Supplementary appendix 1:

Model structure, data sources and model parameters

1 Model framework

This is a discrete time state-transition model[1] to simulate smoking prevalence among the adult population (>=18) in England, using computational language R.[2] The population was categorised into subgroups by age, sex, highest educational qualification, and smoking status. Smoking status was classified as never smokers, current smokers and former smokers. Highest educational qualifications (HEQ) are defined as in Annual Population Survey in the UK.[3] In this study, we classified HEQs into three categories: high, intermediate, and low qualifications. High qualifications (HEQ1) refer degree or higher education or equivalent; intermediate qualifications (HEQ2) include General Certificate of Education (GCE) Advanced level (A-level) or equivalent; and low qualifications (HEQ3) include General Certificate of Secondary Education (GCSE) or lower or no formal qualifications.[3] The modelling updated population subgroups by age, sex, HEG, and smoking status annually according to the probability of death, transitions from low to higher HEQs, and between smoking conditions (Fig 1a, Fig 2a).

Fig 1a. **Possible transitions across qualifications and smoking status**. HEQ1, HEQ2, and HEQ3 refer high, intermediate, and low qualifications, respectively. Ne, Cu, and Ex refer never, current, and former smokers, respectively.

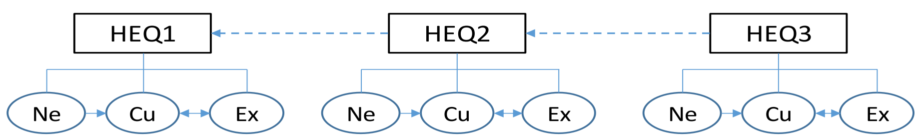
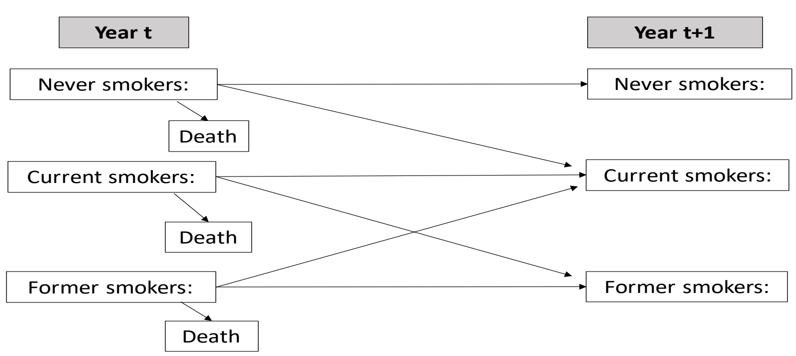


Fig 2a. Possible transitions across smoking status



Song F, *et al. Tob Control* 2021; 30:380–385. doi: 10.1136/tobaccocontrol-2019-055490

In the present study, a modelling approach was used to estimating cessation rate between two adjacent years, given other inflow to and outflow from current smokers, during 2013-2017:

a. First calculating the number of current smokers ($CS'_{t,a}$) aged *a* in year *t*, after considering deaths in current smokers, relapse among former smokers, and newly initiated smokers among never smokers, assuming there were no new quitters, in people aged *a-1* in year *t-1*.
b. Calculating the number of current smokers ($CS_{t,a}$) aged *a* in year *t* based on the observed smoking prevalence
c. Calculating: $Q_{t,a} = CS'_{t,a} - CS_{t,a}$, which was used to estimate quitters in year *t-1*. $Q_{t,a}>0$ indicates that there were new quitters aged *a-1* in year *t-1*.

2 Population and death risk

The mid-year population by age and sex in 2013 in England, and age- sex- specific death rates were obtained from the Office for National Statistics.[4] The simulation was started using the number of adult population at the start of the year 2013 in England. To estimate the size of the population at the start of the year 2013, we subtracted half of the difference in the number of population between 2012 and 2013 from the mid-year population in 2013. Then the model was used to update the population subgroups annually according to the probability of death and the transition between smoking conditions. Without considering the transition between smoking conditions, the adult population in subsequent years was estimated using the following equation:

$$N_{c,s,a+1,t+1} = N_{c,s,a,t} \times (1 - d_{c,s,a,t})$$

Where: $N_{c,s,a,t}$ refers size of the population for a given cigarette smoking status (c), sex (s), and age (a) in year t, and $d_{c,s,a,t}$ refers the annual risk of death for people given smoking status (c), sex (s), and age (a) at the beginning of year (t).

The following equation was used to estimate the death risk (d_a) for the population at the beginning of a year based on central death rates (m_a):

$$d_a \cong \frac{m_a}{1+0.5m_a}$$

The death risk based on population statistics were average estimates for never, current and former smokers. We used the following equation to estimate the death risk for never smokers:

$$d_0 = \frac{d_{all}}{P_{nev}+P_{cur}\times R_{cur}+P_{ex}\times R_{ex}}$$

Where d_{all} is the average death risk; P_{nev}, P_{cur}, and P_{ex} refer the rates of prevalence of never, current and former smokers respectively; R_{cur} and R_{ex} refer the relative risk for current and ex-smokers respectively. The risk of dying for current smokers was then calculated by: $d_1 = d_0 \times R_{cur}$, where d_0 is the death risk for never smokers and R_{cur} is the relative risk of death for current smokers. The risk of dying for ex-smokers was correspondingly calculated by: $d_2 = d_0 \times R_{ex}$, where R_{ex} is the relative risk of death for ex-smokers.

The number of people aged 18 each year was obtained from official statistics,[4] and provided exogenously to the model. The influence of immigration was not considered in this study.

3 Smoking status and data sources

The population were separated into subgroups by age, sex, HEQ, and smoking status. The smoking status of the population were classified as never smokers, current smokers and former smokers. As the cycle length of time in the present study was one year, former smokers were defined as those who stopped smoking for ≥12 months, and were further classified according to time since cessation (quit smoking for 1, 2, 3, 4, 5, 6, 7, 8, 9, and ≥10 years).

3.1 Data on smoking status in England

The rates of prevalence of current and former cigarette smoking by sex, age and HEQ in England were obtained from Annual Population Survey (APS) 2013-2017.[3] The APS has a large sample size of about 320,000 respondents, in which respondents were asked a question: “have you ever smoked cigarettes regularly”? If the answer was “yes”, a second question was asked “and do you smoke cigarettes at all nowadays”? Current smokers were defined as those who answered “yes” to both the questions, and former smokers were those who answered “yes” to the first and “no” to the second question. Before 2016, the first question asked was as: “have you ever smoked a cigarette, a cigar, or a pipe”, which resulted in an over-estimation of the prevalence of former smokers by about 28%.[5] Therefore, a ratio of 0.72 was used to adjust proportions of former smokers estimated using data from APS in 2013-2015. Due to small number of 18-19-year-olds included in the APS, smoking prevalence rates for them were initially not specified by qualifications, although the influence of smoking rates in 20-24-year-olds were taking into account in the modelling process.

Given a general trend of declining smoking prevalence in all age groups, there were some fluctuations in rates of observed smoking prevalence over time, which was likely caused by sampling errors. To facilitate the simulation evaluation, we smoothed the smoking prevalence rates from 2013 to 2017 by simple linear equations (Table 1a).

Song F, *et al. Tob Control* 2021; 30:380–385. doi: 10.1136/tobaccocontrol-2019-055490

Table 1a: Smoothed smoking prevalence rates by sex, age, qualifications during 2013-2017

	HEQ1					HEQ2					HEQ3				
Men	2013	2014	2015	2016	2017	2013	2014	2015	2016	2017	2013	2014	2015	2016	2017
18-19	20.5%	19.0%	17.4%	15.9%	14.3%	20.5%	19.0%	17.4%	15.9%	14.3%	20.5%	19.0%	17.4%	15.9%	14.3%
20-24	16.3%	15.7%	15.2%	14.7%	14.2%	23.2%	22.0%	20.8%	19.6%	18.4%	41.5%	39.7%	37.9%	36.1%	34.3%
25-29	16.8%	15.7%	14.6%	13.5%	12.4%	29.9%	28.1%	26.2%	24.4%	22.5%	42.3%	40.9%	39.5%	38.2%	36.8%
30-34	15.3%	14.3%	13.3%	12.2%	11.2%	28.1%	26.7%	25.3%	23.9%	22.5%	42.3%	40.2%	38.2%	36.1%	34.1%
35-39	14.5%	13.6%	12.8%	11.9%	11.0%	27.0%	25.9%	24.8%	23.8%	22.7%	37.0%	36.0%	34.9%	33.8%	32.8%
40-49	12.6%	11.9%	11.3%	10.6%	9.9%	23.3%	22.5%	21.8%	21.1%	20.4%	32.4%	31.2%	30.1%	28.9%	27.7%
50-59	10.3%	9.9%	9.4%	9.0%	8.5%	21.0%	19.9%	18.8%	17.6%	16.5%	29.1%	27.9%	26.8%	25.6%	24.4%
60-69	7.8%	7.5%	7.2%	6.9%	6.6%	15.5%	14.7%	13.9%	13.1%	12.3%	20.8%	20.2%	19.6%	19.0%	18.4%
70+	3.7%	3.9%	4.2%	4.4%	4.7%	6.6%	6.5%	6.4%	6.3%	6.1%	7.7%	7.2%	6.6%	6.0%	5.4%
total	12.3%	11.6%	10.9%	10.3%	9.6%	21.3%	20.3%	19.2%	18.2%	17.2%	31.3%	30.0%	28.7%	27.4%	26.1%
	HEQ1					HEQ2					HEQ3				
Women	2013	2014	2015	2016	2017	2013	2014	2015	2016	2017	2013	2014	2015	2016	2017
18-19	16.2%	15.2%	14.1%	13.1%	12.1%	16.2%	15.2%	14.1%	13.1%	12.1%	16.2%	15.2%	14.1%	13.1%	12.1%
20-24	13.2%	12.2%	11.2%	10.1%	9.1%	20.5%	19.4%	18.3%	17.2%	16.1%	41.3%	38.2%	35.2%	32.1%	29.0%
25-29	11.7%	11.1%	10.4%	9.8%	9.2%	24.5%	24.2%	23.8%	23.5%	23.2%	37.0%	35.6%	34.2%	32.8%	31.4%
30-34	9.7%	9.2%	8.7%	8.3%	7.8%	19.7%	20.2%	20.7%	21.2%	21.7%	29.3%	28.6%	27.9%	27.1%	26.4%
35-39	9.5%	8.9%	8.4%	7.8%	7.3%	20.8%	19.7%	18.7%	17.6%	16.5%	27.9%	26.9%	25.9%	24.9%	23.9%
40-49	10.5%	9.9%	9.3%	8.7%	8.1%	18.3%	17.7%	17.2%	16.7%	16.2%	27.2%	25.9%	24.6%	23.3%	22.0%
50-59	9.1%	8.9%	8.7%	8.5%	8.3%	17.9%	17.0%	16.1%	15.2%	14.4%	23.6%	22.7%	21.8%	21.0%	20.1%
60-69	7.6%	7.2%	6.8%	6.4%	6.0%	11.8%	11.4%	11.0%	10.6%	10.3%	16.2%	15.8%	15.5%	15.1%	14.8%
70+	5.1%	4.5%	3.9%	3.3%	2.6%	5.3%	5.8%	6.3%	6.8%	7.3%	9.4%	8.9%	8.4%	7.9%	7.4%
total	9.9%	9.4%	8.8%	8.3%	7.8%	18.0%	17.4%	16.7%	16.1%	15.5%	24.7%	23.7%	22.7%	21.7%	20.6%

3.2 Years since cessation for former smokers

In the present study, ex-smokers were defined as those who stopped smoking for ≥12 months, and were further classified according to time since cessation: from 1, 2, 3, · · ·, and ≥10 years. To estimate the distribution of former smokers by time since cessation at baseline in 2013, we used data from Opinion Surveys in England[6] (Table 2a).

Table 2a: Distribution of years since cessation for ex-smokers

	Years since cessation									
Men	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**≥10**
18-19	0.200	0.200	0.200	0.100	0.100	0.100	0.100	0.000	0.000	0.000
20-24	0.240	0.220	0.140	0.200	0.060	0.040	0.050	0.050	0.000	0.000
25-29	0.158	0.175	0.140	0.105	0.070	0.053	0.053	0.053	0.018	0.175
30-34	0.127	0.119	0.087	0.040	0.103	0.064	0.056	0.079	0.032	0.294
35-39	0.097	0.097	0.069	0.063	0.056	0.035	0.049	0.042	0.035	0.458
40-44	0.043	0.043	0.075	0.038	0.081	0.043	0.027	0.043	0.016	0.591
45-49	0.011	0.069	0.057	0.029	0.034	0.051	0.040	0.011	0.017	0.680
50-54	0.029	0.017	0.025	0.038	0.063	0.034	0.021	0.034	0.017	0.723
55-64	0.021	0.028	0.018	0.028	0.023	0.024	0.021	0.021	0.023	0.795
65-74	0.015	0.004	0.012	0.022	0.019	0.015	0.016	0.013	0.006	0.878
75+	0.002	0.005	0.007	0.007	0.016	0.009	0.005	0.020	0.005	0.924
	Years since cessation									
Women	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**≥10**
18-19	0.300	0.400	0.100	0.100	0.100	0.000	0.000	0.000	0.000	0.000

20-24	0.239	0.254	0.194	0.134	0.075	0.045	0.045	0.015	0.000	0.000
25-29	0.091	0.246	0.109	0.109	0.127	0.109	0.055	0.036	0.027	0.091
30-34	0.112	0.094	0.071	0.088	0.100	0.088	0.059	0.077	0.035	0.277
35-39	0.097	0.048	0.091	0.059	0.070	0.048	0.022	0.075	0.027	0.462
40-44	0.041	0.032	0.041	0.064	0.060	0.028	0.055	0.060	0.018	0.601
45-49	0.035	0.030	0.035	0.065	0.060	0.025	0.020	0.045	0.015	0.668
50-54	0.036	0.041	0.046	0.031	0.046	0.010	0.010	0.031	0.041	0.711
55-64	0.036	0.029	0.026	0.035	0.038	0.022	0.024	0.016	0.006	0.769
65-74	0.022	0.042	0.040	0.034	0.030	0.016	0.018	0.026	0.012	0.758
75+	0.011	0.019	0.009	0.015	0.031	0.009	0.011	0.020	0.006	0.869

3.3 Smoking initiation and taking up

New smoking take-ups included 18-year-old smokers, which were exogenously added to the simulated population. Due to small sample size for adults aged 18 in APS, smoking prevalence in 18-year-olds was not specified by HEQ (see Table 1a). Note that smokers aged 18 also include smokers who initiated before the age of 18. Additional new smokers were allowed among adults aged 18-24. Data on smoking initiation by sex and age were very limited. We estimated smoking initiation rates and calibrated according to changes in smoking prevalence between adjacent ages in young adults aged 18-24.[7,8]

3.4 Smoking relapse among former smokers

Relapse rates for former smokers (months since cessation: msc ≥ 12 months) were estimated using the negative exponential model, calibrated according to the relapse rate curve in a paper by Stapleton and West 2012:[9]

Relapse_rate = 5.5*0.05*exp(-0.05*msc).

4. Relative risk for current and former smokers

Relative risks (with 95% confidence intervals) of all-cause mortality by age groups for current vs. never smoking were based on data from a report of the Surgeon General[10]:

	Men	Women
Age 35-54	2.55 (1.86-3.50)	1.79 (1.27-2.53)
Age 55-64	2.92 (2.69-3.18)	2.64 (2.43-2.86)
Age 65-74	3.00 (2.89-3.13)	2.87 (2.76-2.99)
Age≥75	2.36 (2.24-2.48)	2.47 (2.37-2.58)

It was assumed that relative risk of all-cause mortality was reduced from two years since smoking cessation, and it was reduced to as never smokers by 10 years since cessation. The relative risk by years since cessation (RRc) was estimated using a negative exponential model:

RRc = 1+(RR0-1)*exp(-0.4*(c-1)),

Song F, *et al. Tob Control* 2021; 30:380–385. doi: 10.1136/tobaccocontrol-2019-055490

Where RR0 is age-specific and sex-specific relative risk for current smoking, and c is the year since cessation.

5. Sensitivity and uncertain analyses

Different projection scenarios could be evaluated in deterministic sensitivity analyses by changing key input parameters, including assumed smoking initiation, relapse and cessation rates. In addition, the model could also be used to conduct probabilistic (or stochastic) Monte Carlo simulations, according to uncertainty distribution of key input parameters, including relative risks of all-cause mortality, initiation, relapse and cessation rates. The uncertainty in estimates of relative risks was reflected by their 95% confidence intervals. The beta-PERT distribution (with λ=1) was used to evaluate uncertainty in estimates of smoking initiation, relapse and cessation rates. The estimates in deterministic simulations were used as the most likely value, the minimum value was the 50% of the most likely value, and the maximum value was the 150% of the most likely value.

6 Modelling changes in population by HEQ

Highest education qualifications are changing from lower to higher levels for a large proportion of young adults, which considerably increases modelling complexity. (Note: Modelling HEQ is much less complex and less data required than modelling other individual socioeconomic status, such as occupation-based index). In this study, we first estimated age-sex-HEQ-specific proportion of adults, based on data from APS, assuming the distribution of adults by HEQ was constant in England during 2013-2017 (Table 4a). Then, it was assumed that HEQs for adults belonging to different age-sex-specific subgroups could be changed from HEQ3 to HEQ2, or from HEQ2 to HEQ1, among adults aged 18-35 (Fig 1a).

Table 4a: Distribution of Population by age, sex and HEQ according to data from APS during 2013-2017

	Male			Female		
Age	**HEQ1**	**HEQ2**	**HEQ3**	**HEQ1**	**HEQ2**	**HEQ3**
18	2%	41%	57%	2%	48%	50%
19	3%	56%	41%	5%	63%	32%
20	6%	59%	35%	8%	63%	29%
21	17%	49%	34%	20%	50%	30%
22	28%	38%	35%	34%	35%	31%
23	34%	31%	35%	40%	28%	32%
24	38%	27%	35%	43%	26%	32%
25	39%	25%	36%	43%	23%	34%
26	39%	24%	37%	44%	22%	34%
27	38%	24%	38%	44%	20%	35%
28	40%	22%	38%	44%	21%	35%
29	40%	23%	37%	44%	21%	35%
30	40%	23%	37%	46%	20%	35%
31	41%	22%	37%	48%	20%	32%

Song F, *et al. Tob Control* 2021; 30:380–385. doi: 10.1136/tobaccocontrol-2019-055490

32	41%	22%	37%	46%	19%	34%
33	42%	21%	37%	48%	19%	33%
34	42%	21%	37%	50%	18%	33%
35-49	40%	20%	40%	48%	19%	43%
50-59	34%	25%	41%	37%	18%	55%
60+	31%	28%	42%	28%	13%	63%

Changes of HEQs happen along with increase in age from a to a+1, according to different rates HEQs for adults aged a and those aged a+1 (Table 4a). Details of calculations are described using an example of 18-year-old women in year t, and 19-year-old women in year t+1 (Fig 3a). Note that "N" refers the number of total population by age and sex in year t+1 below.

<u>For women with HEQ1</u>

1) Calculate the number of women with HEQ1 in year t+1, if no change in the rate of HEQ1:
   nP1'= 2% N
2) Calculate the true number of women with HEQ1 in year t+1:
   nP1 = 5% N
3) Calculate the difference between the estimates in 1) and 2)
   D1 = 2% N – 5% N = -3% N
4) If D1<0 (otherwise D1=0): add D1 to nP1' according to smoking status in women with HEQ2 in year t.

<u>For women with HEQ2</u>

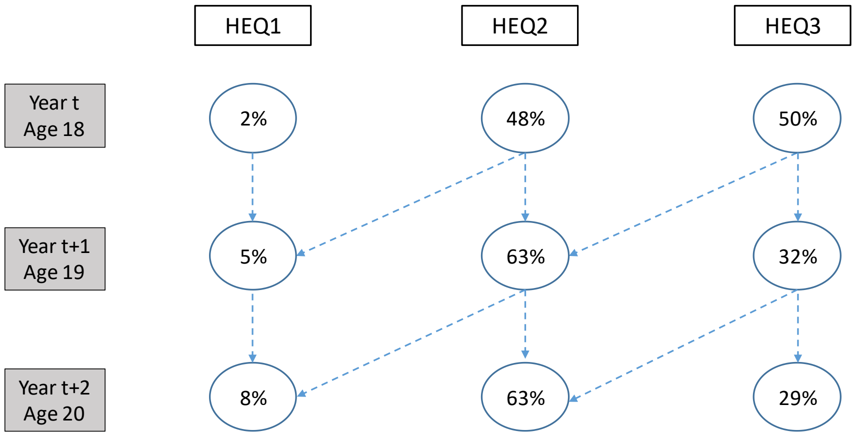
5) Calculate the number of women with HEQ2 in year t+1, after reducing by D1:
   nP2'= (48%-3%) N
6) Calculate the true number of women with HEQ2 in year t+1:
   nP2 = 63% N
7) Calculate the difference between the estimates in 1) and 2)
   D2 = 45% N –63% N = -18% N
8) If D2<0 (otherwise D2=0): add D2 to nP2' according to smoking status in women with HEQ3 in year t.

<u>For women with HEQ3</u>

9) Calculate the number of women with HEQ3 in year t+1, after reducing by D2:
   nP3'= (50%-18%) N
10) Calculate the true number of women with HEQ3 in year t+1:
   nP3 = 32% N
11) Calculate the difference between the estimates in 1) and 2)
   D3 = 32% N –32% N = 0% N
12) D2=0 (always!): no difference between nP3' and nP3.

Song F, *et al. Tob Control* 2021; 30:380–385. doi: 10.1136/tobaccocontrol-2019-055490

Fig 3a: Modelling transition from lower to higher HEQs, an example of women aged 18-20



7 References

1. Mendez D, Warner KE, Courant PN. Has smoking cessation ceased? Expected trends in the prevalence of smoking in the United States. *Am J Epidemiol* 1998; **148**(3): 249-58.
2. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/. 2013.
3. ONS. Labour Force Survey User Guide - Volume 3: Details of LFS Variables 2013. : Office of National Statistics 2014.
4. ONS. User requested data: English population estimates and deaths by sex and single year of age, 1993 to 2013. 2016. https://www.ons.gov.uk/... (accessed February 1st 2017).
5. ONS. Adult smoking habits in the UK: 2016. London: Office for National Statistics 2017.
6. ONS. Opinion Survey Report No. 40. Smoking-related behaviour and attitudes, 2008/09. Newport, South Wales: NHS Information Centre for health and social care. Office for National Statistics. , 2009.
7. Hoogenveen RT, van Baal PH, Boshuizen HC, Feenstra TL. Dynamic effects of smoking cessation on disease incidence, mortality and quality of life: The role of time since cessation. *Cost Eff Resour Alloc* 2008; **6**: 1.
8. Vugrin ED, Rostron BL, Verzi SJ, et al. Modeling the potential effects of new tobacco products and policies: a dynamic population model for multiple product use and harm. *PLoS One* 2015; **10**(3): e0121008.
9. Stapleton JA, West R. A direct method and ICER tables for the estimation of the cost-effectiveness of smoking cessation interventions in general populations: application to a new cytisine trial and other examples. *Nicotine & tobacco research : official journal of the Society for Research on Nicotine and Tobacco* 2012; **14**(4): 463-71.
10. The Surgeon General. The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General.: National Center for Chronic Disease Prevention and Health Promotion (US) Office on Smoking and Health., 2014.