**Creating input data:**

1. File name: creating\_weights.R

Calculates weights for each subcohort (race by smoking status by lifetime quartile of sexual partners, by age) for men aged 15-79, using NHANES 2009-2016 data.

1. File name: nhis2009\_16\_for\_weights\_age\_80up.R

Calculates weights for each subcohort for men aged 80-84, using NHIS 2009-2016 data.

1. File name: create\_splines.R

Calculates the spline terms when using age with a 5 knot spline

1. File name: Logistic\_regression\_new\_partner.sas

Logistic regression model to predict whether an individual has an oral sex partner in the past 12 months

1. File name: negative\_binomial\_number\_of\_partners.R

Negative binomial model of the number of recent female oral sex partners, using data from men in NHANES 2009-16.

1. File name: partner\_age\_distribution.R

Calculates the age distribution of men’s sexual partners, using data from NHANES 2005-6 to get 3 broad categories, then data from NATSAL 3 to get 5-year groups (relative to the age of the man)

1. File name: cervical\_HPV\_prevalence\_median\_quartile\_modelled.R

Models cervical HPV16 prevalence in NHANES 2005-06, using a 3-knot spline for age

1. File name: gridsearch\_code\_upto74.sas

Running a model to get the starting points of a grid search for the probability of developing cancer within the next year from an HPV16 infection, by duration of infection.

1. File name: penetrance.quadratic.gridsearch.R

Create a file containing all the penetrance combinations (probability of developing cancer given an x-year HPV infection)

1. File name: nhis2010\_14\_men\_1yrdeathrates.R

1-year mortality model using data from NHIS 2010-2014

**Running the simulations:**

1. File name: sims.R

Running natural history simulations; uses src/functions.cpp and R/functions.R

1. File name: identifying\_top\_datasets.R

Identify best-calibrated natural history simulations, calibrating to oral HPV prevalence

1. File name: sim\_output.R

Get simulation output for the best-calibrated natural history simulations

1. File name: create\_files\_from\_output\_for\_penetrance.R

Create 3 input files for next step from the output from the natural history simulations: the duration distribution of prevalent HPV16 infections, number of people with HPV16 per 100,000, and total number of people with an HPV16 infection.

1. File name: penetrace.R

Run all combinations of penetrance for each of the best calibrated natural history datasets

1. File name: identifying\_top\_penetrance\_datasets\_gridsearch.R

Identify best-calibrated cancer incidence simulations, calibrating to cancer incidence, and outputting the penetrance values that provided the best calibration for each natural history dataset

1. File name: sims\_cancer.R

Run simulations with vaccination and herd immunity

**Analyzing the output**

1. File name: analysing\_cancer\_output.R

Creating a summary of the total number of cancers under each vaccination scenario, grouped by birth cohort

1. File name: cancer\_reduction\_SQ26\_calc\_for\_each\_natural\_history.R

Calculating the reduction in cervical cancer incidence by birth cohort, relative to under status quo vaccination up to age 26