| **Test description (Please document how the test is conducted, as well)** | **Expected result of the test** | **Observed result of the test** | **Action taken** |
| --- | --- | --- | --- |
| *Pre-analysis calculations* | | | |
| Does the technology (drug/device, etc.) acquisition costs increase with higher prices? | Yes | Yes.  Change input parameter for treatment costs (“tx\_cost\_input”) and total costs increase. | None. |
| Does the drug acquisition cost increase for higher weight or body surface area? | Yes | Not applicable. | None. |
| Does the probability of an event, derived from an odds ratio (OR)/ relative risk (RR) / hazard ratio (HR) and baseline probability, increases with higher OR/RR/HR? | Yes | Not applicable.  Event probabilities are applied directly without applying any OR, RR or HR. Treatment effects are implemented as absolute decrements in baseline characteristics like HbA1c or cholesterol. Changing these characteristics would change the event probabilities. | None. |
| If survival parametric distributions are used in the extrapolations, can the formulae used for the Weibull (generalized gamma) distribution generate the values obtained from the exponential (the Weibull or Gamma) distribution(s) under some parameter transformations? | Yes | Partially.  Survival parametric distributions are fixed in the UKPDS equations, and so are in our model. When an exponential distribution was assumed to model some events, a Weibull equation with shape parameter = 1. | None. |
| In a partitioned survival model, does the progression free survival curve or the time on treatment curve crosses the overall survival curve? | No | Not applicable.  The model is not PSM. | None. |
| If survival parametric distributions are used in the extrapolations or time-to-event calculations, can the formulae used for the Weibull (generalized gamma) distribution generate the values obtained from the exponential (the Weibull or Gamma) distribution(s) after replacing/transforming some of the parameters? | Yes | Yes.  This happens for the Weibull and Exponential distributions as mentioned above. | None. |
| Is hazard ratio calculated from Cox proportional hazards model applied on top of the parametric distribution extrapolation found from the survival regression? | No, it is better if the treatment effect that is applied to the extrapolation comes from the same survival regression in which the extrapolation parameters are estimated. | Not applicable.  Cox PH model not applied. | None. |
| For the treatment effect inputs, if the model uses outputs from WINBUGs, are the OR, HR and RR values all within plausible ranges? (should be all non-negative and the average of these WINBUGs outputs should give the mean treatment effect) | Yes | Not applicable.  As mentioned above, treatment effects are applied as absolute decrements in baseline characteristics like HbA1c or cholesterol. | Checks should be included to make sure that the units lie within the logical range of values. |
| *Event-state calculations* | | | |
| Calculate the sum of the number of patients at each health state | Should add up to the cohort size | Not applicable.  This is a patient-level model, so individual patients are run through the simulation. | None. |
| Check if all probabilities and number of patients in a state are greater than or equal to zero | Yes |  | This can be partially tested for example with the file “validation\_ukpds.R”. However, since equations depend on patent characteristics, this leaves a potentially infinite number of equations. We might use lower and upper bounds for patient characteristics. |
| Check if all probabilities are smaller than or equal to one | Yes |  | This can be partially tested for example with the file “validation\_ukpds.R”. However, since equations depend on patent characteristics, this leaves a potentially infinite number of equations. We might use lower and upper bounds for patient characteristics. |
| Compare the number of dead (or any absorbing state) patients in a period with the number of dead (or any absorbing state) patients in the previous periods? | Should be larger | Not applicable.  This is not a stat/transition model. | None. |
| In case of lifetime horizon, check if all patients are dead at the end of the time horizon | Yes | Not applicable.  This is a patient-level simulation model and the simulation stops when patients die. | None. |
| *Discrete event simulation specific:* sample one of the “time to event” types used in the simulation from the specified distribution. Plot the samples and compare the mean and the variance from the sample | Sample mean and variance & the simulation outputs should reflect the distribution it is sampled from. |  | This can be partially tested for example with the file “validation\_ukpds.R”. However, since equations depend on patent characteristics, this leaves a potentially infinite number of equations. We might use lower and upper bounds for patient characteristics. |
| Set all utilities to one  Set all utilities to zero | The QALYs accumulated at a given time would be the same as the life years accumulated at that time  No utilities will be accumulated in the model | Input utilities set to 1 for all ages and gender, and decrements set to 0. No, there seems to be an error here since QALYs are higher than life expectancy. The difference seems to be always 0.5 years, so probably it has to do with the assumption that the initial duration of the treatment is 6 months.  Yes, setting utilities to 0 works as expected. | Utility decrements are now read from a csv file (they were “hardcoded”).  Explore and correct. |
| Decrease all state utilities simultaneously (but keep event-based utility decrements constant) | Lower utilities will be accumulated each time | Unclear how to test this since we do not have state utilities. Decreasing age/gender-dependent utilities works as expected. | None. |
| Set all costs to zero | No costs will be accumulated in the model at any time | Change different input cost files (complications, future costs, informal care and productivity).  Results as expected. | Informal care and productivity costs input parameters are now read from a csv file (they were “hardcoded”). **TO DO**. |
| Put mortality rates to 0 | Patients never die | The model works with annual probabilities. These can be set manually to 0 in the code (but not as input parameters). Patients die in the simulation when they become 99 as this was set as a condition in the simulation. | None. |
| Put mortality rate extremely high | Patients die in the first few cycles | The model works with annual probabilities. When these are set manually to high values (e.g., 0.9) in the code, the code returns an error when sampling from binomial distributions.  To test this assumption, we replace the object “current\_DEATH\_prob” by a large number in these lines of code:  # Sampling "dead" status  current\_DEATH\_event <- rbinom(1, 1, current\_DEATH\_prob)  This works as expected. | Unclear if we have to do something. |
| Set the effectiveness, utility and safety related model inputs for all treatment options equal | Same life years and QALYs should be accumulated for all treatment at any time | This is achieved by assuming no treatment effect input parameters (i.e., by running the “comparator” arm). | None. |
| In addition to the inputs above, set cost related model inputs for all treatment options equal | Same costs, life years and QALYs should be accumulated for all treatment at any time | Same as above + equal treatment costs. | None. |
| Change around the effectiveness, utility and safety related model inputs between two treatment options | Accumulated life years and QALYs in the model at any time should be also reversed | Not relevant for our model since effectiveness is modelled “on top” of a reference scenario. | None. |
| Check if the number of alive patients estimate at any cycle is in line with general population life table statistics | At any given age, the % alive should be lower or equal in comparison to the general population estimate | Not relevant for this model but it can be checked what’s the average life expectancy for the general population at the same (average) baseline age. | Check life expectancy for the general population. |
| Check if the QALY estimate at any cycle is in line with general population utility estimates | At any given age, the utility assigned in the model should be lower or equal in comparison to the general population estimate | Not applicable.  This is the case by definition. Utilities have been calculated for the diabetes population. Therefore, it was already assumed that these were lower than those for the general population. | None. |
| Set the inflation rate of the previous year higher | The costs (which are based on a reference from previous years) assigned at each time will be higher | Unsure how to test this. | Unclear. |
| Calculate the sum of all ingoing and outgoing transition probabilities | Both should be one. | Not applicable.  There are no transition probabilities. | None. |
| Calculate the number of patients entering and leaving a tunnel state throughout the time horizon | Numbers entering = Numbers leaving | Not applicable.  There are no health states. | None. |
| Check if the time conversions for probabilities were conducted correctly. | Yes |  |  |
| *Decision tree specific:* calculate the sum of the expected probabilities of the terminal nodes | Should sum up to one | Not applicable.  This is not a decision tree model. | None. |
| *Patient-level model specific:* check if common random numbers are maintained for sampling for the treatment arms? | Yes | Yes, random seeds are used for this purpose. | None. |
| *Patient-level model specific:* check if correlation in patient characteristics is taken into account when determining starting population? | Yes | Not applicable at this moment.  We don’t know yet how the relevant population will be modelled. | Unclear. |
| Increase the treatment acquisition cost | Costs accumulated at a given time will increase during the period when the treatment is administered | Yes, treatment costs are applied at the beginning of the simulation only. | None. |
| *Population model specific:* set the mortality and incidence rates to zero | Prevalence should be constant in time | Not applicable. | None. |
| *Result calculations* | | | |
| Check the incremental life years and QALYs gained results. Are they in line with the comparative clinical effectiveness evidence of the treatments involved? | If a treatment is more effective, it generally results in positive incremental LYs and QALYs in comparison with the less effective treatments | Yes, this is in line. However, we have noticed that the treatment effects that we have implemented so far have a really minor impact on the results. | Further exploration of the regression equations, for example in the validation UKPDS R file: for example, what’s a reasonable treatment effect and what’s the expected impact on the annual risk equation? |
| Check the incremental cost results. Are they in line with the treatment costs? | If a treatment is more expensive, and if it does not have much effect on other costs, it generally results in positive incremental costs. | Costs are closely linked to the costs of events and, therefore, linked to the effectiveness. Treatment costs are not really that important in this model. | None. |
| Total life years > total quality adjusted life years | Yes | Yes. | There was the issue of the first 0.5 years mentioned above. This needs to be resolved. |
| Undiscounted results > discounted results | Yes | Yes, but it seems that life expectancy is reported undiscounted all the time. | Not sure if we have to do something here. |
| Divide undiscounted total QALYs by undiscounted life years. | This value should be within the outer ranges (maximum and minimum) of the all utility value inputs. | Utilities are age and sex dependent; therefore, it is difficult to properly test this assumption. | None. |
| Subgroup analysis results: How do the outcomes change if the characteristics of the baseline change? | Better outcomes for better baseline health conditions and worse outcomes for worse health conditions are expected. |  | To do. |
| Could you generate all the results in the report from the model (including the uncertainty analysis results)? | Yes | At this moment, there is no report. | None. |
| Do the total life years, QALYs and costs decrease if a shorter time horizon is selected? | Yes | This is not applicable.  Simulations end when patients die or become older than 99 years old. This could be tested by forcing the code to stop after x years. This would certainly imply less costs and QALYs since only the first x years would be considered in the calculations. | None. |
| Is the reporting and contextualization of the incremental results correct? | The use of the terms such as: “dominant”/ “dominated”/ “extendedly dominated”/ “cost-effective” etc. should be in line with the results.  In the incremental analysis table involving multiple treatments, ICERs should be calculated against the next non-dominated treatment. | At this moment, there is no report. | None. |
| Are the reported ICERs in the fully incremental analysis non-decreasing? | Yes | At this moment, there is no report. | None. |
| If disentangled results are presented, do they sum up to the total results? (e.g. different cost types sum up to the total costs estimate) | Yes |  |  |
| Check if half cycle correction is implemented correctly (total life years with half cycle correction should be lower than without) | The half cycle correction implementation should be error free. Also check if it should be applied for all costs, for instance if a treatment is administered at the start of a cycle, half cycle correction might be unnecessary. | Not applicable in a patient-level model. | None. |
| Check the discounted value of costs/QALYs after 2 years | Discounted value=undiscounted/(1+r)2 | There is no way to test this in the model (unless we force the model to stop after 2 years). | None. |
| Set discount rates to zero | The discounted and undiscounted results should be the same | Yes. | None. |
| Set mortality rate to zero | The undiscounted total life years per patient should be equal to the length of the time horizon | This was already tested above. | None. |
| Put the consequence of adverse event/discontinuation to zero. (zero costs and zero mortality/utility decrements) | The results would be the same as the results when AE rate is set to zero. | Yes. | None. |
| Divide total undiscounted treatment acquisition costs by the average duration on treatment. | This should be similar to treatment related unit acquisition costs | Not applicable in the current version of the model. | None. |
| Set discount rates to a higher value | Total discounted results should decrease | ERROR: costs decrease but QALYs increase. |  |
| Set discount rates of costs/effects to an extremely high value | Total discounted results should be more or less the same as the discounted results accrued in the first cycles | Cannot be tested in a patient-level model.  However, we saw that QALYs still increase like in the previous test. | None. |
| Put adverse event/discontinuation rates to zero and then to extremely high level. | Less costs higher QALYS/LYs when adverse event rates are 0, higher costs and lower QALYS/LYs when AE rates are extreme | High rates: Higher complication costs and lower QALYs but TOTAL costs are lower because patients die earlier. | None. |
| Double the difference in efficacy and safety between new intervention and comparator and report the incremental results. | Approximately twice of the incremental effect results of the base case. If this is not the case: report and explain the underlying reason/ mechanism | This is not easy (not sure if possible) to test because efficacy is implemented through changes in patient characteristics that have an effect on the annual probability of experiencing an event. As such, I don’t think there is any way to “double” the difference in efficacy and safety. | None. |
| Do the same for a scenario in which the difference in efficacy and safety is halved. | Approximately halve of the incremental effect results of the base case. If this is not the case: report and explain the underlying reason/ mechanism | Same as the previous test. | None. |
| *Uncertainty analysis calculations* | | | |
| Are all parameters subject to uncertainty included in the one-way sensitivity analysis (OWSA)?  Check if the OWSA includes any parameters associated with joint uncertainty (e.g., parts of a utility regression equation, survival curves with multiple parameters). | Yes  No | No.  We’re not interested in OWSA as such (e.g., tornado diagram) but the focus will be on scenario analyses. | None. |
| Are the upper and lower bounds used in the one-way sensitivity analysis used confidence intervals based on the statistical distribution assumed for that parameter?  Are the resulting ICER, incremental costs/QALYs with upper and lower bound of a parameter plausible and in line with a priori expectations? | Yes  Yes | Not applicable. | None. |
| Check that all parameters used in the sensitivity analysis have appropriate associated distributions  - upper and lower bounds should surround the deterministic value (i.e., Upper bound ≥ mean ≥ Lower bound)  - standard error and not standard deviation used in sampling  - Lognormal / gamma distribution for hazard ratios and costs/ resource use  - Beta for utilities and proportions/probabilities  - Dirichlet for multinomial  - Multivariate normal for correlated inputs (e.g., survival curve or regression parameters)  - Normal for other variables as long as samples don’t violate requirement to remain positive when appropriate | Yes | To do.  There is a problem here: the parameters used for the regression equations: UKPDS only reports mean + SD but there is no covariance matrix. Without the covariance matrix, it is not possible to account for dependencies between the regression coefficients. | Investigate whether the covariance matrix is available in the UKPDS model. |
| Check PSA output mean costs, QALYs and ICER compared to the deterministic results. Is there a large discrepancy? | No (in general) |  |  |
| If you take new PSA runs from the excel model do you get similar results? | Yes |  |  |
| Is(are) the CEAC line(s) in line with the CE scatter plots and the efficient frontier? | Yes |  |  |
| Does the PSA cloud demonstrate an unexpected behaviour or has an unusual shape? | No |  |  |
| Is the sum of all CEAC lines equal to 1 for all WTP values? | Yes |  |  |
| Are the explored scenario analyses provide a balanced view on the structural uncertainty? (i.e., not always looking at more optimistic scenarios) | Yes |  |  |
| Are the scenario analysis results plausible and in line with a priori expectations? | Yes |  |  |
| If a certain seed is used for random number generation (or previously generated random numbers are used), check if they are scattered evenly between 0-1 when they are plotted? | Yes |  |  |
| Compare the mean of the parameter samples generated by the model against the point estimate for that parameter, use graphical methods to examine distributions, functions | The sample means and the point estimates will overlap, the graphs will be similar to the corresponding distribution functions (e.g., Normal, Gamma, etc.) |  |  |
| Check if sensitivity analyses include any parameters associated with methodological/ structural uncertainty (e.g., annual discount rates, time horizon). | No |  |  |
| Value of information analysis if applicable: Was this implemented correctly?  Which types of analysis? Were aggregated parameters used? Which parameters are grouped together? Does it match the write-up’s suggestions?  Is EVPI larger than all individual EVPPI?  Is EVPPI for a (group of) parameters larger than the EVSI of that (group) of parameter(s)?  Are the results from EVPPI in line with OWSA or other parameter importance analysis (e.g., ANCOVA)? | Yes |  |  |
| Did the electronic model pass the black-box tests of the previous verification stages in all PSA iterations and in all scenario analysis settings? (additional macro can be embedded to PSA code, which stops the PSA when an error such as negative transition probability, is detected) | Yes |  |  |
| Check the correlation between 2 PSA results (i.e., costs/QALYs under the SoC and costs/QALYs under the comparator) | Should be very low (very high) if different (same) random streams are used for different arms |  |  |
| OWSA=one-way sensitivity analysis; ICER = incremental cost-effectiveness ratio; PSA = probabilistic sensitivity analysis; WTP = willingness to pay; CE = cost-effectiveness; CEAC = cost-effectiveness acceptability curve; LY = life years; QALYs = Quality adjusted life years; OR = odds ratio; RR= relative risk; HR = hazard ratio | | | |