character(time_horizon))
```