E-cigarette longitudinal paper- supplementary material

Sample characteristics

eTable 1 Baseline never smokers -Sample characteristics

Variable		Valid cases (%)	Missing (%)
Ever used e-cigarette 2015	Yes	263/2800 (9.4%)	201/3001(6.7%)
Susceptible to smoking 2015	Yes	771/2927 (26.3%)	74/3001 (2.5%)
A family member smokes (mother/father/sister/brother) 2015	Yes	980/2992 (32.8%)	9/3001 (0.3%)
Number of friends who smoke 2015	Most of them	22/2581 (0.9%)	420/3001 (14.0%)
	About half	17/2581 (0.7%)	420/3001 (14.0%)
	Some of them	574/2581 (22.2%)	420/3001 (14.0%)

	None	1968/2581 (76.3%)	420/3001 (14.0%)
Sex	Female	1477/2992 (49.4%)	6/3001 (0.2%)
Ethnic group	Non-white	226/2991 (7.6%)	10/3001 (0.3%)
Family Affluence Scale 2015	Low	965/3001 (32.2%)	0
	Medium	1008/3001 (33.6%)	0
	High	1028/3001 (34.3%)	0
School year group –	year 1 mean 12.5 yrs	682/3001 (22.7%)	0
	year 2 mean 13.5 yrs	716/3001 (23.9%)	0
	year 3 mean 14.6 yrs	557/3001 (18.6%)	0

	year 4 mean 15.5 yrs	514/3001 (17.1%)	0
	year 5 mean 16.5 yrs	334/3001 (11.1%)	0
	year 6 mean 17.5 yrs	198/3001 (6.6%)	0
School 1	Urban/ medium low deprivation	858/3001 (28.6%)	0
School 2	Accessible small town/medium, low deprivation	738/3001 (24.6%)	0
School3	Other urban/ high deprivation	672/3001 (22.4%)	0
School4	Urban/ high deprivation	733/3001 (24.5%)	0

Information on loss to follow-up

In common with most longitudinal samples subject attrition differentially affected males, young people from lower socioeconomic groups and those with more smokers in their social circle. This probably means we underestimate the proportions of young people who initiate smoking during the follow up year. However, there is no difference in baseline e-cigarette status between those who were and were not lost to follow up. Therefore number of smoking initiators is probably underestimated in both groups.

eTable 2 Comparison of characteristic of cases in Model 2 (Table 3 main paper) versus those with missing data

	Missing group (n=1195)	Analysed group (n=1806)	Total never smokers 2015 (n=3001)
Sex -male	649 54.3%	867 48.0%	1479 49.3%
FAS-low	415 34.7%	550 30.5%	965 32.2%
FAS-medium	436 36.5%	572 31.6%	1008 33.6%
FAS-high	344 28.8%	684 37.9%	1028 34.3%
White ethnic group	1072 89.6%	1690 93.8%	2762 92.1%
Ever e-cigarette use 2015-yes	103 8.6%	160 8.9%	263 8.8%
Any family member smoker	435 36.4%	545 30.2%	980 32.7%
At least one friend smokes	246 20.6%	367 20.3%	613 20.4%
Susceptible to smoking	313 26.2%	458 25.4%	771 25.7%

Imputation procedure

Data was assumed to be Missing at Random (MAR) (Rubin 1976). The data was tested for Missing Completely At Random (MCAR) by conducting multiple logistic regressions on the 'missingness' of each variable using all the other variables in the model. The proportion of significant results suggested that the hypothesis of MCAR should be rejected. 'Missingness' was in the dependent variable was associated with the observed variables 'age' and 'low family affluence' and therefore likely to be MAR.

All dependent and independent variables in the estimation models were included in the imputation model with the following auxiliary variables: mother in employment, father in employment, entitlement to free school meals, whether lives with both parents. One hundred imputed datasets were produced with a burn-in of 10. Trace plots were inspected to confirm stationarity of each chain by the end of the specified burn-in period.

eTable3

Variable	Model 1-adjusted main effects model n=1806			Model 1 fully imputed n=3001		
	OR	CI	p	OR	CI	p
E-cigarette ever use 2015	2.42	1.63 to 3.60	<0.001	2.38	1.66 to 3.42	<0.001
Susceptibility to smoking 2015	3.65	2.70 to 4.94	<0.001	3.46	2.63 to 4.55	<0.001
Any family member smokes 2015	1.89	1.40 to 2.56	<0.001	1.99	1.53 to 2.59	<0.001
'At least some' friends smoke 2015	1.33	0.93 to 1.82	0.151	1.31	0.95 to 1.80	0.097
Age centred	1.12	1.01 to 1.25	0.032	1.09	0.98 to 1.20	0.102
Ethnic group-white	1.11	0.63 to 1.98	0.715	1.01	0.62 to 1.64	0.981
Sex- female	0.85	0.64 to 1.12	0.243	0.87	0.68 to 1.11	0.266

Family affluence scale low	1			1		
medium	1.56	1.09 to 2.25		1.50	1.10 to 2.04	0.010
high	1.40	0.96 to 2.05		1.51	1.08 to 2.12	0.017
School 1				1.18	0.84 to 1.67	0.332
2				1		
3				1.28	0.90 to 1.81	0.177
4				1.71	1.18 to 2.47	0.004

For the models with interaction terms two methods of imputation were attempted- treating the interaction term as 'just another variable' and imputing 'by' the interaction terms. The 'just another variable approach was unsuccessful due to convergence issues. The second method does not allow the data to be fully imputed. In the third model below we tested the effects of removing the imputed values of the dependent variable from the analysis. eTable 4 below shows the imputed model with interaction terms (column1) and the same analysis with the missing values of the dependent variable excluded (column 2).

eTable 4

Variable	Model 3 adjusted imputed model with interaction effects n=2520	Model 3 adjusted imputed model with interaction effects missing dependent variable excluded

				n=1894		
	OR	CI	p	OR	CI	p
E-cigarette ever use 2015	6.64	3.60 to 12.26	<0.001	6.06	3.16 to 11.64	<0.001
Susceptibility to smoking 2015	5.19	3.74 to 7.21	<0.001	4.13	2.99 to 5.71	<0.001
Any family member smokes 2015	1.83	1. 37 to 2.44	<0.001	1.81	1. 34 to 2.43	<0.001
'At least some' friends smoke 2015	1.51	1.07 to 2.14	0.020	1.53	1.06 to 2.20	0.022
Interaction between e-cig and any friends smoke	0.52	0.25 to 1.09	0.082	0.49	0.23 to 1.07	0.075
Interaction between e-cig and susceptibility	0.42	0.20 to 0.88	0.021	0.43	0.19 to 0.97	0.041

*Adjusted for age, sex, family affluence scale, school and ethnic group.

The table below (eTable 5) shows the difference between the fully imputed model with no interaction terms (column 2) and the fully imputed model with no interaction terms and the missing values of the dependent variable excluded.

eTable 5

Variable	Model 1-with imputed dependent variable cases dropped n=2271			Model 1 fully imputed n=3001		
	OR	CI	p	OR	CI	p
E-cigarette ever use 2015	2.37	1.64 to 3.43	<0.001	2.38	1.66 to 3.42	<0.001
Susceptibility to smoking 2015	3.50	2.67 to 4.59	<0.001	3.46	2.63 to 4.55	<0.001
Any family member smokes 2015	2.02	1.55 to 2.64	<0.001	1.99	1.53 to 2.59	<0.001
'At least some' friends smoke 2015	1.32	0.96 to 1.83	0.092	1.31	0.95 to 1.80	0.097
Age centred	1.10	1.00 to 1.21	0.032	1.09	0.98 to 1.20	0.102
Ethnic group-white	0.98	0.60 to 1.61	0.932	1.01	0.62 to 1.64	0.981
Sex- female	0.87	0.68 to 1.12	0.243	0.87	0.68 to 1.11	0.266
Family affluence scale low	1			1		
medium	1.63	1.19 to 2.24	0.003	1.50	1.10 to 2.04	0.010

high	1.56	1.12 to 2.19	0.009	1.51	1.08 to 2.12	0.017
School 1	1.31	0.91 to 1.87	0.145	1.18	0.84 to 1.67	0.332
2	1			1		
3	1.34	0.93 to 1.93	0.120	1.28	0.90 to 1.81	0.177
4	1.90	1.30 to 2.77	0.001	1.71	1.18 to 2.47	0.004

Effect of wide age range

We tested whether the wide age range employed in this study affected the results. The sample was split at age 14 and the models re-run (eTable 6).

eTable 6

Logistic regression Ever smoked 2016	Under 14s Number of obs = 920 OR		95% CI	p	14 and over Number of obs =882 OR		95% CI	p
E-cigarette ever use 2015	3.46	1.80	6.68	<0.001	2.32	1.40	3.87	0.001
Susceptibility to smoking 2015	4.00	2.44	6.55	<0.001	3.28	2.21	4.87	<0.001
Any family member smokes 2015	1.91	1.20	3.03	0.006	1.7	1.12	2.56	0.012
'At least some' friends smoke 2015	1.40	0.79	2.50	0.25	1.36	0.90	2.07	0.143
Age centred	0.90	0.61	1.31	0.574	1.27	1.01	1.60	0.039
Ethnic group -other	1.27	0.56	2.89	0.562	1	0.44	2.27	0.995
Sex- female	1.46	0.93	2.27	0.098	1.01	0.70	1.47	0.954
Family affluence scale- low	1							
medium	1.33	0.79	2.25	0.286	1.83	1.09	3.06	0.022
high	1.10	0.62	1.95	0.745	1.65	0.98	2.77	0.058
School 1	0.72	0.34	1.50	0.376	1.76	1.09	2.86	0.022
2	1							
3	1.69	0.89	3.23	0.11	0.94	0.54	1.64	0.828
4	2.15	1.12	4.12	0.022	1.21	0.66	2.21	0.533