Incidence Sim Script

1. Define function to simulate age at which cancer found
2. Sample a time (age) at which cancer will occur based on prob of getting cancer at different ages
3. Determine when in the year the cancer found based on uniform distribution
4. In baseline assume cancer would be clinically detected
5. Determine if cancer was actually found by a screen. If cancer appeared at an age within the screening interval then draw a random number between 0 and 1 from a uniform distribution. If this number is less than the proportion who are screen detected then set the cancer as screen detected.
6. Determine size of cancer at detection
7. For clinically detected draw from a normal distribution based on input parameters for size of screen detected cancers
8. If size is less than 4 then set at 4 to prevent unrealistically small tumours
9. If size is greater than or equal to 9 then set at 8.99 to stop unrealistically large tumours
10. This number is actually the number of doublings so recode variable as start size \*2 to the power of the number of doublings
11. If screen detected the repeat with different input parameters
12. Check that screen detected size is not greater than the clinically detected size and correct if true
13. Generate all-cause mortality time by sampling a future age based on the probability of dying at that age
14. Put variables into a results vector and report
15. Compile function

NPI by size function

1. Define function with arguments for tumour size and whether it was screening detected
2. Define empty variable for NPI category
3. Set NPI category to 4 (DCIS) if cancer screening detected and a random number drawn between 0 and 1 is less than the proportion of cancers which are DCIS
4. If cancer size is less than or equal to 25 then m size is 25
5. If cancer size >25 then m size is a ceiling function of cancer size
6. Set m size to a maximum of 85
7. Look up the probability of cancer being metastatic given the m size and set NPI category to 5 if a random number drawn between 0 and 1 is less than the probability
8. If NPI category still not assigned then need to sample from a matrix of probabilities of category 1,2 or 3 based on cancer size.
   1. Determine which of 6 intervals the cancer size fits in to
   2. Sample an NPI category (1,2,3) with probability based on the cancer size interval
9. Store NPI category
10. Return NPI category
11. Compile function

Screening result function

1. Define function with arguments of cancer size, volpara density group, whether MRI or US screening being used)
2. Define sensitivity of screening conditional on size of cancer at screening appointment. Uses logistic function to predict sensitivity but sets maximum sensitivity at 95%.
3. Define odds ratio for sensitivity if woman has dense breasts
4. Define sensitivity of screening for women with dense breasts as sensitivity \*OR for breast density
5. Redefine sensitivity as the value from 4
6. Determine if cancer detected at screen and at what stage
   1. Set random number from uniform distribution
   2. If random number less than sensitivity then cancer is detected at screen by mammography.
   3. If cancer not detected by mammography but MRI used then calculate supplementary odds of finding cancer using MRI + mammography vs mammography alone
   4. If random number less then new sensitivity then cancer is screen detected and detected by MRI
   5. Repeat c and d for US screening
7. Define result which is a vector containing whether cancer is screen detected and at what stage it was detected
8. Compile function

Survival time function

1. Define function which relies on arguments of NPI category, age, age at which person would die without cancer, age at which cancer diagnosed.
2. Define gamma NPI which is a vector of survival curves for cancer with different NPI categories
3. Define metastatic survival which is a vector of survival curves for metastatic cancer starting at different ages
4. If NPI category is less than 4 then draw a survival time appropriate to the NPI category from the gamma NPI vector
5. If cancer diagnosed above age 65 then apply an odds ratio to survival time reflecting additional mortality
6. If survival time is greater than 10 years, assume patients die at their all-cause mortality age. Need to set survival time as all-cause mortality death age minus current age
7. If NPI is five then draw a survival time from vector of survival times for metastatic disease based on age category
8. If the person will survive to be over 100, their survival time is recalculated as the time horizon minus their cancer diagnosis age
9. If NPI category is 4 then survival time is their all-cause mortality death age minus their cancer diagnoses age i.e. there is no effect on mortality
10. If current age plus survival time is greater than time horizon then set survival time to time horizon minus cancer diagnosis age
11. Store result which is cancer diagnosis age plus survival time
12. Return result
13. Compile functions