CONFIDENTIAL

Appendix - Clinical impact and cost-effectiveness of primary human papilloma virus testing

A1 Parameterisation of Sexual Behaviour

Sexual debut

Entry into the HPV-susceptible population is determined by age of sexual debut. Responses from the National Survey of Sexual attitudes and Lifestyles 2010 (NATSAL-3) are used to construct the cumulative probability distributions of age at sexual debut.

The data allow us to directly determine the cumulative fraction women that are sexually active from age 16 to 75 years. For individuals aged 16 years and under, we use the distribution of reported age at first sex, S_d, (for those that report sexual activity before the age of 16 years), and scale this to the known fraction of individuals that are active by age 16 years, to determine the probability of sexual debut from age 10 onwards ($P[S_d=d] = P[S_d=d \,|\, S_d<=16]P[S_d <=16]$).

A hill function is used to provide a smooth monotone description of the empirical cumulative distribution of sexual debut age before and after the age of 16 years (shown in Figure A1.

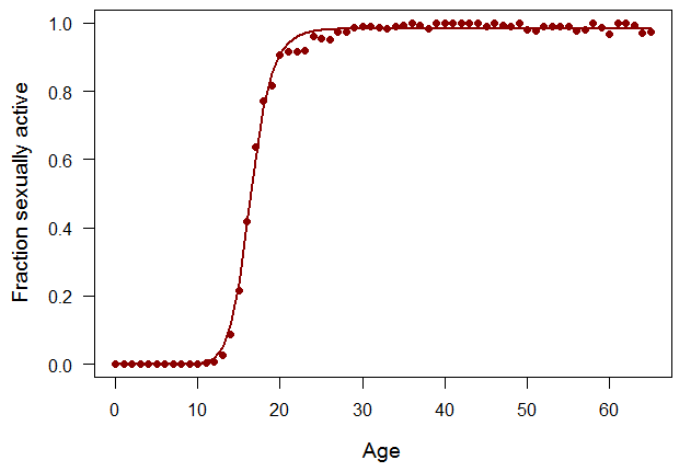


Figure A1: Cumulative probability of sexual debut as a function of age; age reported in NATSAL-3 (points) and smooth monotone function that best describes the data (line).

Partnership acquisition

The number of new partners acquired in the last year, as reported in NATSAL-3, reveals a trend towards decreasing partner acquisition with age with significant variation between individuals at the population level.

We divide the population into four sexual behavioural categories according to behaviour in the recent past; S_i, for i={1,2,3,4}, where increasing i reflects increasing sexual activity, and represent women in the 0-40th, 41-80th, 81-97th and 98-100th percentile of a given age-band, with respective to number of partners in the past five years.

We use a poisson distribution to describe the number of partners, S_N, acquired over the last 12 months.

$$P[S_N=n] = \exp(-\mu)\,\mu^n/n!$$

We divide the population into five year age bands (16-20, 21-25, 26-30, 31-35, 36-40, 41-45, 46-50, 51-55, 56 an over) and use a nelder-mead optimisation to identify the rate of partner acquisition, μ , that best describes each sub-population within ageband *a* (S_{1_a}, S_{2_a}, S_{3_a}, S_{4_a}).

An individual within the model remains associated with a given behavioural category throughout their lifetime, however, the rate of partner acquisition associated with each behavioural category decreases with age.

Partnership duration

A survival model is used, in combination with the NATSAL-3 dataset, to parameterise the cumulative probability of relationship 'survival' as a function of time and age of women at the start of relationship. In the natsal survey, individuals reported on the three most recent partners with sexual activity within the past 5 years. To counter the bias towards observing longer-relationships in a fixed-window sampling scenario, the survival of relationships is calculated using a modified Kaplan-Meier estimator that explicitly accounts for truncation (as described by Burington and colleagues[1])

$$S(x) = \prod_{y_i<x} 1 - d(y_i)/R(y_i)$$

where $d(y_i)$ represents the number of uncensored events (relationships) of length exactly y_i ;and the denominator R counts the total number of relationship events lasting more than or equal to y_i months, but excluding events that began more than y_i months before the start of the sampling window.

$$R(y_i) = \sum_j I(t_j < y_i) - \sum_j I(y_j < y_i)$$

where t_j is the time between the start of the relationship and the start of the sampling window, measured as 5 years before the date of interview according to the natsal questionnaire design, (t_j is 0 for partnerships that began after the start of the sampling period), and $I(x)$ is the indicator function, value 1 if x is true, 0 otherwise.

A partnership is defined as complete when there has been no sexual activity for 3 months. Data from the first, second and third most recent heterosexual partner is combined.

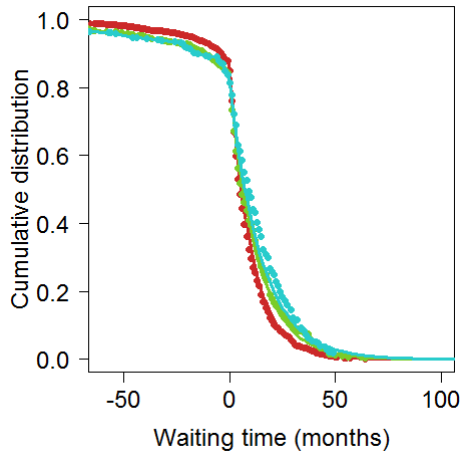
Missing partnerships: The sampling of detailed partnership information from the three most recent partnerships in the past 5 years can lead to a bias towards longer relationships and those with a large gap between relationships as they are most likely to be 'most recent' at time of interview. We compare the total number of partners in the past 5 years to the number of partners for whom we have detailed information for individual, to determine the number of missing partnerships in our data sample. It is assumed that missing partnerships are complete and therefore will be most similar to completed partnerships that have been reported in detail by the same individual.

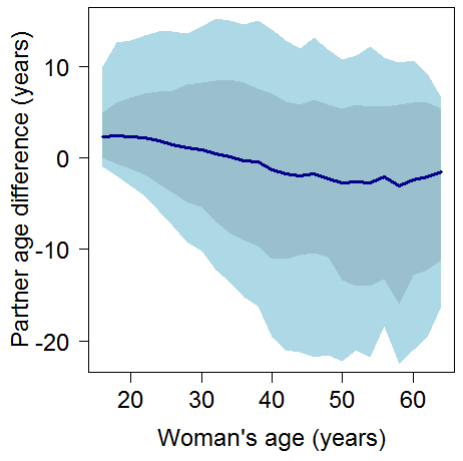
The detailed information on complete partnerships for individual *i* can be weighted by W, $W = 1 + (H_{5,i} - T_i)/C_i)$, where $H_{5,i}$ is the number of partnerships reported in the last five years; T_i is the number of partners for whom detailed partnership data is available; C_i is the number of complete partnerships for whom data is available for subject i.

Relationship survival curves exhibit a biphasic decay; with a large number of short term relationships and a smaller number of very long partnerships. The data reveals that the fraction of relationships falling into the short-term category increases as a function of age at start of relationship; the five year survival for a relationship is 19%, 7% and 3,5% for women aged 16 to 20 years, 31 to 35 years, and 51 to 55 years, respectively, at the start of relationship.

Age mixing

An age-mixing matrix is generated by directly sampling from the age of most recent partners reported by female participants in NATSAL-3. We stratify the data according to the age of female respondent at the start of relationship to reveal an increasing variance in partner age for older women. This approach is preferred to a more traditional approach to partner-matching that assumes a constant age difference distribution or an approximation of +/- 3 years, as it better captures the complexities of HPV transmission; in particular, the role of novel HPV infections in older women versus long term persistent infection. The shift in age difference is illustrated in Figure A2.





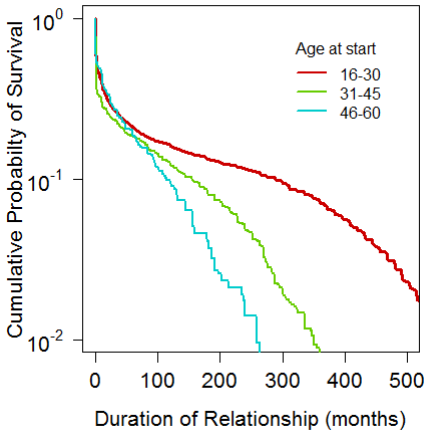


Figure A2: Sexual behaviour parameters are identified using data collected in natsal-3. (A) Simulated cumulative probability of a waiting time to next partner after the end of previous relationship; (B) cumulative probability relationship survival as a function of time since start of relatiosnhip; (C) age preference of male partners as a function of female age at start of relationship.

Frequency of sex acts

The model simulates HPV transmission by introducing a probability of transmission per sexual contact. We use data from NATSAL-3 to quantify the number of sex acts per month for individuals in an active relationship (defined as those participants who reported a sexual encounter in the last 3 months with the most recent partner). As above, a weighting is added to response data that scales with total 5 year partner count to remove the bias towards reporting of characteristics from long term relationships (Figure A3).

We also derive the fraction of new relationships that result in a single sexual contact only. We distinguish between recent ongoing relationships that may yet lead to further contacts and those that are complete by assuming that relationship is complete if there have been no contact in the past three months.

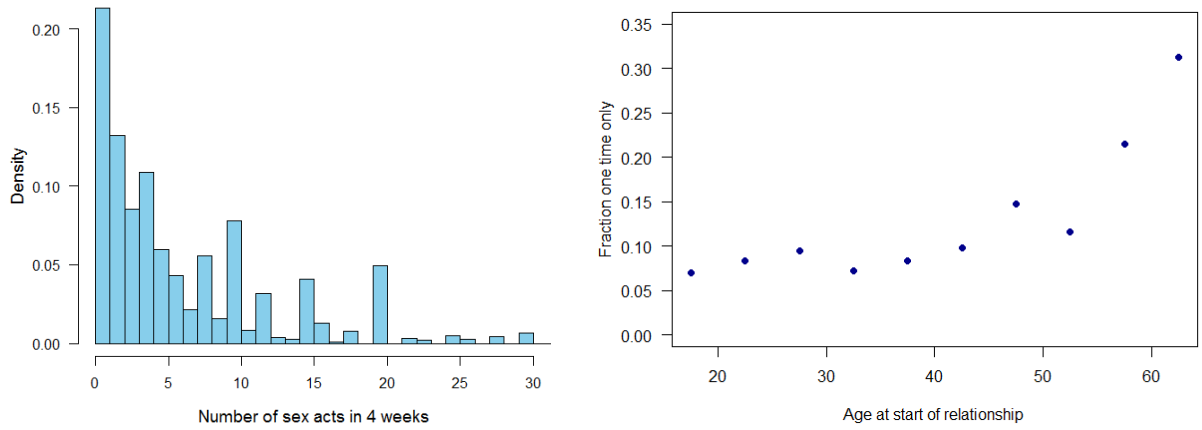


Figure A3: Left panel: Distribution of reported sex acts in a 4 week period for individuals in an active relationship. Right panel: Fraction of new partnerships that result in a single contact only.

A2 Parameterisation of HPV infection

The sexual behaviour characteristics described above are used to generate the age of sexual debut and formation and dissolution of subsequent sexual partnerships for a population of 100,000 women from birth to age sixty five years.

The HPV status of a new male partner is randomly generated using the distribution of age of new partners, reported in NATSAL-3, according to the age of the woman at start of partnership, and the prevalence of HPV among men of the preferred partner age.

HPV infection is modelled by introducing a per-sex-act probability of transmission of HPV. The probability of infection by each strain is assumed to be independent. In this work we consider HPV-16, 18, 31, 33, 45, 51 and 52. Rates of clearance and transmission of HPV were parameterised using HPV prevalence data in women and sero-prevalence data in males.

HPV-strain specific prevalence was determined using surveillance data collected by Public Health England from residual samples taken from the NHSCSP pre-immunisation

for women aged 25-65 years[2]. For younger women between the ages of 16 and 24 years, HPV prevalence was measured in residual samples taken from the national chlamydia screening programme (NCSP), pre-HPV-immunisation[3]. Data from these younger women is important for characterising the peak of HPV infection, however, the selective nature of women attending the NCSP means that the sample reflects a higher sexual risk group than the general population; NCSP data is accompanied by data with number of partners reported in the past 12 months which is higher than that predicted by NATSAL-3 for women aged16-24 years. We introduce a weighting to resample the NCSP data such that the number of partners reported in the past 12 months matches that observed in natsal-3 for women aged 16-24 years, and introduce a sub-population of sexually-inactive women in the same age range (as predicted in NATSAL-3 but not present in the NCSP dataset) who are expected to be HPV-naïve. HPV prevalence is recalculated in this re-weighted population and this new comparable prevalence is merged with the NHSCSP predicted prevalence.

HPV prevalence data was not available for a sufficiently large population in England. Instead, national surveillance data describing male sero-prevalence of HPV-16 and HPV-18 in England, collected by PHE[4], was used to estimate prevalence of HPV among males. A study of sero-prevalence in the Netherlands revealed that sero-prevalence levels were similar in HPV-,33, 45 and 52, but approximately two-fold higher in HPV-31[5] ; a similar result was found in the German population[6]. In the parameterisation that follows, sero-prevalence of hpv 33, 45 and 52 among males was constrained by observed sero-prevalence of hpv-18, in accord with levels of hpv prevalence of these strains observed in women in England. Sero-prevelance of HPV strains 31 and 51 was estimates by scaling the observed hpv-18 sero-prevalence according to the ratio of hpv prevalence of hpv-31: hpv-18 and hpv-51: hpv-18 observed in women.

HPV infection in males

A static model of transmission is applied in which male prevalence is assumed to be constant throughout the duration of these simulation; we argue that the introduction of primary HPV DNA testing in cervical screening will not have an effect on male prevalence of HPV.

A simple three compartment differential equation model is used to analyse the sero-prevalence data and extract HPV prevalence for each HPV type. We consider individuals that are (i) infected but sero-negative (I); (ii) sero-positive, that is they have detectable HPV antibodies (S), and (iii) infected and HPV-DNA positive (H).

$$\dot{I} = f(t) - (c + k)I(t)$$

$$\dot{S} = k\, I(t) - wS(t)$$

$$\dot{H} = f(t) - cH(t)$$

Where, *f(t)* is the number of new infections at time *t*; *c* is the rate of clearance of male infection; *k* is the rate of sero-conversion; and *w* is the rate of HPV antibody waning.

The size of the infected population, H, can be estimated using numerical methods to solve the following equation:

$$\dot{H} = \frac{\ddot{S}}{k} + (w + c + k)\frac{\dot{S}}{k} + w(c + k)\frac{S}{k} - cH(t)$$

where, the observed sero-prevalence, S[t], can be described by a polynomial, and it is assumed that the half life of antibodies is at least 20 years, that is the rate of waning (w) is constrained to be less than 0.05 (/year).

The rates of sero-conversion and clearance for each male HPV strain are identified, together with the rates of female clearance and transmission, using the observed HPV sero-prevalence in males and prevalence in females. Described in detail below.

Parameterisation

The disease transmission was parameterised independently for each HPV-subtype. In this parameterisation we assume that HPV prevalence is not sensitive to screening strategy. The reasoning is that (i) the number of women treated for cervical lesions is small relative to the number of women that are infected with HR-HPV, ~10% of population at large; and (ii) not all treatment of lesions leads to clearance of HPV-infection. As a result, we can identify the rate of disease transmission and clearance using a simplified individual-based model without screening, in a computationally tractable parameterisation.

A Markov Chain Monte Carlo simulation, using an adaptive Metropolis algorithm, was implemented using the FME package in R to simultaneously identify the (i) HPV clearance parameters in females and males (c_1, c_2, c_m); (ii) per contact probability of transmission from males to females; and (iii) rate of sero-coversion in males. Each chain was run for a length of 20,000 and 100 parallel chains were generated using randomly generated starting values, for each strain of HPV.

A thinning interval of 50 was used to remove auto-correlation within each chain. Convergence was identified using the Geweke test statistic, a test of equality of the means of the first 10% and last 50% of the markov chain. A Gelman convergence diagnostic was then used to confirm convergence of the MCMC output in the parallel chains; a comparison of the empirical variance of each parameter within each chain should be comparable to the variance from all chains combined. The final parameter distribution reflects the joint distribution of the parallel chains.

Clearance of HPV infection

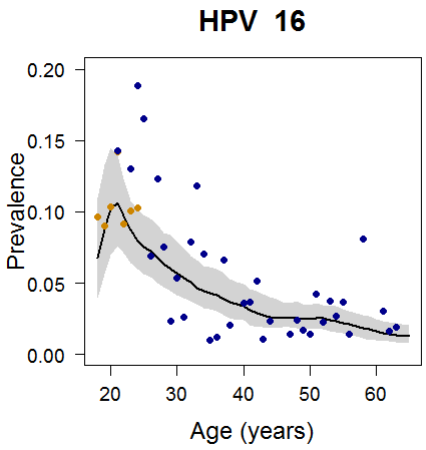
We model the waiting time to clearance of infection using a weibull distribution C~W[c_1, c_2] , where a c_1 value of less than 1 gives a decreasing rate of clearance with time and determines the scale of the distribution. We find that all HPV types are well described by a decreasing rate of clearance with increasing time since infection, that is the shape parameter lies below 1. The scale parameters lead to the differences in the time of clearance, and the analysis suggests that HPV types 18 and 51 are cleared most rapidly, with just 16% and 25% of women expected to remain infected for 12 months, respectively, compared to 54% and 69% of women infected with HPV types 16 and 31

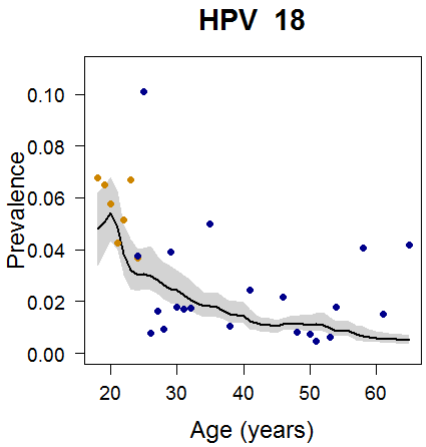
respectively. The most persistent long-term infections are found to be associated with hpv –strains 31, 33 and 45.

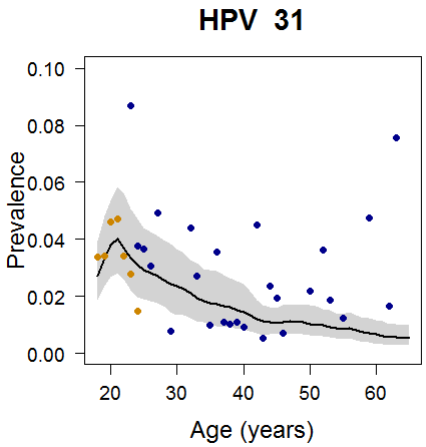
The rate of male clearance is not well constrained, but the 95% confidence interval of rates suggest a half-life of male infection of at most 7 months across all HPV types. Under the assumption that HPV antibodies waning results in a half-life of at least 20 years, the model predicts low rates of male sero-conversion of 0.043 (/year) for HPV-16 and 0.011 (/year) for HPV-18.

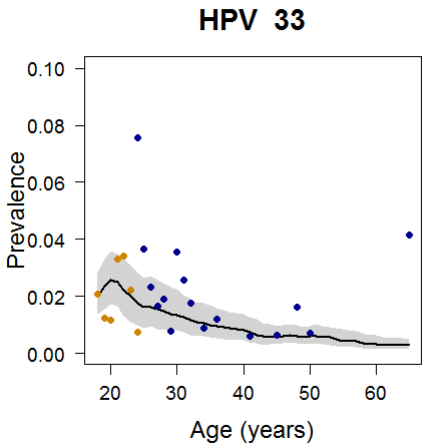
HPV transmission probability (per-sex act)

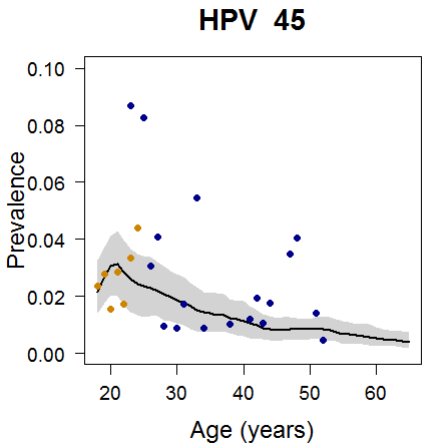
We find that the transmission probability per contact is not well defined. One explanation for this is that, according to the sexual behaviour data, the majority of partnerships result in multiple contacts; where the probability of contracting HPV from an infected partner, 1-(1- *Transmission.Probability*)^*'Sex.Acts'*, becomes decreasingly sensitive to the *Transmission.Probability* as the number of Sex.Acts increases. We accept the broad range of values suggested for transmission probabilities as they are able to reproduce the observed HPV prevalence within the context of known sexual behaviour.

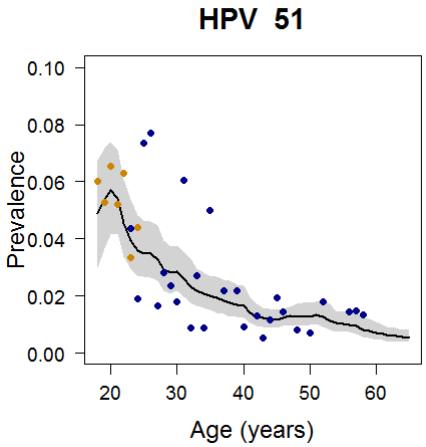












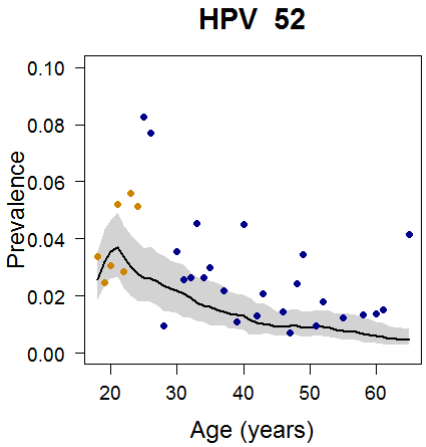
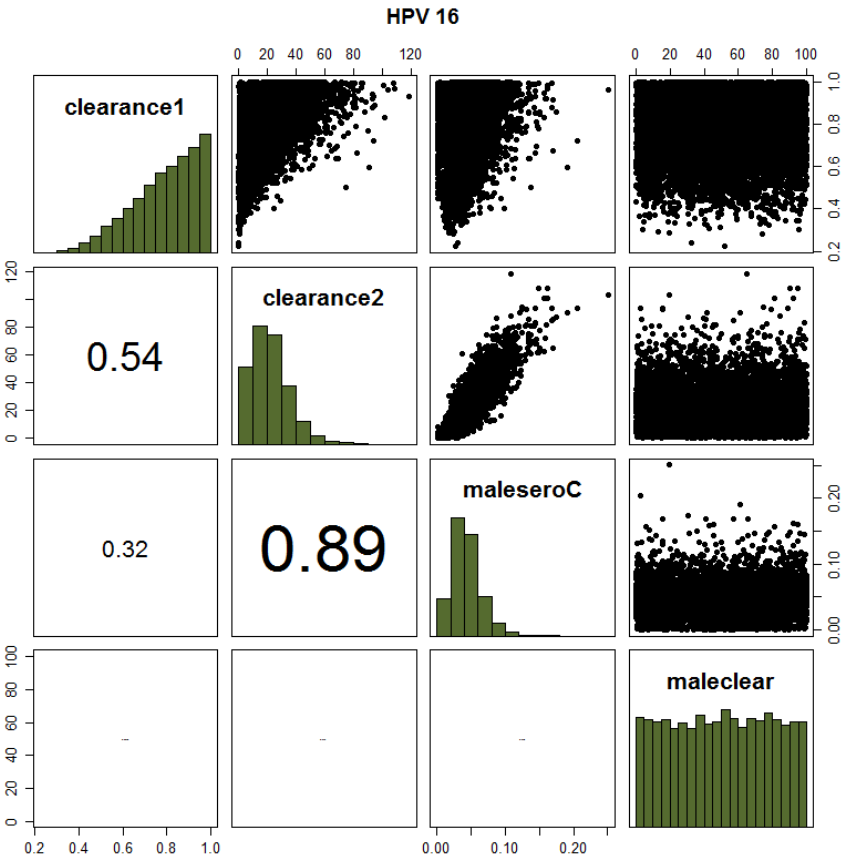


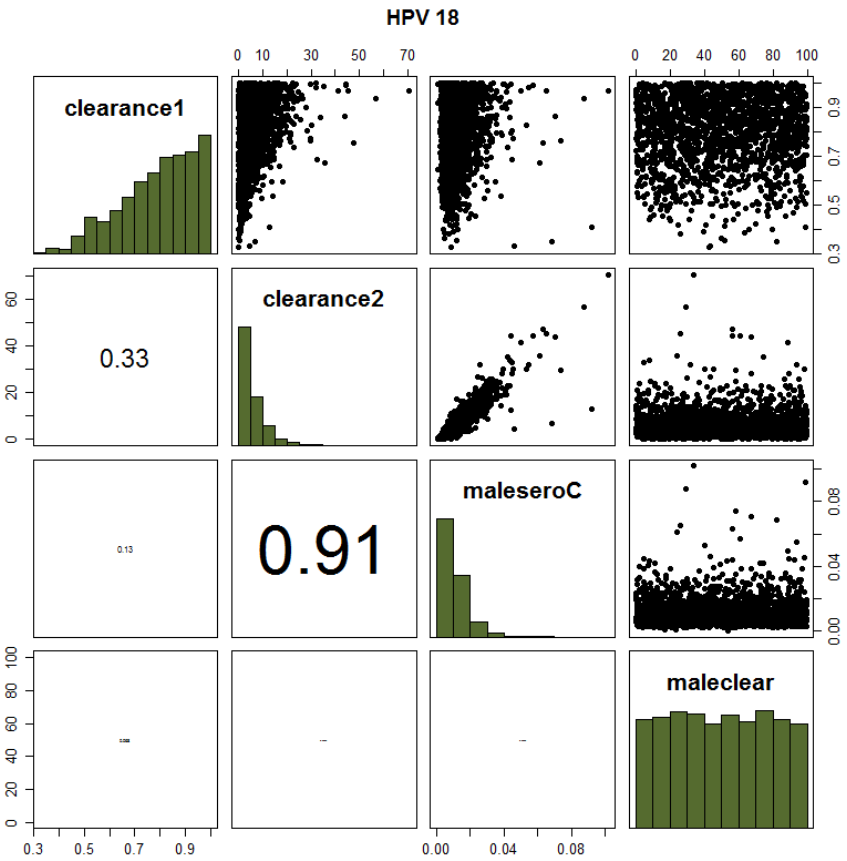
Figure A4: Observed type-specific HPV prevalence measured in residual samples from the NCSP (age 16-24 years – orange points) and NHSCSP (age 24-65 years – blue points) and best-fitting model predictions – mean (solid black line), upper and lower 95% interval (grey shaded region).

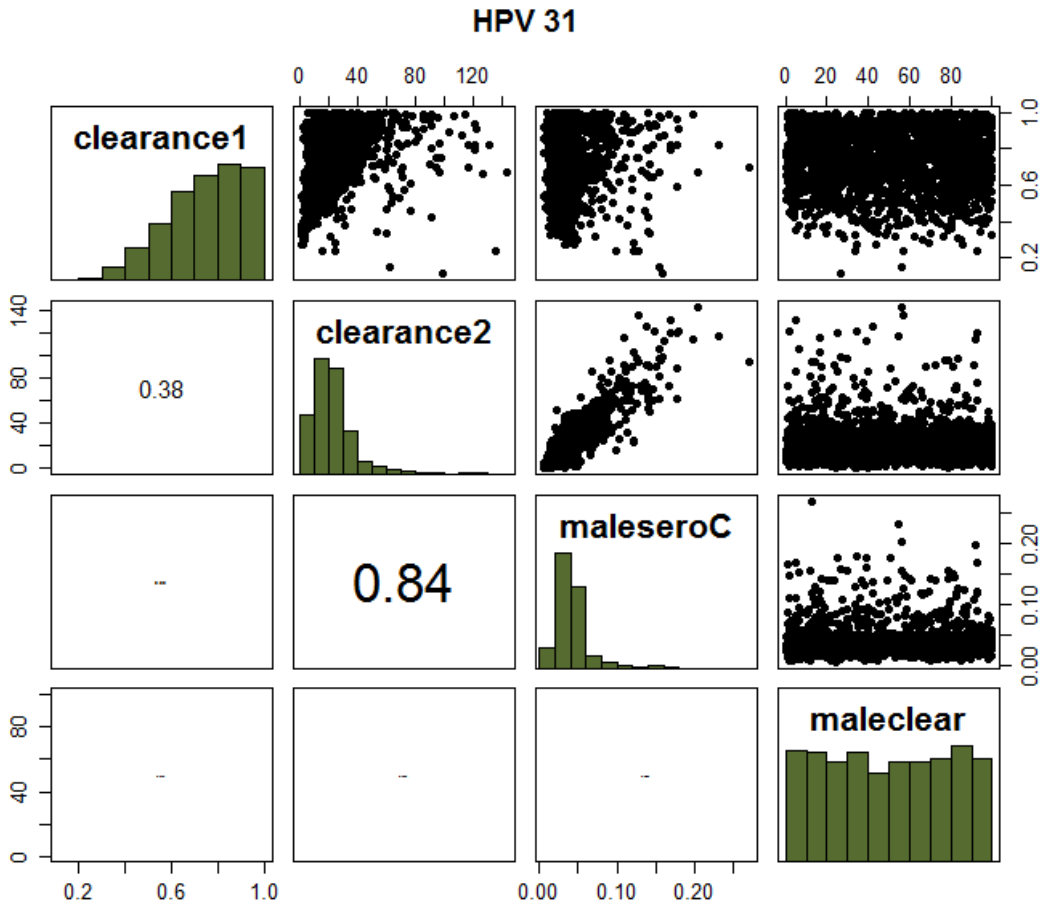
	Per-contact probability of transmission	c1	c2	Fraction of women remain infected at 12 months	Fraction of women remain infected at 24 months	Male seroconversion (/year)	Male rate of clearance (/year)
HPV 16	0.536	0.802	22.274	0.495	0.321	0.042	51.001
	(0.045 ,0.98)	(0.459 ,0.993)	(1.183 ,53.791)	(0.007 ,0.778)	(0 ,0.614)	(0.002 ,0.09)	(1.975 ,97.415)
HPV 18	0.580	0.790	5.259	0.137	0.051	0.010	49.922
	(0.051 ,0.985)	(0.475 ,0.991)	(0.223 ,17.243)	(0 ,0.484)	(0 ,0.271)	(0.003 ,0.03)	(2.61 ,96.616)
HPV 31	0.452	0.735	18.404	0.458	0.287	0.035	50.609
	(0.023 ,0.95)	(0.395 ,0.983)	(2.756 ,37.403)	(0.15 ,0.698)	(0.058 ,0.503)	(0.014 ,0.05)	(2.538 ,97.065)
HPV 33	0.422	0.716	10.116	0.304	0.155	0.037	51.348
	(0.012 ,0.966)	(0.363 ,0.986)	(1.18 ,22.975)	(0.046 ,0.568)	(0.006 ,0.355)	(0.014 ,0.049)	(2.776 ,96.175)
HPV 39	0.499	0.786	14.462	0.374	0.209	0.021	50.820
	(0.027 ,0.978)	(0.47 ,0.992)	(1.354 ,38.19)	(0.023 ,0.702)	(0.002 ,0.513)	(0.005 ,0.045)	(3.367 ,97.786)
HPV 45	0.418	0.728	14.340	0.388	0.226	0.036	49.239
	(0.019 ,0.974)	(0.404 ,0.987)	(1.625 ,31.474)	(0.049 ,0.645)	(0.008 ,0.452)	(0.01 ,0.05)	(3.561 ,97.703)
HPV 51	0.518	0.789	7.630	0.205	0.093	0.012	50.723
	(0.042 ,0.975)	(0.47 ,0.99)	(0.272 ,26.266)	(0 ,0.606)	(0 ,0.397)	(0.002 ,0.032)	(2.848 ,98.391)
HPV 52	0.454	0.764	17.917	0.460	0.279	0.037	48.677
	(0.024 ,0.97)	(0.428 ,0.99)	(3.463 ,34.4)	(0.163 ,0.683)	(0.057 ,0.485)	(0.016 ,0.049)	(2.464 ,97.512)
HPV 58	0.486	0.776	13.805	0.355	0.200	0.020	49.638
	(0.037 ,0.965)	(0.43 ,0.99)	(1.047 ,37.687)	(0.007 ,0.707)	(0 ,0.509)	(0.004 ,0.044)	(3.281 ,96.428)
HPV 59	0.529	0.814	6.993	0.195	0.074	0.011	50.930
	(0.041 ,0.981)	(0.48 ,0.993)	(0.43 ,19.868)	(0 ,0.526)	(0 ,0.305)	(0.003 ,0.025)	(2.452 ,97.32)
HPV 66	0.510	0.789	8.938	0.243	0.113	0.012	48.510
	(0.04 ,0.985)	(0.445 ,0.994)	(0.471 ,29.398)	(0 ,0.624)	(0 ,0.422)	(0.003 ,0.036)	(1.951 ,97.546)

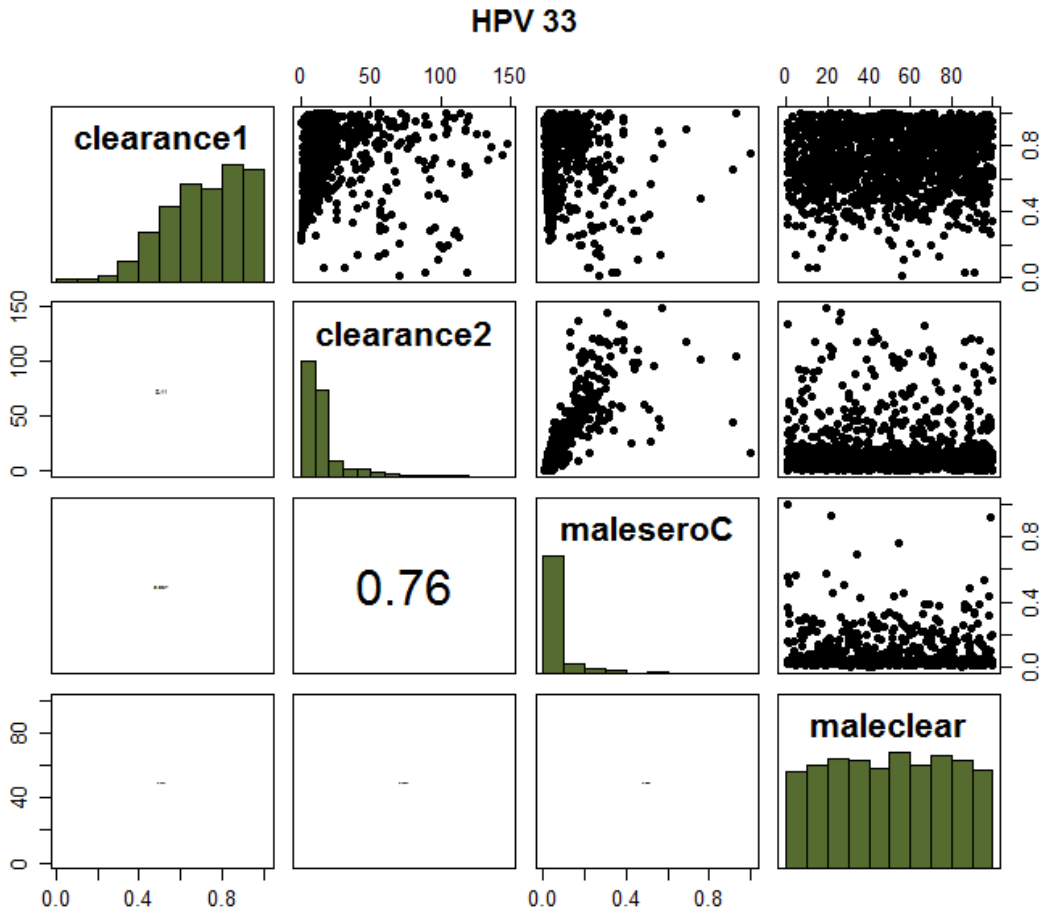
Table A1: Best fitting clearance and transmission parameters for HPV-16, 18, 31, 33, 39, 45, 51, 52, 58, 59 and 66

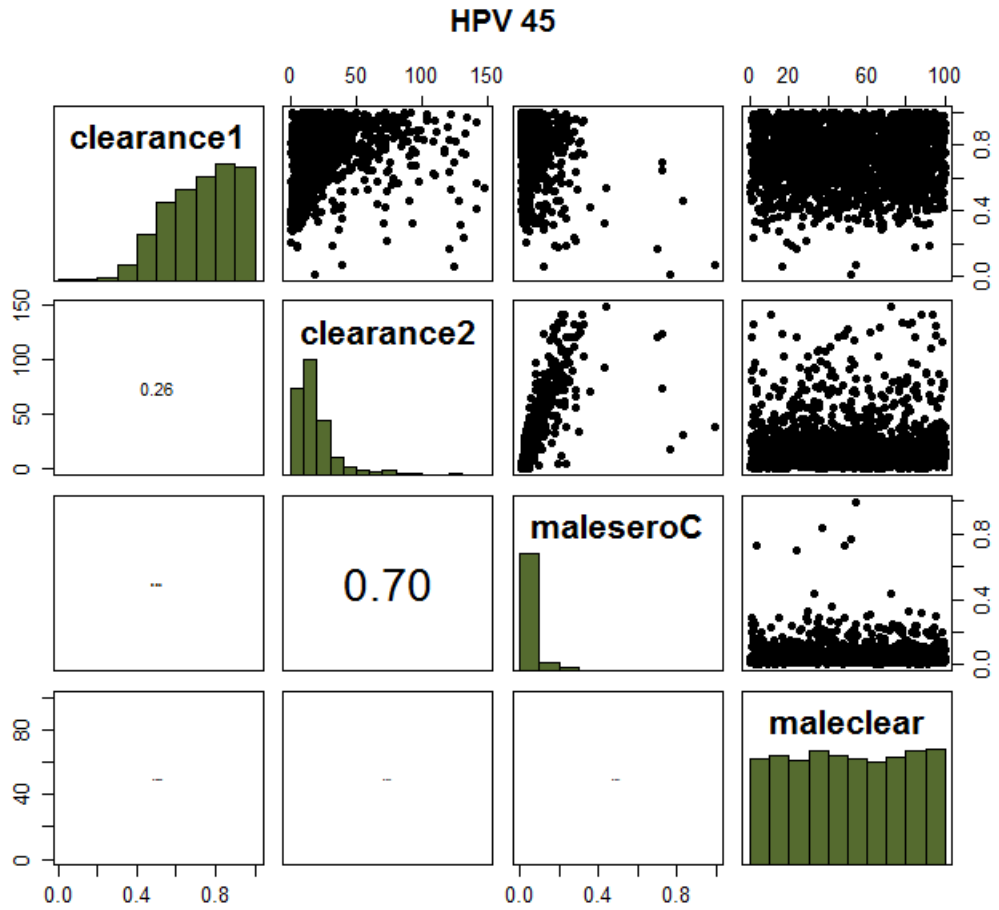
Figure A5: Posterior distributions of best fitting clearance and transmission parameters for HPV-16, 18, 31, 33, 45, 51 and 52

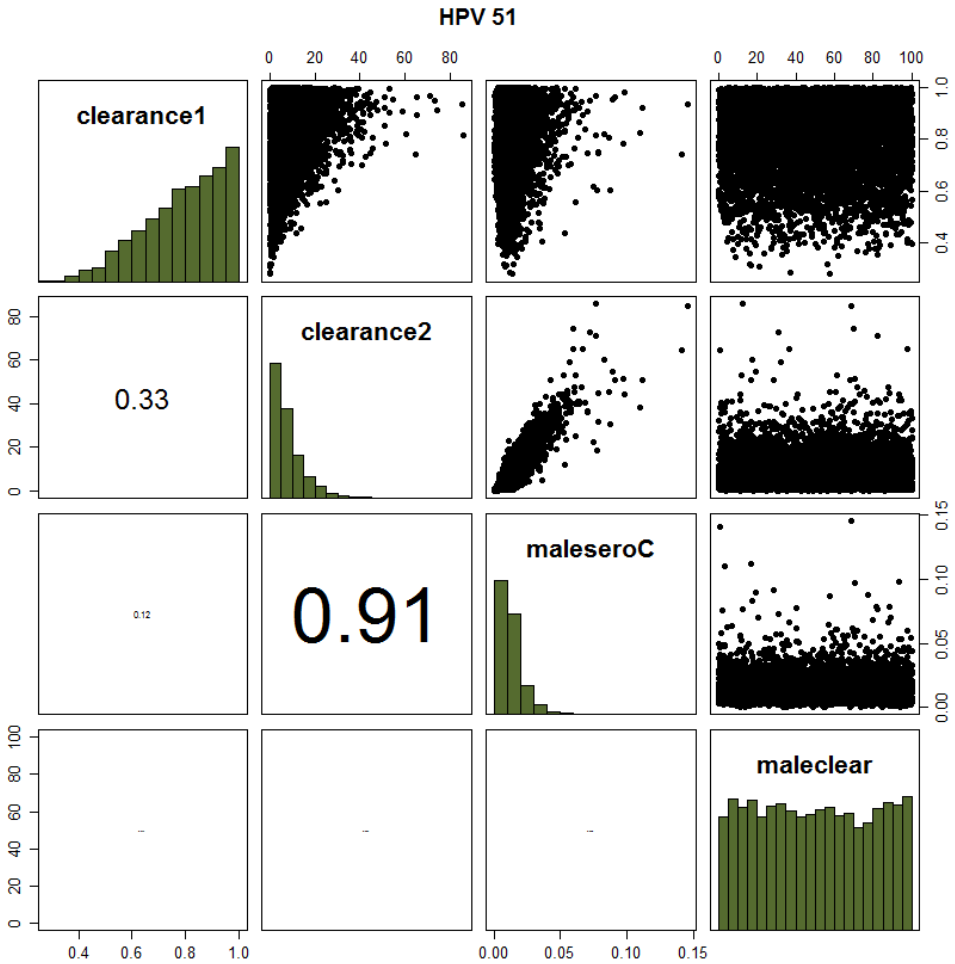


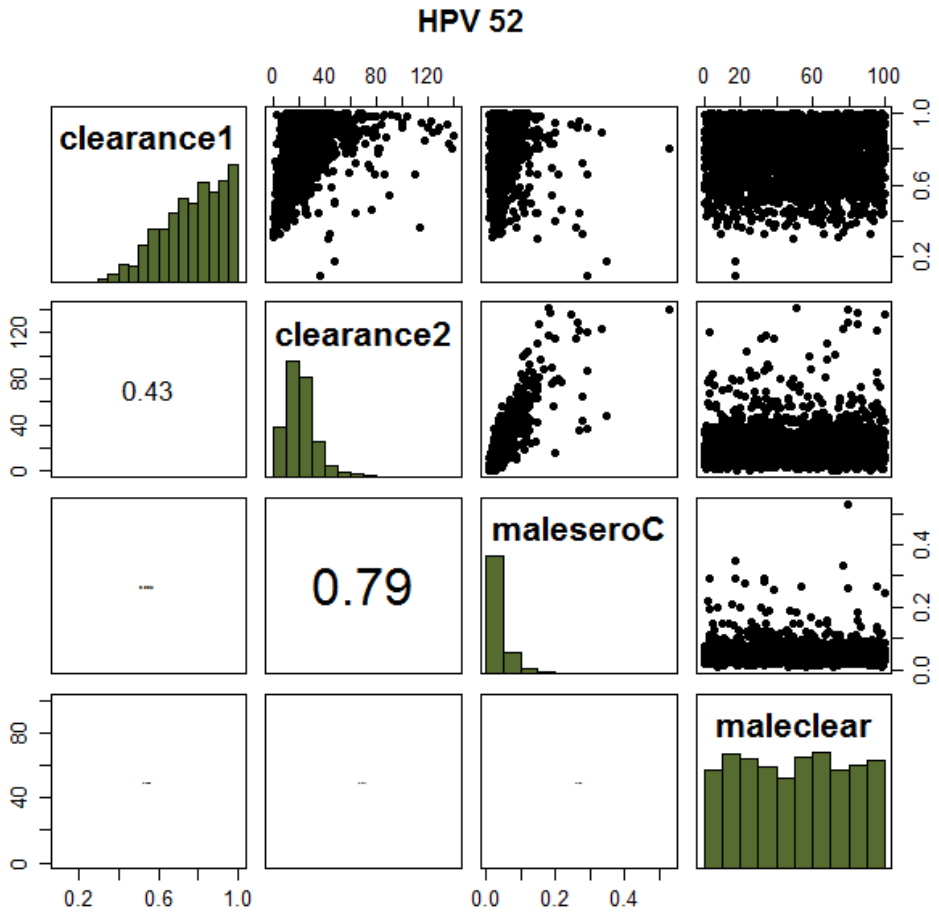



HPV 31
clearance1
clearance2
maleseroC
maleclear
0.38
0.84


HPV 33
clearance1
clearance2
maleseroC
maleclear
0.76


HPV 45
clearance1
clearance2
maleseroC
maleclear
0.26
0.70


HPV 51
clearance1
clearance2
maleseroC
maleclear
0.33
0.12
0.91


HPV 52
clearance1
clearance2
maleseroC
maleclear
0.43
0.79
0 40 80 120
0 20 60 100
0 40 80 120
0.2 0.6 1.0
0.0 0.2 0.4
0 40 80
0.2 0.6 1.0
0.0 0.2 0.4

A3 Parameterisation of Disease Progression

The disease progression was parameterised using surveillance data collected by PHE; residual LBC samples from the NHSCSP were used to identify cytological outcome, is measured as a function of age and hpv strain. We assume that each infection within an individual will progress according to the hpv strain and time since infection. Multiple HPV infections can lead the development of multiple lesions with independent progression rates. We explicitly model cytological outcomes, rather than discrete CIN states, in order to directly calibrate the model against observed cytology outcomes. The advantage of this approach is that we avoid incorporating assumptions about the sensitivity and specificity of colposcopy and cytological testing into the underlying model parameters.

We use a nested conditional probability structure to generate a model in which the probability of distinct cytological outcomes varies as a continuous function of time since infection for each strain of HPV, rather than distinct disease states.

We introduce a flexible mixed exponential model structure, in which the probability of a normal cytological outcome can decrease and then increase as a function of time post-infection (T), according to choice of parameters p_norm1 and p_norm2:

$$P[N|T=t] = 1- (1-Exp(-p\_norm1*t))Exp(-p\_norm2*t)$$

For subsequent cytological outcomes, we define a structure that results in an increasing probability of a severe outcome with time. Given an abnormality, we model the probability of a borderline outcome as a decreasing function of time since infection,

$$P[B|\ Norm^c, T=t] = Exp(-p\_bord*t)$$

Similarly, the probability of a mild and moderate outcomes given that the outcome was not normal nor borderline, and, not normal nor borderline nor mild, respectively, is described by:

$$P[Mild|\ Norm^c \text{ and } Bord^c, T=t] = Exp(-p\_mild*t)$$

$$P[Mod|\ Norm^c \text{ and } Bord^c \text{ and } Mild^c, T=t] = Exp(-p\_mod*t)$$

Finally, the probability of a severe outcome is modelled as:

$$P[Sev] = 1 – P[Norm] – P[Bord]- P[Mild]-P[Mod]$$

This leads to waves of disease progression with time. The model was fitted simultaneously for each HPV-strain, in a simulation that incorporated screening under the existing primary cytology protocol.

A Markov Chain Monte Carlo simulation, using an adaptive Metropolis algorithm, was implemented using the FME package in R to simultaneously identify the rate of change in probability of a given cytology outcome with time. Each chain was run for a length of 20,000. Clearance and transmission parameters were sampled from the posterior distributions derived previously for each HPV type; 200 distinct combinations were used in total. 50 parallel chains were generated for each clearance-transmission parameter-combination using randomly generated starting values.

As before, a thinning interval of 50 was used to remove auto-correlation within each chain. Convergence was identified using the Geweke test statistic, a test of equality of the means of the first 10% and last 50% of the markov chain. A Gelman convergence diagnostic was then used to confirm convergence of the MCMC output in the parallel chains; a comparison of the empirical variance of each parameter within each chain should be comparable to the variance from all chains combined. The final parameter distributions reflect the joint distribution of the parallel chains generated using all 200 clearance parameter-combinations.

Cancer progression

Cervical lesions can be dissected into sections of different grades, each infected with a unique high risk HPV strain, suggesting that HPV infection with different strains can lead the development of multiple lesions with independent growth rates. HPV typing was carried out on residual tissue sections from routinely obtained diagnostic biopsies of cervical cancers archived in NHS pathology laboratories [n=555] by Howell-Jones and colleagues[2]. The observed distribution of HR-HPV types and co-infections in this sample was used to estimate the number of adenocarcinomas and squamous cell carcinomas that are associated with each strain, nationally.

Co-infection of cancer–causing strains, as defined in our model, was observed in ~7.1%, and 10%, of tissue samples taken from cervical cancers, and adenocarcinomas, respectively; ~3%, and 6%, of samples were positive for both hpv-16 and hpv-18 in cervical cancers, and adenocarcinomas, respectively. We generate national cancer incidence for each model HPV-type alone plus co infection of hpv 16 and hpv 18 by scaling the incidence values with the observed distribution of types. There are not sufficient data to accurately project the co-infection with other strains; instead, we distribute the remaining joint infection cases according to the number of observed cancers associated with a single infection of each type involved.

Adenocarcinoma and Squamous cell carcinoma

The progression of women to adenocarcinoma is modelled independently form squamous cell carcinoma. It is assumed that both conditions might arise independently. We identify the waiting time to both types of invasive cancer using reported cancer incidence in combination with data form the cervical cancer audit identifying the distribution of adenocarcinoma and squamous cell carcinoma according to age. The hazard of both squamous cell carcinoma and adenocarcinoma incidence is assumed to increase as a function of time post-infection, and we model the waiting times for each hpv type using a gamma distribution. We assume that 75.9% of cancer cases are diagnosed in FIGO stage 1, and the remainder are assumed to be stage 2+ [source: audit of invasive cervical cancer, July 2011].

Age-dependent cancer survival rates

Cancer mortality rates are calculated using 1 and 5 years survival rates published by ONS. We describe the survival using an exponential decay function following diagnosis of cancer and estimate an age-dependent mortality hazard inTable . Rates are identified using a nelder-mead optimisation in R.

CONFIDENTIAL

Age at diagnosis	Rate of mortality
15-39	0.03
40-49	0.05
50-59	0.10
60-69	0.14
70-79	0.28
80-99	0.50

Table A2: Age dependent mortality rate following diagnosis of cervical cancer

Table A3: Model parameters that best describe the occurance of cytological abnormalities with time since infection.

	p.norm1	p.norm2	Expected wait to abnormality (years)	p.bord	p.mild	p.mod
HPV 16	0.334 (0.08, 1.574)	0.001 (0, 0.004)	3.0	0.723 (0.091, 1.907)	0.658 (0.046, 1.912)	0.737 (0.048, 1.906)
HPV 18	1.381 (0.233, 1.982)	0.005 (0, 0.038)	0.7	1.098 (0.13, 1.944)	0.848 (0.045, 1.939)	0.954 (0.057, 1.939)
HPV 31	0.187 (0.018, 0.938)	0 (0, 0.001)	5.3	0.575 (0.03, 1.911)	0.606 (0.023, 1.893)	0.845 (0.037, 1.948)
HPV 33	0.421 (0.121, 1.671)	0 (0, 0.003)	2.4	0.451 (0.103, 1.685)	0.123 (0.021, 0.491)	0.838 (0.065, 1.945)
HPV 45	0.15 (0.006, 0.203)	0.001 (0, 0.007)	6.6	0.91 (0.039, 1.928)	0.86 (0.031, 1.945)	0.937 (0.047, 1.939)
HPV 51	0.998 (0.126, 1.965)	0.004 (0, 0.039)	1.0	0.827 (0.062, 1.87)	0.723 (0.026, 1.887)	0.933 (0.046, 1.953)
HPV 52	0.144 (0.028, 0.43)	0.001 (0, 0.005)	6.9	0.678 (0.039, 1.901)	0.677 (0.021, 1.916)	0.841 (0.036, 1.942)

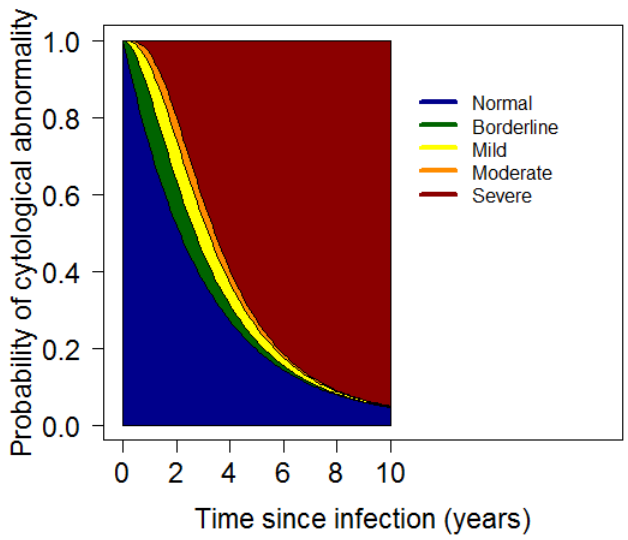
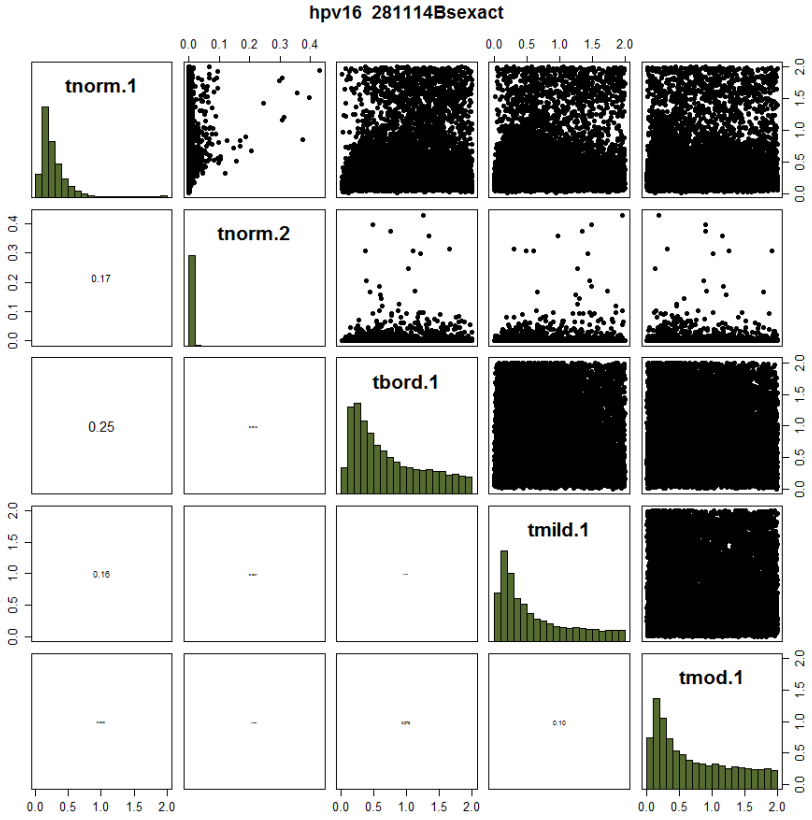
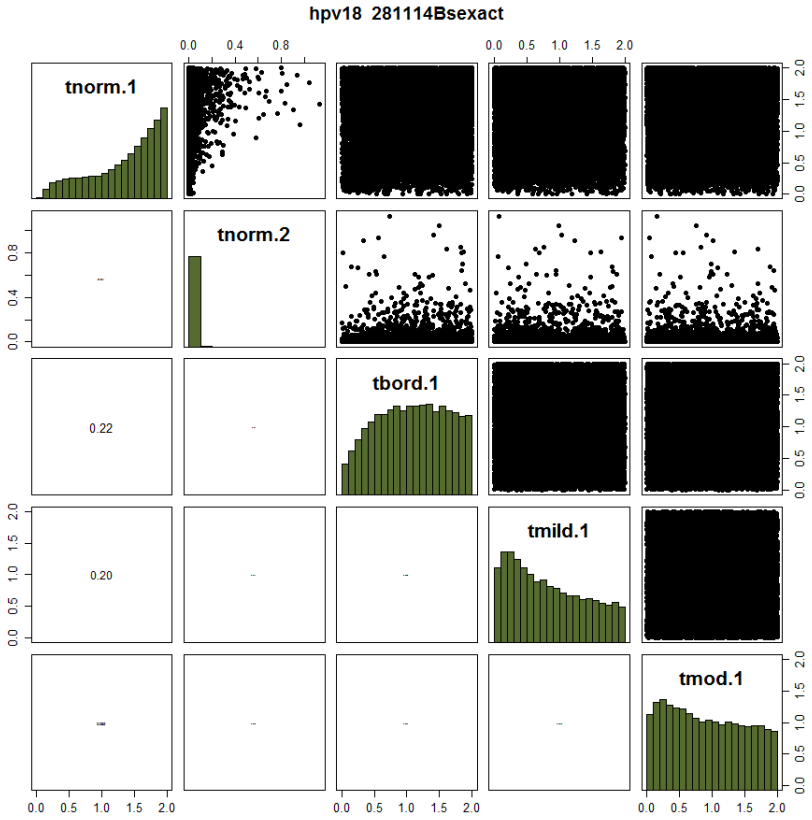
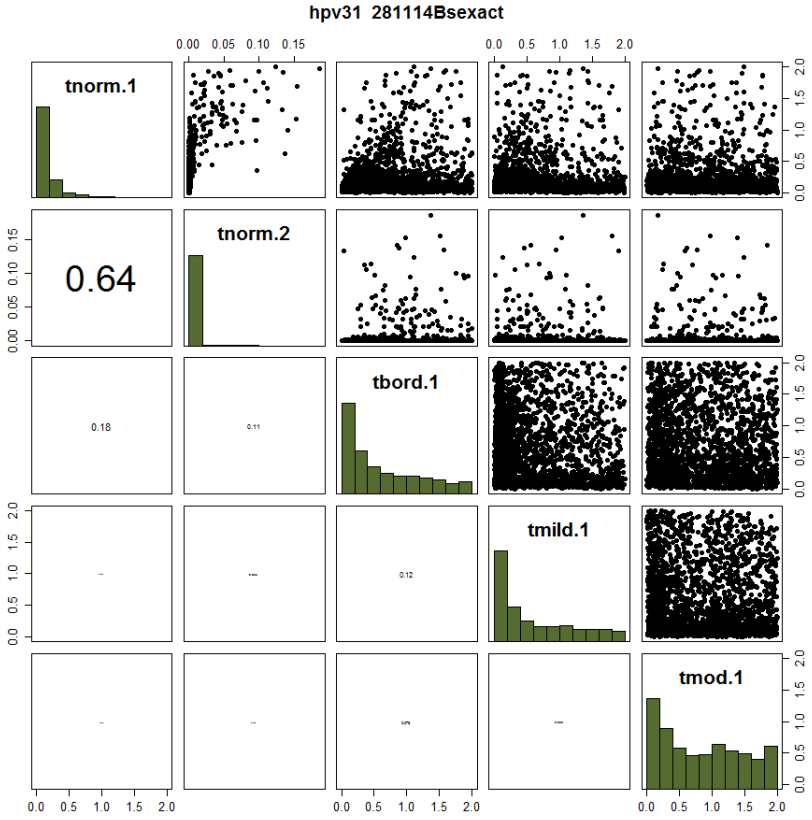


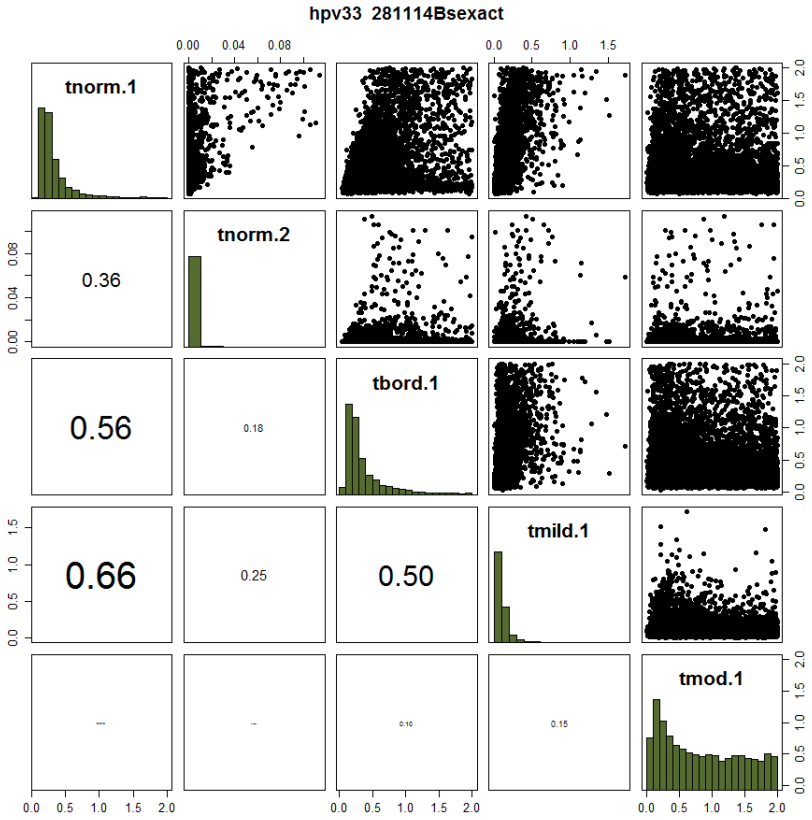
Figure A6: Illustration of probability of cytological abnormality as a function of time post- infection for HPV-16. This coincides with an decreasing probability of natural clearance with time post-infection. The time dependent probabilities are calibrated for each HPV strain using observed cytology outcome and HPV status measured as a function of age in residual samples collected by the NHSCSP.

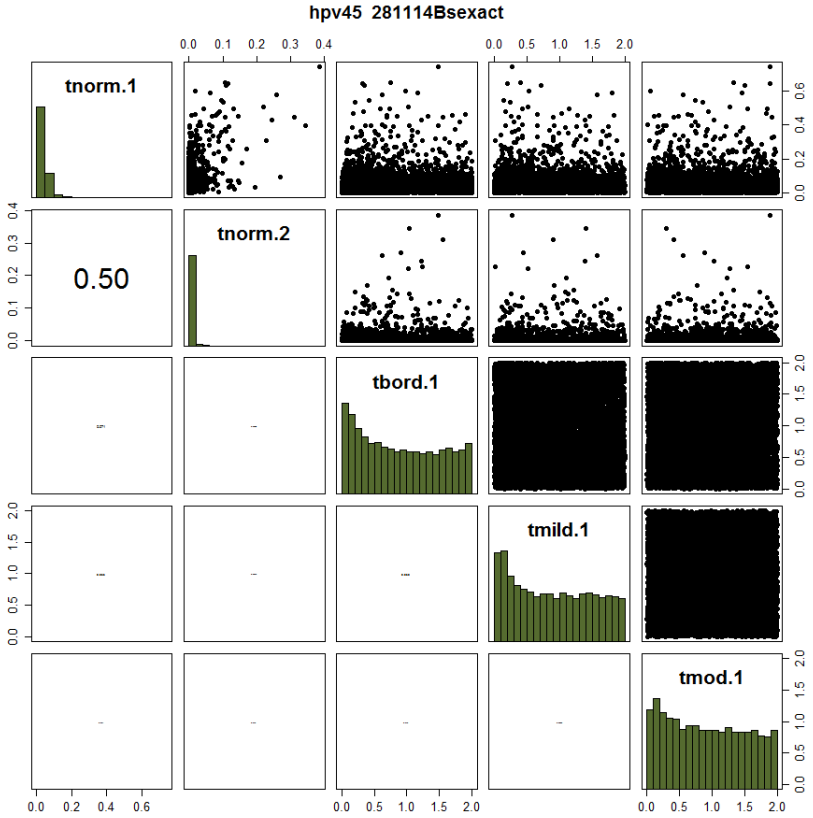
Figure A7: Posterior distributions of best fitting parameters describing cytology outcomes as a function of times for HPV-16, 18, 31, 33, 45, 51 and 52

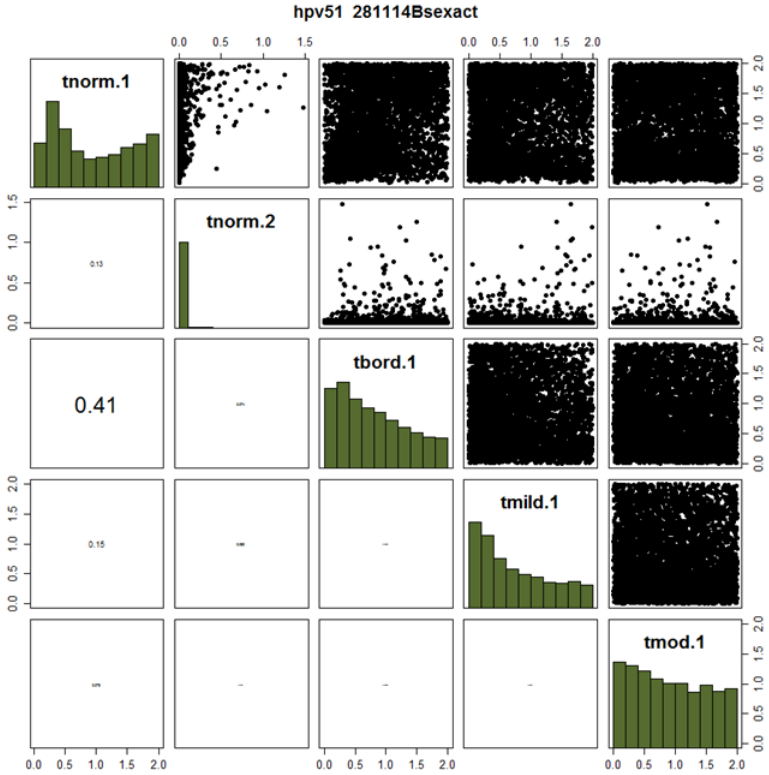

hpv16 281114Bsexact
tnorm.1
tnorm.2
tbord.1
tmild.1
tmod.1
0.17
0.25
0.16
0.10

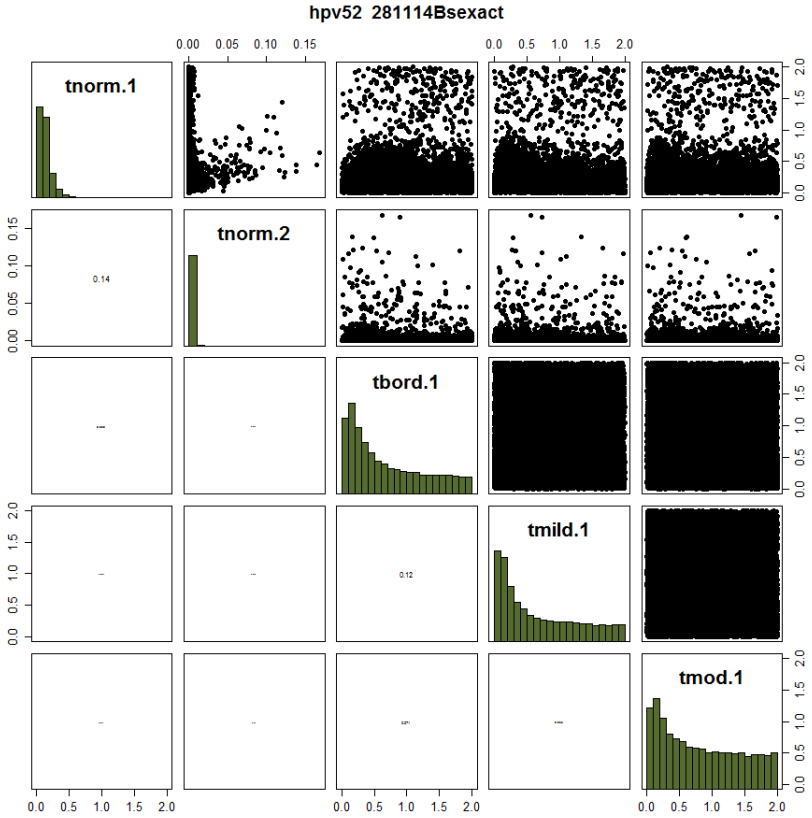

hpv18 281114Bsexact
tnorm.1
tnorm.2
tbord.1
tmild.1
tmod.1
0.22
0.20


hpv31 281114Bsexact
tnorm.1
tnorm.2
tbord.1
tmild.1
tmod.1
0.64
0.18
0.11
0.12


hpv33 281114Bsexact
tnorm.1
tnorm.2
tbord.1
tmild.1
tmod.1
0.36
0.56
0.18
0.66
0.25
0.50
0.10
0.15


hpv45 281114Bsexact
tnorm.1
tnorm.2
0.50
tbord.1
tmild.1
tmod.1


hpv51 281114Bsexact
tnorm.1
tnorm.2
0.41
tbord.1
tmild.1
tmod.1



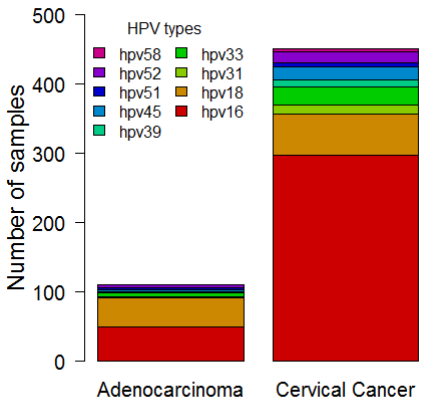


Figure A8: Hpv-type breakdown of observed adenocarcinomas and squamous cell carcinomas.

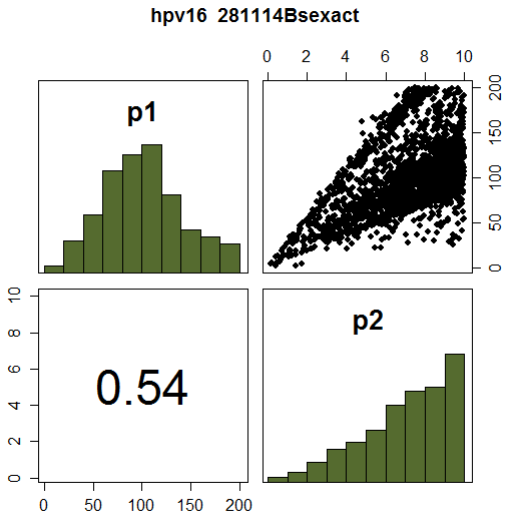
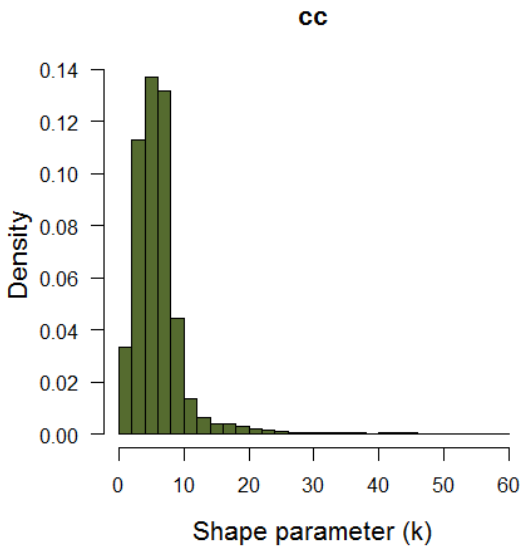
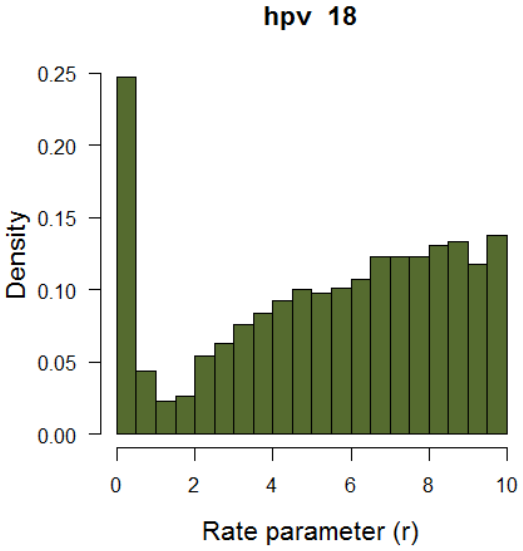
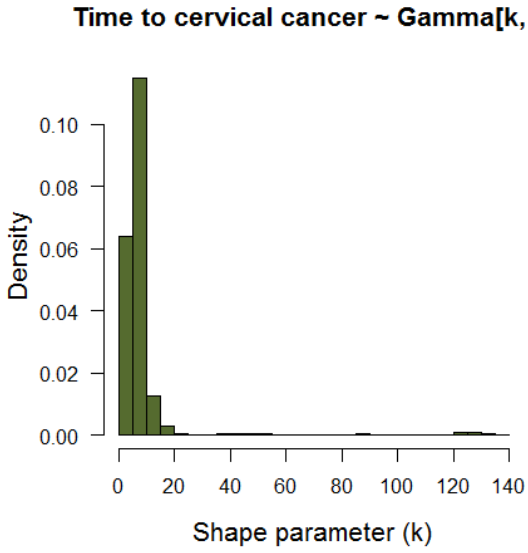
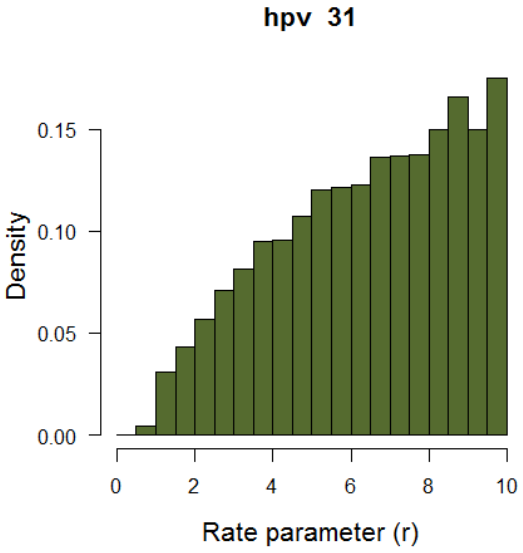


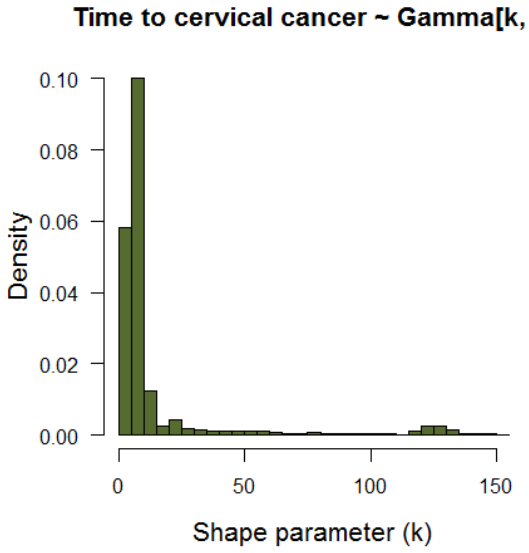
Figure A9 : Paramterisign progression to cancer using a gamma distribution, with increasing hazard with time. The shape is represented by parameter (p1), and the rate parameter is represented by p2. The subsequent plots show the posterior distribution derived for the waiting time to squamous cell carcinoma and adenocarcinoma in hpv type 16, 18,, 31, 33, 45, 51 and 52.

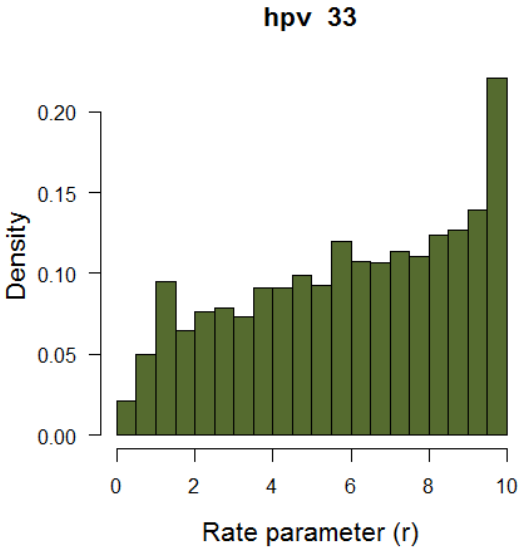


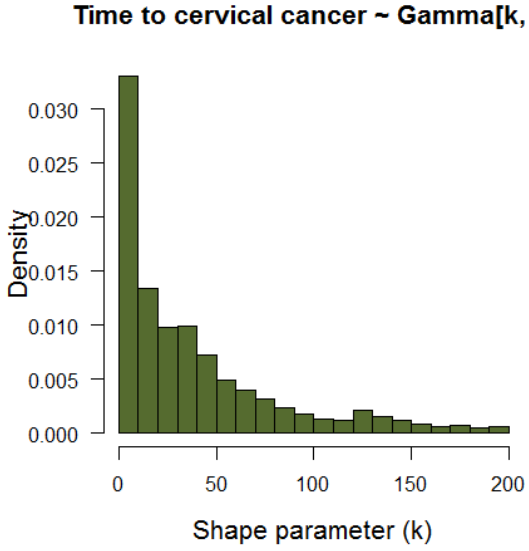


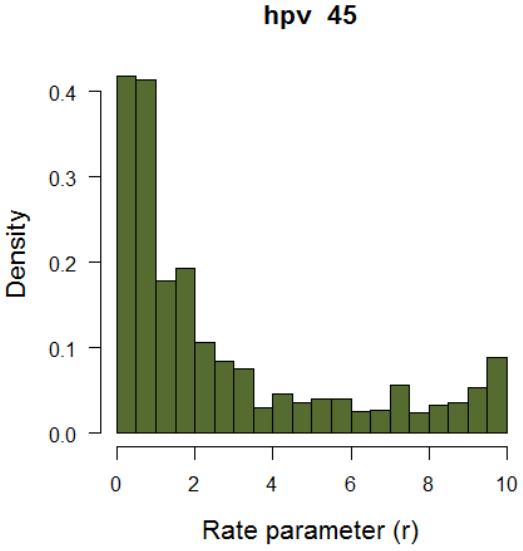


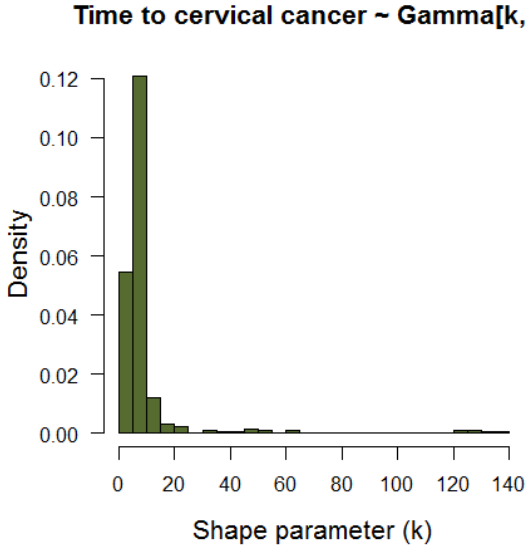


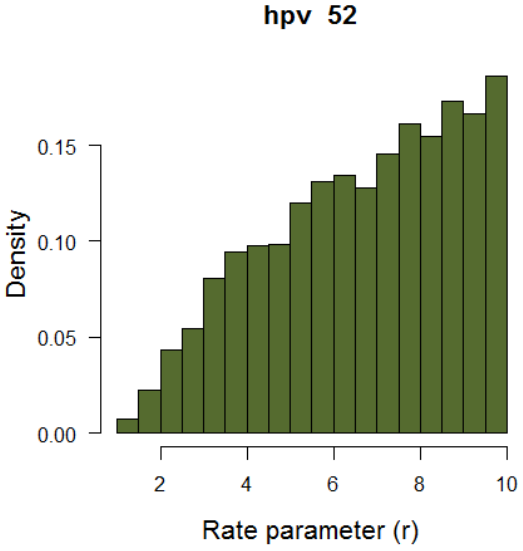

Time to cervical cancer ~ Gamma[k, r]
Density
0.10
0.08
0.06
0.04
0.02
0.00
0
50
100
150
Shape parameter (k)

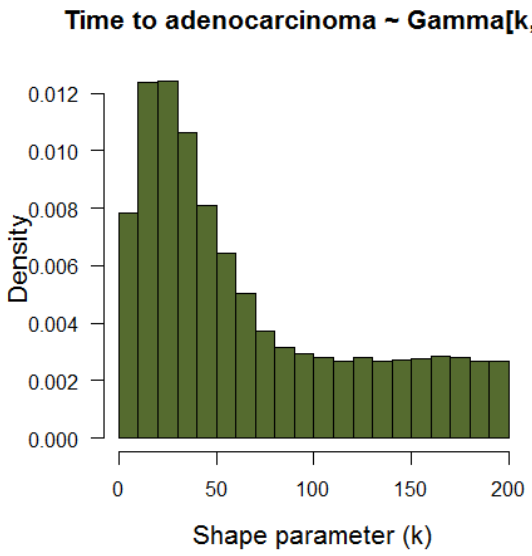

hpv 33
Density
0.20
0.15
0.10
0.05
0.00
0
2
4
6
8
10
Rate parameter (r)

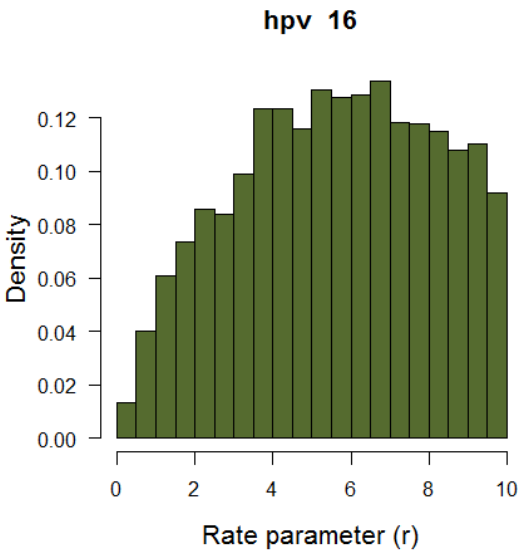

Time to cervical cancer ~ Gamma[k, r]
Density
0.030
0.025
0.020
0.015
0.010
0.005
0.000
0
50
100
150
200
Shape parameter (k)

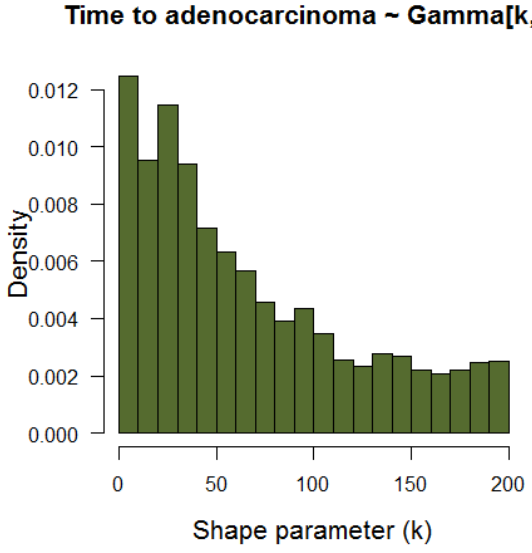

hpv 45
Density
0.4
0.3
0.2
0.1
0.0
0
2
4
6
8
10
Rate parameter (r)

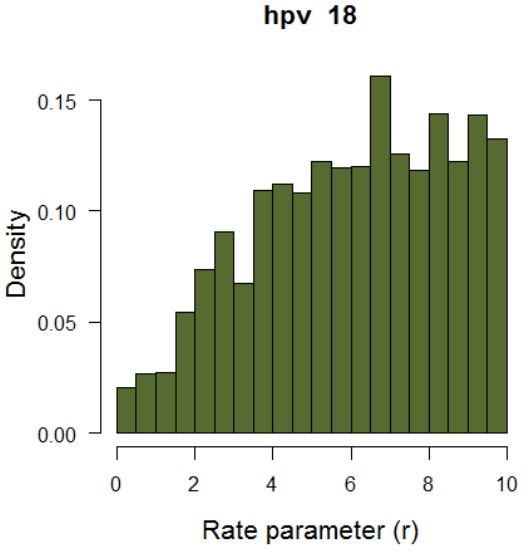

Time to cervical cancer ~ Gamma[k, r]
Density
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0
20
40
60
80
100
120
140
Shape parameter (k)

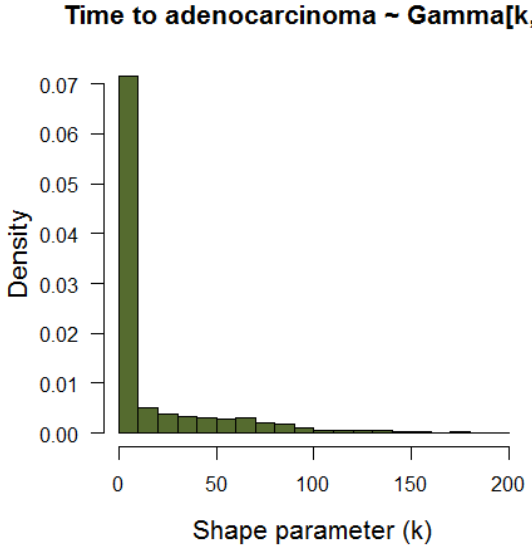

hpv 52
Density
0.15
0.10
0.05
0.00
2
4
6
8
10
Rate parameter (r)

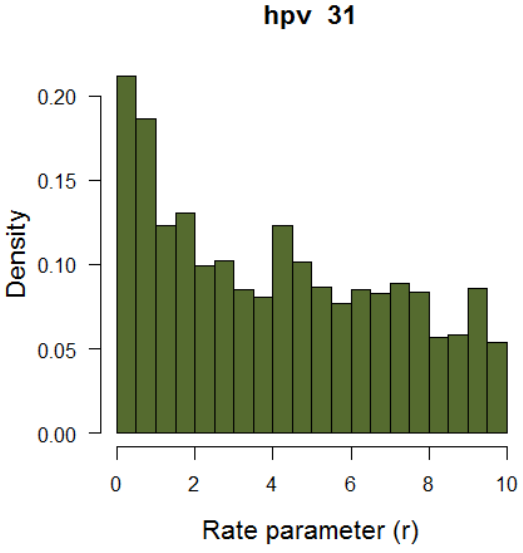

Time to adenocarcinoma ~ Gamma[k, r]
Density
0.012
0.010
0.008
0.006
0.004
0.002
0.000
0
50
100
150
200
Shape parameter (k)

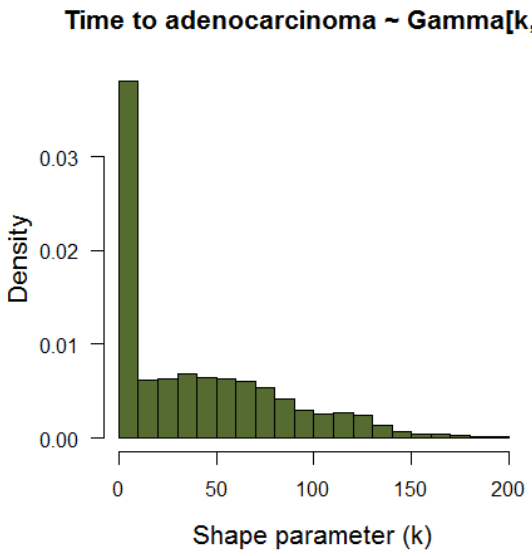

hpv 16
Density
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0
2
4
6
8
10
Rate parameter (r)

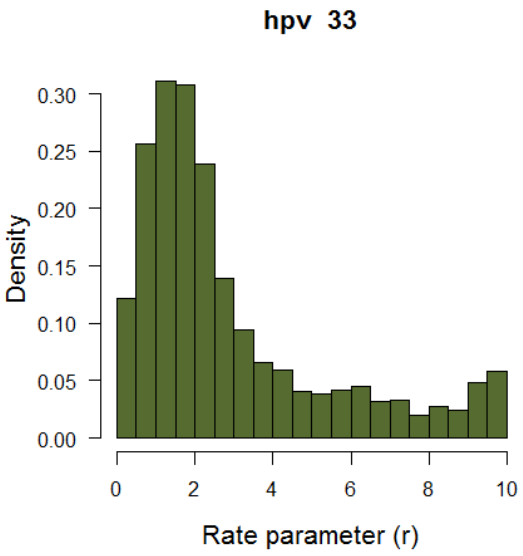

Time to adenocarcinoma ~ Gamma[k, r]
Density
0.012
0.010
0.008
0.006
0.004
0.002
0.000
0
50
100
150
200
Shape parameter (k)

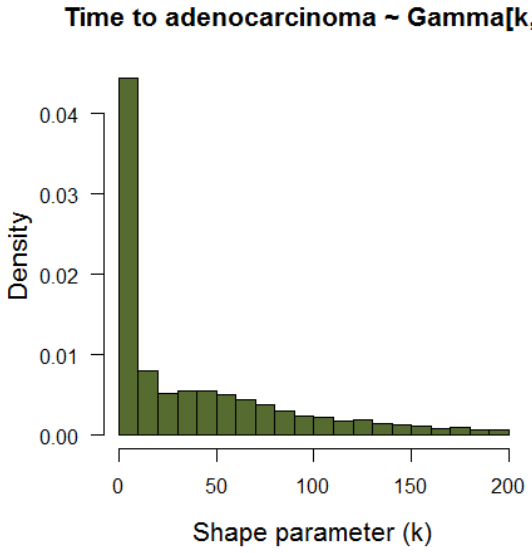

hpv 18
Density
0.15
0.10
0.05
0.00
0
2
4
6
8
10
Rate parameter (r)

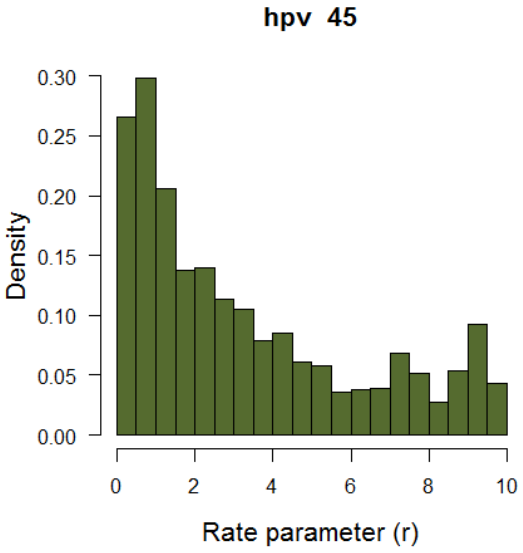

Time to adenocarcinoma ~ Gamma[k, r]
Density
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
0
50
100
150
200
Shape parameter (k)

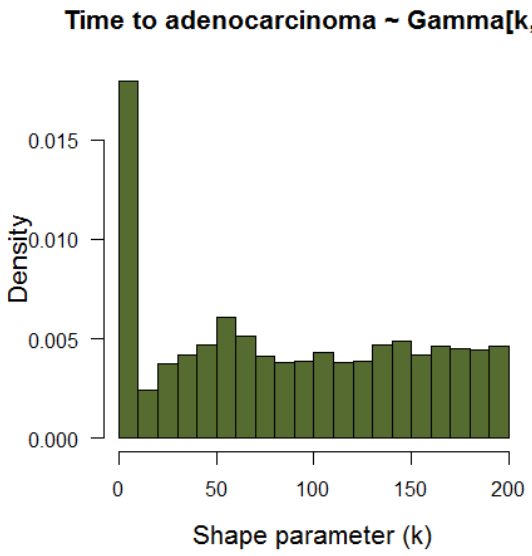

hpv 31
Density
0.20
0.15
0.10
0.05
0.00
0
2
4
6
8
10
Rate parameter (r)

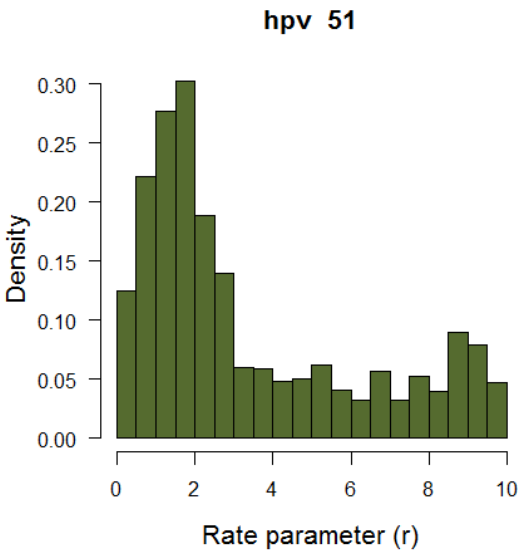

Time to adenocarcinoma ~ Gamma[k, r]
Density
0.00
0.01
0.02
0.03
0
50
100
150
200
Shape parameter (k)

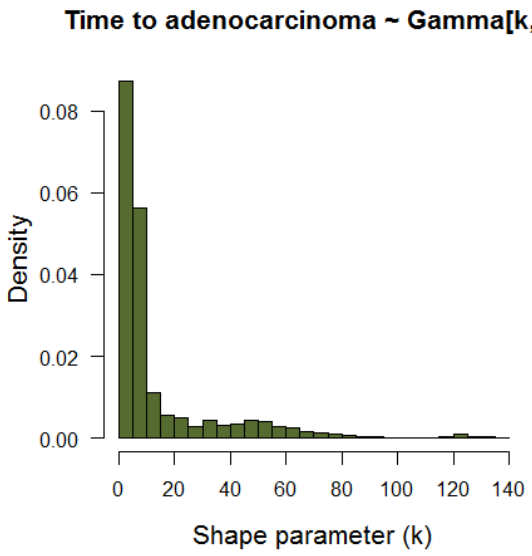

hpv 33
Density
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
2
4
6
8
10
Rate parameter (r)

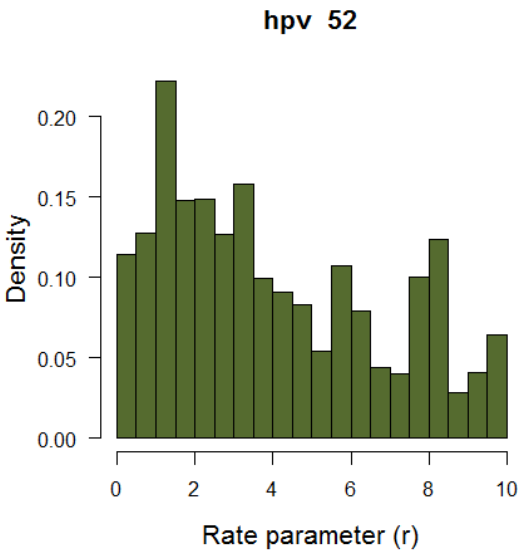

Time to adenocarcinoma ~ Gamma[k, r]
Density
0.00
0.01
0.02
0.03
0.04
0
50
100
150
200
Shape parameter (k)


hpv 45
Density
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
2
4
6
8
10
Rate parameter (r)


Time to adenocarcinoma ~ Gamma[k, r]
Density
0.000
0.005
0.010
0.015
0
50
100
150
200
Shape parameter (k)


hpv 51
Density
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
2
4
6
8
10
Rate parameter (r)


Time to adenocarcinoma ~ Gamma[k, r]
Density
0.00
0.02
0.04
0.06
0.08
0
20
40
60
80
100
120
140
Shape parameter (k)


hpv 52
Density
0.00
0.05
0.10
0.15
0.20
0
2
4
6
8
10
Rate parameter (r)

A4 Parameterisation of Screening Behaviour

Age at First Screen

A lognormally distribution is used to characterise the age at which a women attends her first cervical screen (for age 24.5 years and above) (Figure A10). A nelder-mead optimization is used to identify the distribution of age at first cervical screen that is best able to describe the observed fraction of women that have never attended screening with age [source: Cervical Screening Programme 2011-2012].

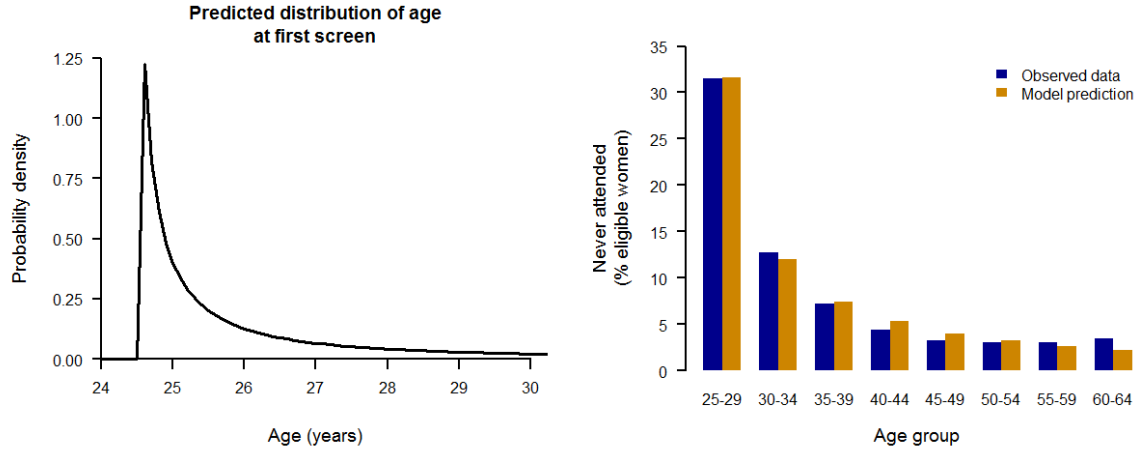
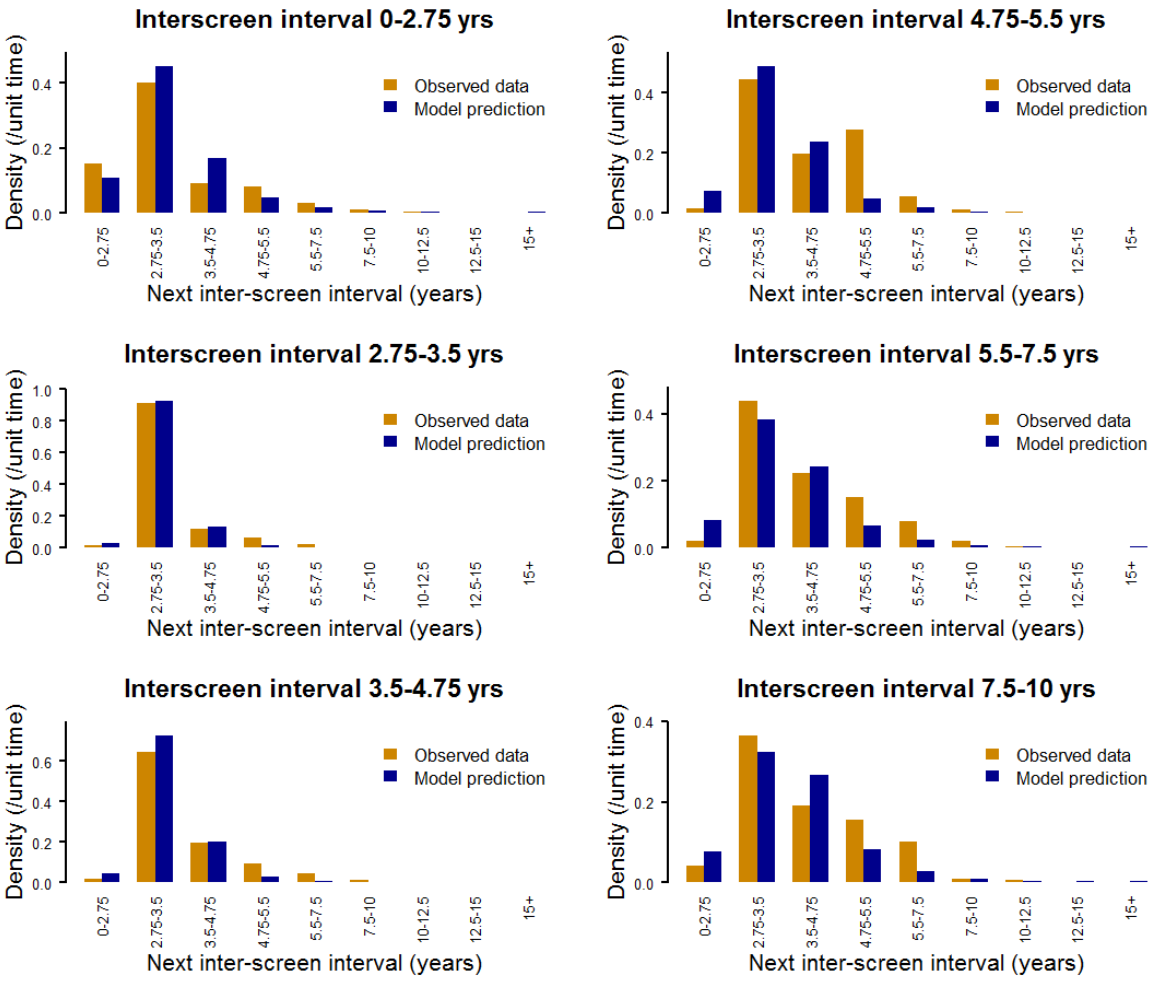


Figure A10: Age at first cervical screen**. Left panel**: Best-fit probability density (likelihood) of first attending screening at given age. **Right panel**: Model predictions and observed values for percentage of women that have never attended for screening, as a function of age.

Screening adherence

We use the time between two successive screens to identify long-term behavioural screening pattern. The data is restricted to women on routine recall with no history of abnormalities. A log-cauchy distribution of waiting time best describes the observed interval between successive screens. The interval between the last and penultimate screen is studied in women under 50 year with a prescribed interval of 3 years and data are stratified according to the previous inter-screening time (between screen (*n-2*) and *(n-1)*) – under 2.75, 2.75-3.5, 3.5-4.75, 4.75-5.5, 5.5-7.5, 7.5-10, 10-15 and 15 plus years and never screened). The observed distributions are fitted simultaneously to identify best-fitting log-cauchy scale and location parameters as a function of previous inter-screen. There is a slight increase in mean waiting time with increasing inter-screening time. This shift in expected value is accompanied by an accelerated increase in the variation of subsequent waiting times as individuals diverge form the 'prescribed' 3 year routine recall. In figure A11 we interpolate between the predicted mean and 90% interval to give a smooth distribution of inter-screening waiting that is then used to predict the time to next screen given an individual's screening history.



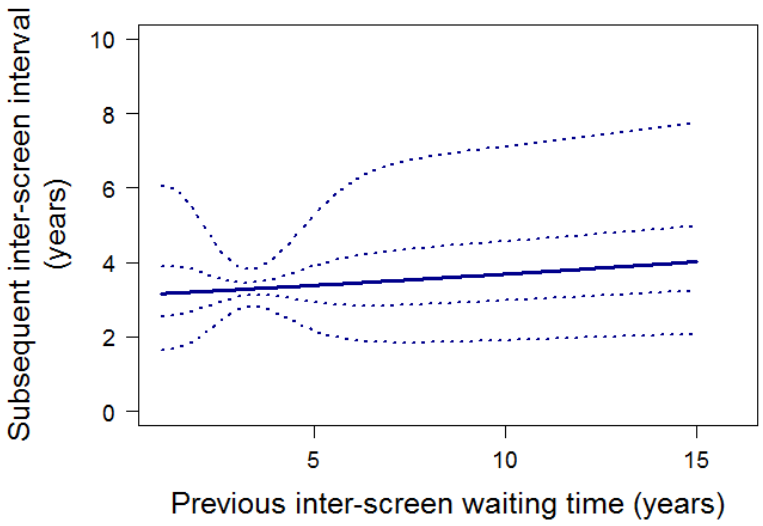


Figure A11: Distribution of waiting time to next screen. **Upper panel**: Observed and predicted distribution of 'next' inter-screen intervals when the previous interval was known to lie in the range 0-2.27, 2.75-3.5, 3.5-4.75, 4.75-5.5, 5.5-7.5 or 7.5-10 years. **Lower panel**: Best-fitting waiting time percentiles (10^{th}, 25^{th}, 50^{th}, 75^{th}, 90^{th}) as a function of previous inter-screen interval.

A5 Economic costs

	Karnon (2003 HTA) Liquid based-cytology in cervical screening	Moss (2004) Evaluation of HPV/LBC Cervical Screening Pilot Studies	Kitchner (2011 HTA) MAVARIC - a comparison of automation-assisted and manual cervical screening	Legood (BMJ 2012) Cost -effectiveness of HPV in test of cure	Kitchner (2014 HTA) Clinical effectiveness and cost effectiveness of primary HPV screenign	Primary HPV pilot site 2014	NHS supplier chain recommended price 2014	Mean cost (sd)
Consultation / sampling	**10.25**	**13.63**	**16.35**	**18.99**	**17.31**			**15.31 (3.43)**
Cytology testing								
Consumables	5.17	5.09	3.96	7.78	5.45/15.4[i]	4.37		
Equipment and labour process slides	1.00	0.91						
Smear reading costs	2.57	2.26	1.37			13.63		
Other costs (overheads , staff time spent on admin, admin costs in lab, transport)	10.88	10.53	[redacted]	[redacted]	[redacted]			
Lab cost per cytology slide excluding other costs	8.74	8.27	5.33	7.78	6.445 / 8.44 [ii]			
Total lab cost per slide	**19.62**	**18.80**	***16.03****	***18.48****	***16.97/18.96**[#]*	**18**[iii]		**18.15 (1.3)**
HPV testing								
Consumables						6.5	6.31 6.35 6.19 7.49	
Equipment costs			17.22	13.88				
Staff costs - admininstration / sample preparation / performing test / quality control					4.5 / 9.38[iv]	3.25		
Other costs (overheads , staff time spent on admin, admin costs in lab, transport)			[redacted]	[redacted]			[redacted]	
Lab cost per sample excluding other costs			17.22	13.88				
Total lab cost per sample			***20.47****	***17.13****	**4.5 / 9.38**[iv]	**9.75**	***9.56** 9.6* 9.44* 10.74**	**9.75/12.01**[v] **(0.5/4.31)**

Table A4: Comparison of screening costs from historic economic analyses of the NHS cervical screening programme in England[7–11], inflated to 2014 values. *Missing costs (overheads /other lab costs) are estimated using average of available values. Notes: (i) costs quoted for normal / abnormal outcomes; (ii) costs for primary and triage calculated using expected proportion of abnormal outcomes; (iii) using a SurePath cytology assay, final value includes equipment, staff costs and all other lab overheads and costs; (iv) manufacturers estimated cost/ sentinel sites study cost; (v) costs using values from primary hpv testing studies only / including historical hpv triage test costs.

Treatment	**Cancer stage at diagnosis**					
	1A	**1B**	**2**	**3**	**4**	**1B+**
None	4.6%	5.4%	8.6%	12.3%	19.6%	19.8%
Cone	69.6%	18.1%	0.7%	1.0%	1.4%	15.8%
trachelectomy	1.0%	5.6%	0.2%	0.0%	0.0%	1.0%
hysterectomy only	20.4%	46.0%	7.6%	1.5%	1.4%	19.8%
radiotherapy (+/- hysterectomy)	1.5%	6.9%	20.7%	24.1%	32.4%	8.9%
chemotherapy (+/- hysterectomy)	0.4%	2.0%	3.8%	6.9%	10.1%	4.0%
chemo-radio therapy (+/- hysterectomy)	2.6%	16.1%	58.4%	54.2%	35.1%	30.7%

Table A5: Treatment of cancers according to stage at diagnosis (source: Cervical Cancer Audit, 2010)

Age at diagnosis	**Cancer stage at diagnosis**					
	1A	**1B**	**2**	**3**	**4**	**1B+**
25	47.8%	35.4%	6.2%	1.8%	3.5%	5.3%
25-49	48.9%	35.7%	7.7%	2.6%	1.3%	3.8%
50-64	21.3%	33.9%	17.9%	11.3%	7.3%	8.3%
65 and over	6.6%	27.7%	26.6%	17.9%	12.9%	8.4%

Table A6: Observed state of cancer progression, according to age at diagnosis (source: Cervical Cancer Audit, 2010)

References

1. Burington B, Hughes JP, Whittington WLH, et al. Estimating duration in partnership studies: issues, methods and examples. *Sex Transm Infect*. 2010;86:84-89. doi:10.1136/sti.2009.037960.

2. Howell-Jones R, Bailey a, Beddows S, et al. Multi-site study of HPV type-specific prevalence in women with cervical cancer, intraepithelial neoplasia and normal cytology, in England. *Br J Cancer*. 2010;103(2):209-216. doi:10.1038/sj.bjc.6605747.

3. Howell-Jones R, de Silva N, Akpan M, et al. Prevalence of human papillomavirus (HPV) infections in sexually active adolescents and young women in England, prior to widespread HPV immunisation. *Vaccine*. 2012;30(26):3867-3875. doi:10.1016/j.vaccine.2012.04.006.

4. Desai S, Chapman R, Jit M, et al. Prevalence of human papillomavirus antibodies in males and females in England. *Sex Transm Dis*. 2011;38(7):622-629. doi:10.1097/OLQ.0b013e31820bc880.

5. Scherpenisse M, Mollers M, Schepp RM, et al. Changes in Antibody Seroprevalence of Seven High-Risk HPV Types between Nationwide Surveillance Studies from 1995-96 and 2006-07 in The Netherlands. *PLoS One*. 2012;7(11). doi:10.1371/journal.pone.0048807.

6. Michael KM, Waterboer T, Sehr P, et al. Seroprevalence of 34 human papillomavirus types in the German general population. *PLoS Pathog*. 2008;4(6). doi:10.1371/journal.ppat.1000091.

7. Karnon J, Peters J, Platt J, Chilcott J, McGoogan E, Brewer N. Liquid-based cytology in cervical screening: an updated rapid and systematic review and economic analysis. *Health Technol Assess (Rockv)*. 2004;8(20):1-90.

8. Moss SM, Gray A, Marteau T, Legood R, Henstock E, Maissi E. Evaluation of HPV / LBC Cervical Screening Pilot Studies Report to the Department of Health (Revised October 2004). 2004;(October).

9. Kitchener HC R. B, H. C, et al. MAVARIC - A comparison of automation-assisted and manual cervical screening: A randomised controlled trial. *Health Technol Assess (Rockv)*. 2011;15(3):1-176.

10. Legood R, Smith M, Lew J-B, et al. Cost effectiveness of human papillomavirus test of cure after treatment for cervical intraepithelial neoplasia in England: Economic analysis from NHS Sentinel Sites Study. *BMJ*. 2012;345(7882):e7086.

11. Kitchener HC, Canfell K, Gilham C, Sargent A, Roberts C, Desai M. The clinical effectiveness and cost-effectiveness of primary human papillomavirus cervical screening in England. *Health Technol Assess (Rockv)*. 2014;18(23). doi:10.3310/hta18230.