MANC-RISK-SCREEN Basic User Guide

Running the model

In the MANC-RISK-SCREEN model, the modelling process is split into two sections: generating cost and outcomes data for a sample of women; and analysing the data to produce results. Generating the data involves running the MANC\_RISK\_SCREEN\_main\_script. This script simulates a sample of women going through the screening strategies chosen by the user and saves a series of tables containing the costs and outcomes for women in each strategy. The Base Case Analysis script in the Analysis folder then synthesises the data and produce cost-effectiveness planes and a results table.

The strategies to simulate can be chosen by changing the elements in the “controls” list at the start of the main script. There are a number of choices the user can make which affect how the model runs and which strategies are evaluated. The elements of the controls list that can be changed are described below:

1. **Strategies**

This element contains a vector of numbers relating to which strategies are to be evaluated. The model will loop through each of the strategies included. The table below shows the strategies which are currently programmed into the model and their codes:

|  |  |
| --- | --- |
| **Code** | **Strategy** |
| 0 or any other number not listed | No screening (only clinical diagnosis) |
| 1 | A risk-stratified screening approach using the PROCAS (Tyrer-Cuzick + Volpara breast density) strategy: high risk women (>8% 10 year risk) screened annually, moderate risk (5-8% 10 year risk) screened biannually, all others screened 3-yearly |
| 2 | Risk-tertiles whereby women are divided evenly into 3 risk groups using 10 year risk provided by the Tyrer-Cuzick questionnaire and Volpara breast density: the highest risk third receive annual screening, the middle group receive bi-annual screening, the lowest group receive 3-yearly screening |
| 3 | Universal 3-yearly screening |
| 4 | Universal 2 yearly screening |
| 5 | Universal 5-yearly screening |
| 6 | Universal 10-yearly screening (at age 50 and 60) |
| 7 | Risk-stratified approach, reducing screening for women at low risk of cancer (<1.5% ten year risk) such that these women receive 5-yearly screening and all other receive 3-yearly screening |
| 8 | Risk-stratified approach, reducing screening for women at low risk of cancer (<1.5% ten year risk) such that these women receive 6-yearly screening and all other receive 3-yearly screening |
| 9 | A fully risk-stratified screening approach using the PROCAS (Tyrer-Cuzick + Volpara breast density) strategy: high risk women (>8% 10 year risk) screened annually, moderate risk (5-8% 10 year risk) screened biannually, low risk (<1.5% ten year risk) screened 5-yearly, all others screened 3-yearly |

1. **Gensample**

Gensample tells the model whether to create a new sample of women to simulate. The first time you run the model this will need to be set to TRUE to create an initial sample. Note that if multiple strategies are chosen then the sample is only generated once and then re-used for each strategy.

If set the FALSE the last generated sample will be used instead of generating a new sample. This would potentially be useful when comparing the results for different strategies with or without misclassification or chemoprevention (see sections below).

NOTE: when changing the desired\_case (sample size), mcruns, or intervals you will need to set gensample to TRUE to update the number of women in the sample. If errors occur in the model after changing the sample size despite setting gensample to TRUE then try manually deleting the contents of the Risksample and Risksamplewithmisclassification folders in the repository.

1. **MISCLASS**

When predicting a woman’s risk of breast cancer, there will be uncertainty in the prediction. This uncertainty may mean that a woman’s risk is predicted to be higher or lower than it should be and may therefore receive more or fewer mammograms than would be intended if risk were perfectly predicted.

The user can ask the model to account for this uncertainty by setting MISCLASS to TRUE in the controls. If alternatively MISCLASS is set to FALSE then it will be assumed that the risk prediction is a perfect prediction.

1. **PREVENTATIVE\_DRUG**

One of the potential benefits of risk stratified screening is that by predicting risk it will be possible to offer risk-reducing medication to women at higher risk. The user can evaluate the impact of adding preventative medication (sometimes called “chemoprevention”) to risk-based screening programmes by setting PREVENTATIVE\_DRUG to TRUE or can omit this by leaving it set to FALSE.

1. **PSA**

When turned to TRUE the model will generate data using monte carlo simulation with differing draws for core model parameters. Note that this is very computationally intensive as it will involve analysis of a sample of (desired\_case/0.12)\*mcruns women. Generally using the PSA function to conduct an actual PSA is not recommended due to this computational burden and instead a meta-model is fitted to the data to help in analysing uncertainty. This is explained in the intervals section below.

1. **Intervals**

The intervals argument is used to generate data on which to estimate a generalised additive model for conducting PSA. As the model is too complex for traditional PSA, monte carlo simulation is used but with a smaller number of draws than would be required for a full PSA. The parameters are then also drawn from much wider distributions (hence the name wide intervals).

With the data generated from an intervals run of the model a generalised additive model can be fit which predicts the QALYs and Costs for each strategy as a function of the uncertain parameters. The models are then applied to a sample of parameter values sampled using the true PSA distributions to create a sample of costs and QALYs for each strategy which can the be analysed.

1. **Desired\_cases**

This value sets the sample size in terms of the number of cancer cases included in the sample. We recommend this value is set to 300,000 to produce a sample large enough to produce stable differences between the risk-based strategies. However, this can be set to a lower value for more rapid analysis i.e. when testing out changes to model code.

1. **Chunks**

The sample of women created by gensample results in a very large data.frame usually using multiple GB of RAM. Holding the whole sample in memory would significantly slow down analysis or make it infeasible for very large samples. As such the larger sample is split into multiple smaller samples (chunks) by the model which are then all separately analysed.

Testing suggests that approximately 10 chunks are optimal for the standard analysis. Few chunks results in larger data.frames which slow down analysis. However, too many chunks also slows down the analysis as there is a fixed cost of loading and initialising the parallelisation for each chunk.

1. **Mcruns**

This sets the number of monte carlo simulations to run in the PSA or intervals analysis.

1. **Numcores**

The model runs using parallel processing. This parameter defines the number of computer cores to use to run the analysis. It is recommended that users use a maximum of one core less than the number of cores in the computer to run the analysis.