**Input file formats:**

Also used in natural history simulations (<https://doi.org/10.1093/jnci/djad009>):

1. longitudinal\_weights.csv: Each row represents an age. The columns show the proportion of men of each age with each combination of race, smoking status, and quartile of lifetime sexual partners.
2. pop\_size.csv: Each row gives the number of men alive in 2021, from age 84 (row 1) to age 15 (row 70).
3. mortality32.csv: Each row represents an age, with age 15 in row 1 and age 84 in in row 70. Each column gives the proportion of men expected to die in 1-year from all causes, for each combination of race, smoking status, and quartile of lifetime sexual partners.

Results from natural history simulations (<https://doi.org/10.1093/jnci/djad009>):

1. calibrated\_seeds.csv: A list of the seeds which provided the best calibration to oral HPV16 prevalence and HPV16-positive oropharyngeal cancer incidence in the original natural history model (Landy et al, JNCI 2023, <https://doi.org/10.1093/jnci/djad009>)
2. screening\_data51\_", seed, ".csv": The oral HPV16 natural history each year from ages 15-84 for each simulated individual under status quo vaccination, for the given natural history seed. 0 means no infection, integer values give the duration of an infection, 0.5 represents an acquired infection which clears before the next year.

Screening-associated input files:

1. AJCC7.by.age.csv: Each row represents an age from 15 – 84. Column A gives the proportion of cancers diagnosed in AJCC v7 Stage I, column B gives the cumulative proportion of cancers diagnosed in AJCC v7 Stage I or II, and column C gives the cumulative proportion of cancers diagnosed in AJCC v7 Stage I, II or III.
2. stage.progression.csv: Each row represents a different progression speed scenario. Column B gives the coefficient for an exponential distribution for the time spent in Stage I; column D gives the coefficient for an exponential distribution for the time spent in Stage II, Column F gives the coefficient for an exponential distribution for the time spent in Stage III, Column H gives the coefficient for an exponential distribution for the time spent in Stage IV. Columns A, C, E and G are not used.
3. AJCC7 TNM proportions\_numeric.csv: Cancer characteristics.

Column A: AJCC v7 Stage (1=Stage I, 2=Stage II, 3=Stage III, 4=Stage IV)

Column B: AJCC v7 T (1=1, 2=2, 3=3, 4=4, 5=all (for M=M1))

Column C: AJCC v7 N (1=1, 2=2a, 3=2b, 4=2c, 5=3)

Column D: AJCC v7 M (0=M0, 1=M1)

Column E: Cumulative proportion of AJCC v7 cancers of the given stage with these characteristics

Column F: proportion of tumors with these characteristics that have tumor size >5mm

Column G: proportion of tumors with these characteristics that have a node >6cm

Column H: proportion of tumors with these characteristics that have a node 3-6cm

Column I: proportion of tumors with these characteristics that have a node <3cm

Column J: AJCC v8 Stage (1=Stage I, 2=Stage II, 3=Stage III, 4=Stage IV)

Column K: AJCC v8 T (1=1, 2=2, 3=3, 4=4, 5=all (for M=M1))

Column L: AJCC v8 N (1=1, 2=2a, 3=2b, 4=2c, 5=3)

Column M: AJCC v8 M (0=M0, 1=M1)

Column N: not used

Column O: proportion of tumors with these characteristics that have tumor size <10mm

Column P: proportion of tumors with these characteristics that have tumor size 10-19mm

1. screening\_scenarios.csv: Each row represents a screening scenario. Column A is the start age for screening, Column B says how many years between screens (or 0 when a set number of annual screens are performed, see Column C), column C says how many annual screens should be performed if screening is not continued until age 79/exit from screening/cancer diagnosis.
2. cumulative.cause.specific.survival.by.smoking.csv: Column A shows smoking status (0 for never/former, 1 for current), Column B shows Stage at diagnosis (0 = Stage I, 1 = Stage II, 2 = Stage III, 3 = Stage IV), Column C shows 1-year survival, Column D shows 2-year survival, Column E shows 3-year survival, Column F shows 4-year survival, Column G shows 5-year survival.
3. mortality\_denominator\_51\_", seed, "\_1.csv: The number of men alive in each calendar year at each age, for the specified natural history seed, to calculate cancer incidence rates.