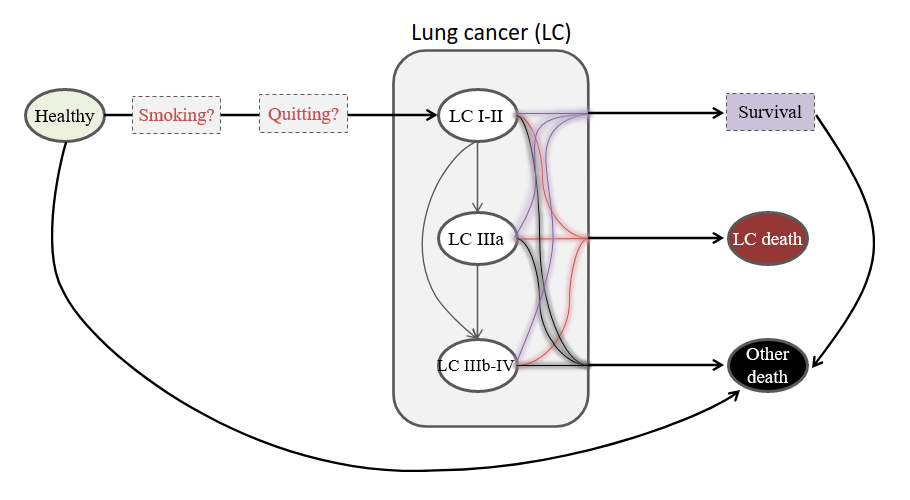
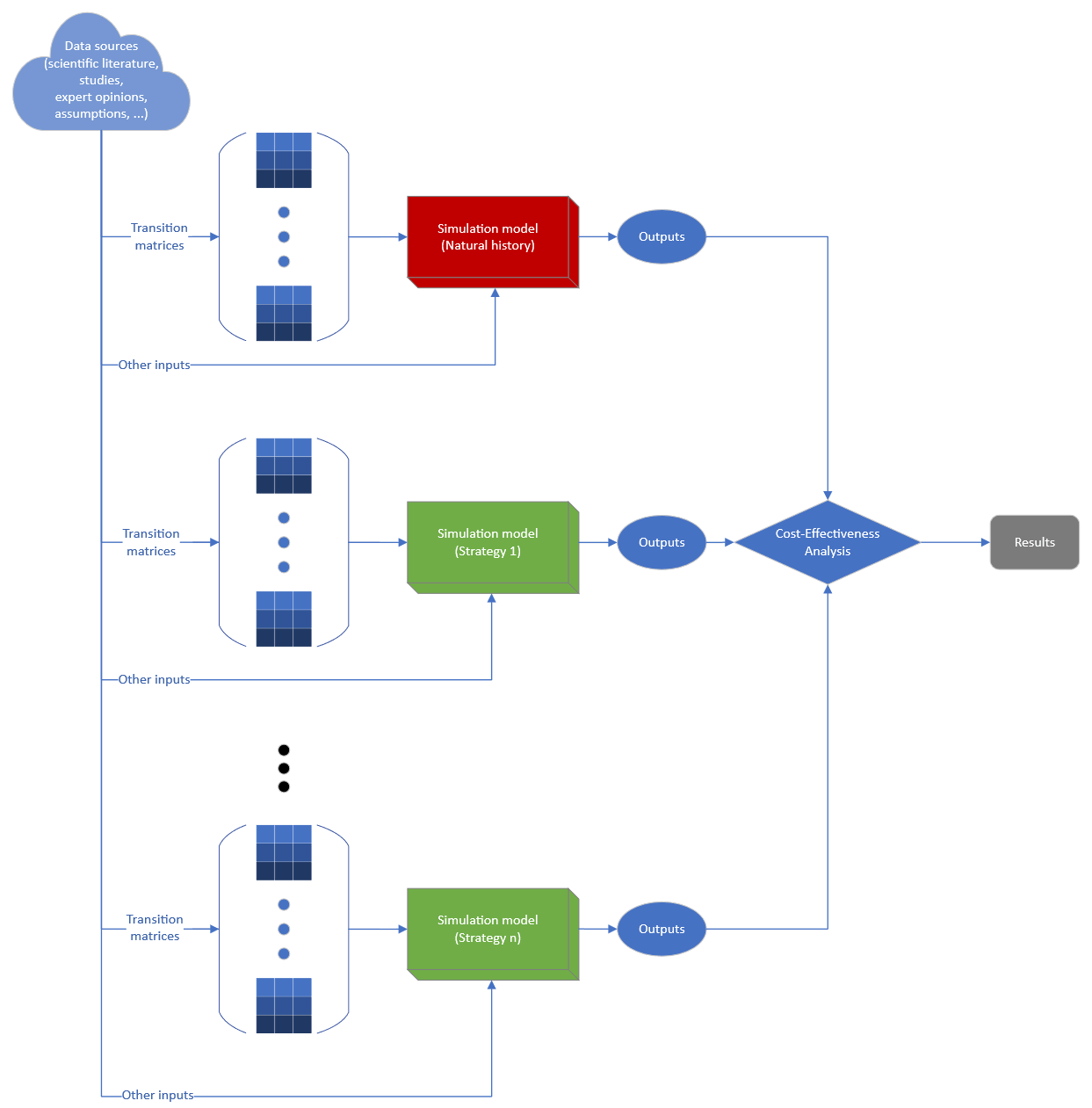
Calibration of cost-effectiveness models

# Background: Cost-Effectiveness Models in Healthcare

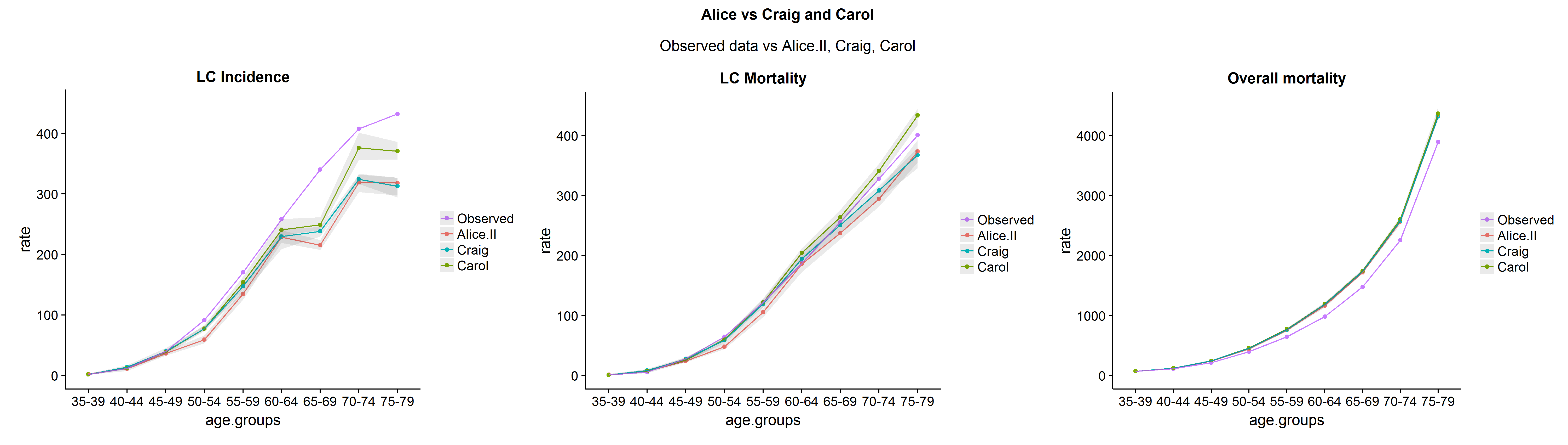
* **Objective**: Compare different strategies for detection/treatment of a disease, from health and economic point of view.
* **Methodology**: Simulation model to mimic the strategies and compare the outputs for each strategy to determine which strategies are worth considering. A special strategy called the natural history describes the progression of the disease without any planned interventions, and it is used to calibrate some the inputs that will be used in the rest of strategies (see *calibration* below).
* **Type of model**: decision tree, markov model, microsimulation, …



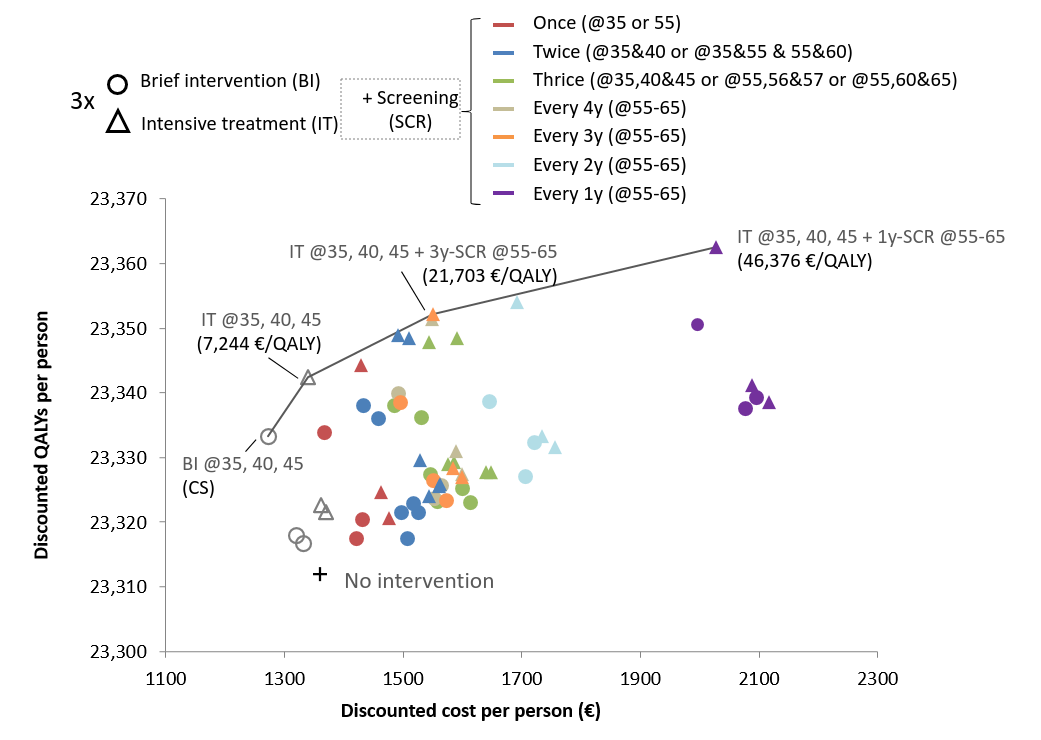
* **Input**: Parameters extracted from the scientific literature, studies, expert opinions, assumptions, … An interesting intermediate input are the transition matrices that show the probabilities of transitioning between health states.
* **Outputs**: For each strategy:
  + Effectiveness measure (e.g. **Quality-Adjusted Life Years, QALYs**)
  + Cost measure (e.g. euros, €)
  + Other general measures of interest: incidence, mortality, …
  + Other domain-dependent measures: e.g. number of hysterectomies, number of high-grade lesions, …



1. **Usual methodology**
2. **Calibration**: Before starting the base analysis, we calibrate the transition matrices in the natural history by slightly modifying the original probabilities so that the output of our model (e.g. incidence, mortality, …) fits an observed value based on evidence. These calibrated probabilities can then be used by the rest of the strategies in the base analysis. See *Calibration Workflow* for more details.

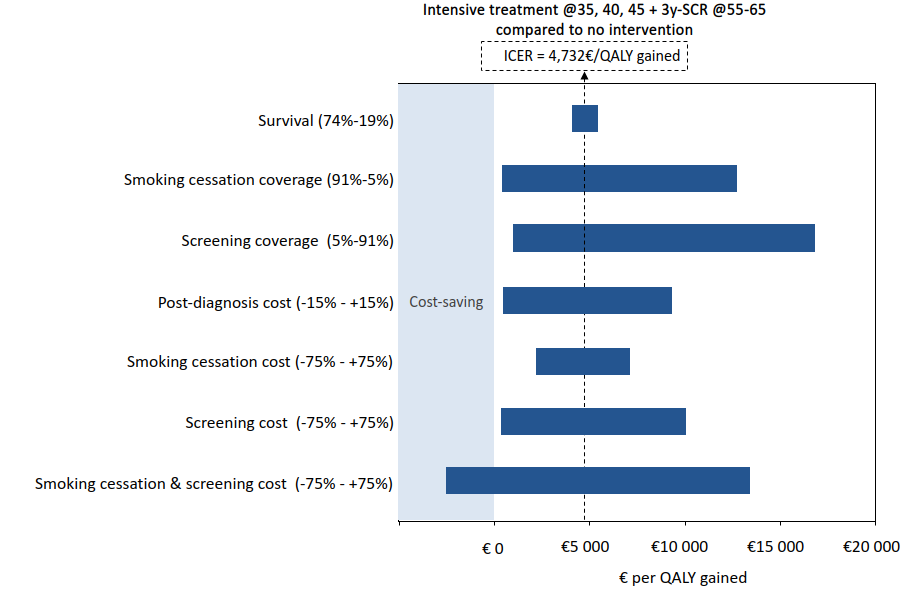


1. **Base analysis**: Each strategy is plotted in the Cost and Effectiveness axes and the efficiency curve shows the strategies that are cost-effective, the rest are dominated by them and they are not considered cost-effective.

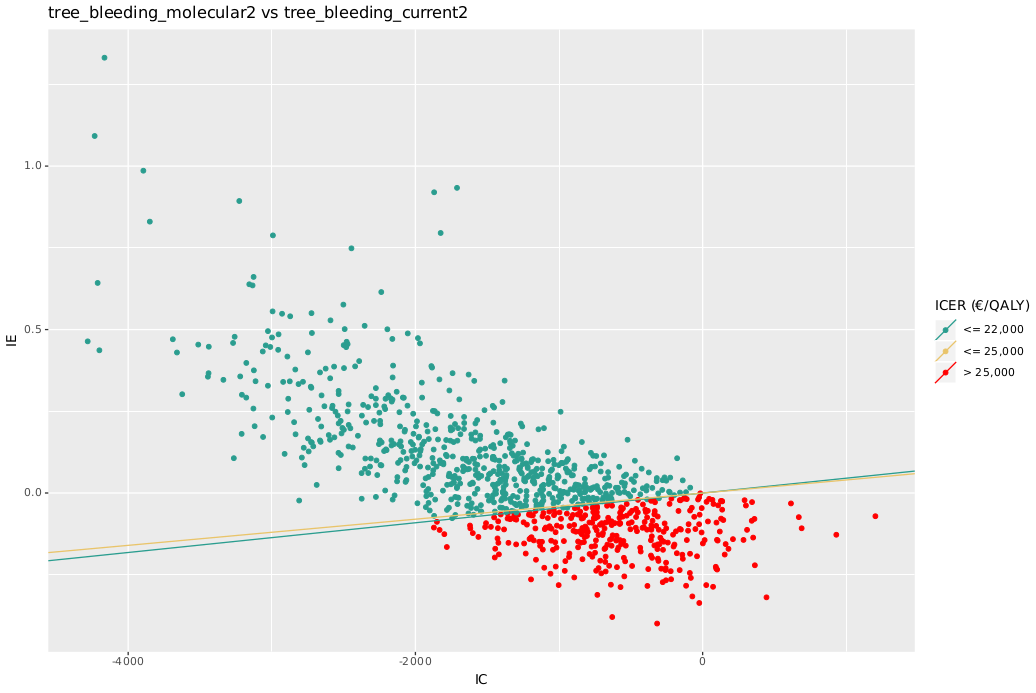


We can compare each strategy in relation to another calculating the Incremental Cost-Effectiveness Ratio as . If the ICER is below the Willingness-To-Pay (**WTP**) threshold (i.e. the maximum amount of money a country/region is willing to pay per additional QALY) the second strategy is more cost-effective than the first. If the ICER is greater than the WTP the strategy’s benefits are not considered cost-effective (i.e. the increased health benefit does not justify the increment of cost). Negative ICERs imply that one strategy dominates the other one.

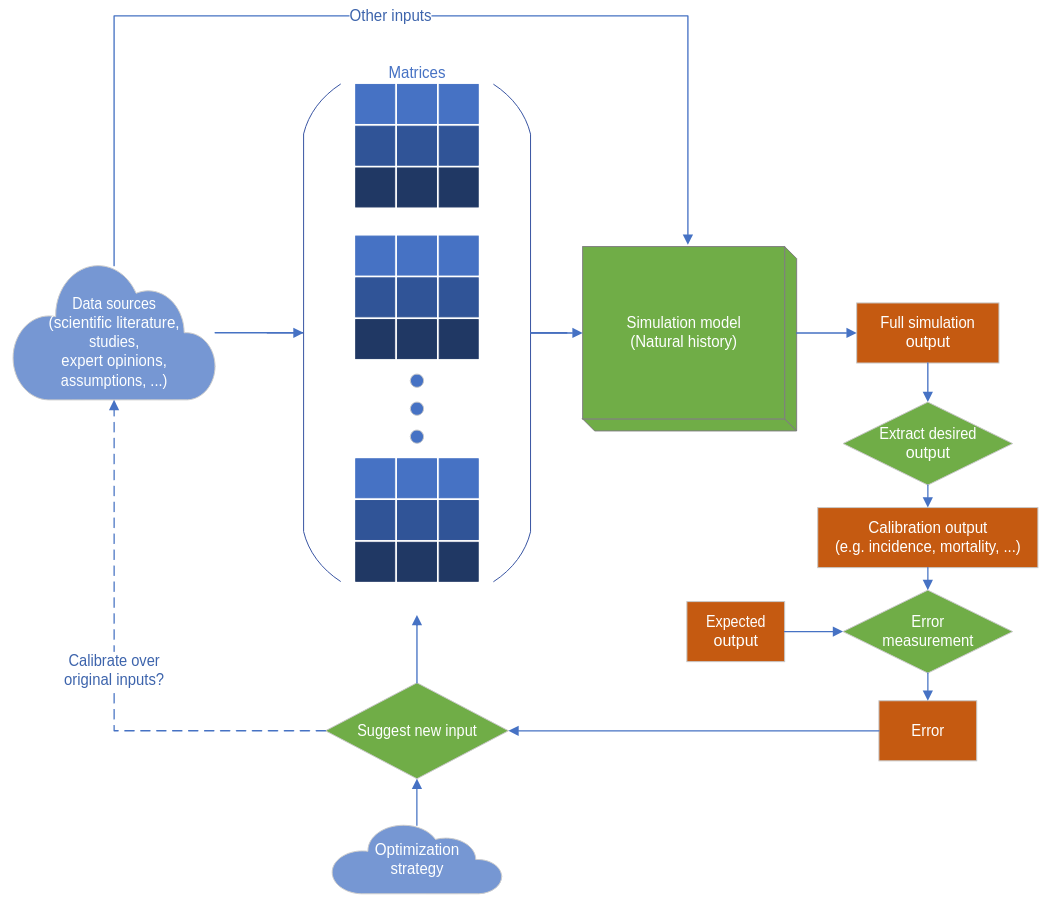
1. **Sensitivity analysis**: Once the base analysis is performed we evaluate the uncertainty of the used parameters to check the robustness of the results. We modify the values of the parameters of interest to see how they affect the output of the model.
   1. **Deterministic sensitivity analysis (DSA):** A sweep is performed over a range (e.g. +-15% of the base value) for the parameters of interest, to see how the ICER changes and whether the cost-effectiveness decision is different.



* + **Probabilistic sensitivity analysis (PSA):** Each parameter of interest is modeled as a probabilistic distribution (e.g. Beta for probabilities, Gamma/Lognormal for costs, …) with the base value as the mean and a standard deviation dependent on the amount of uncertainty. We sample from these distributions (univariate or multivariate) to run a number of random simulations to check the percentage of simulations that show a cost-effective result (i.e. the percentage of simulations below the line ICER=WTP).



# Calibration workflow



# Test model: Lung cancer (LC)

* Total of 9 matrices (one per age group: 35-39, 40-44, …, 75-79).
* 7 health states: healthy, stages I-II, stage IIIa, stage IIIb, LC survival, death from LC, death from other causes → 7x7 matrices. Sometimes the LC survival state is excluded from the calibration resulting in 6x6 matrices.
* Matrices represent monthly steps in the simulation. Since they are applied for 5-year groups, each matrix is used in 5 \* 12 = 60 iterations in the model.
* A **simplified calibration** can be performed without running the model, only the matrices are used. This is a fast approximation since we are not considering some factors of the full model (e.g. prevalence of smoking):
* LC survival state is excluded → 6x6 matrices
* From the 6x6=36 probabilities per matrix, only 11 probabilities are allowed to change. The rest are either constant (zeroes, ones) or one minus the sum of the rest of the row.
* The error measurement is a weighted sum of the absolute differences of the LC incidence, LC mortality and mortality from other causes. The weights are 0.45, 0.45 and 0.10 respectively.

## Test #1: Simplified lung model calibration using only first matrix

* Source file: *models/lung/calibration\_wrapper.R* (N\_MATRICES set to 1)
* Only the first age group is being calibrated (35-39): 1x11 = **11 parameters**.
* Results:

|  |  |  |
| --- | --- | --- |
| **Algorithm** | **Nelder-Mead** | **Bayesian Optimization** |
| **Best solution** | [1.11792560e-06  4.40025000e-05  2.85500424e-02  5.58688151e-01  2.33054382e-02  2.56198878e-06  6.26189872e-02  2.35346590e-02  1.89438555e-06  2.33023545e-02  1.84300020e-06] | [1.11743209e-06  4.40025000e-05  3.49317322e-02  3.52614061e-01  2.33019105e-02  2.16691398e-06  9.78486120e-02  2.33015791e-02  2.81935900e-06  2.33023544e-02  1.69119083e-06] |
| **Error** | 0.6633085653748201 | 0.6629851533965986 |
| **Time (s)** | 0.7594297569958144 | 114.89221513200027 |
| **Model evaluations** | 252 | 21 |

## Test #2: Simplified lung model calibration using first and second matrices

* Source file: *models/lung/calibration\_wrapper.R* (N\_MATRICES set to 2)
* The first and second age groups are being calibrated (35-39 and 40-44): 2x11 = **22 parameters**.
* Results:

|  |  |  |
| --- | --- | --- |
| **Algorithm** | **Nelder-Mead** | **Bayesian Optimization** |
| **Best solution** | [1.11772134e-06  4.42306248e-05  2.85017410e-02  3.45979312e-01  2.33460108e-02  2.57661105e-06  1.02035604e-01  3.28843601e-02  2.83664387e-06  2.33070955e-02  1.44304635e-06  7.87229723e-06  7.22628698e-05  1.36209733e-02  3.68460579e-01  2.95497778e-02  1.71769333e-06  6.01628601e-02  3.01599686e-02  2.48081674e-06  2.36671750e-02  1.08325532e-06] | [1.11743209e-06  4.40025000e-05  3.29773169e-02  5.01644723e-01  2.33019105e-02  2.37534043e-06  1.04074200e-01  2.33015791e-02  2.95611906e-06  2.33023544e-02  1.77700528e-06  7.86577500e-06  7.22100000e-05  1.76876690e-02  4.21488733e-01  3.07201770e-02  1.69078343e-06  6.25333450e-02  3.88364300e-02  1.95655409e-06  2.33027557e-02  8.71881206e-07] |
| **Error** | 0.7333381543348456 | 0.7287116136105645 |
| **Time (s)** | 4.444447371002752 | 421.6934125780026 |
| **Model evaluations** | 2092 | 70 |

## Test #3: Simplified lung model calibration using all matrices

* Source file: *models/lung/calibration\_wrapper.R* (N\_MATRICES set to 9)
* All age groups are being calibrated: 9x11 = **99 parameters**.
* Standard bayesian optimization takes too much time and the process was aborted before completion. Other strategies could be attempted: calibrate matrices sequentially, restrict number of parameters, optimize gaussian process regression (see phd\_proposal.docx), ...
* Results:

|  |  |  |
| --- | --- | --- |
| **Algorithm** | **Nelder-Mead** | **Bayesian Optimization** |
| **Best solution** | \* | <Aborted due to excessive computation time> |
| **Error** | 4.021066119701601 | - |
| **Time (s)** | 57.58928211599414 | - |
| **Model evaluations** | 19800 | - |