



RESEARCH ARTICLE

**REVISED** Investigating gateway effects using the PATH study

[version 2; peer review: 2 approved]

Peter Lee <sup>1</sup>, John Fry<sup>2</sup>

<sup>1</sup>P.N.Lee Statistics and Computing Ltd, Sutton, Surrey, SM2 5DA, UK

<sup>2</sup>RoeLee Statistics Ltd, Sutton, Surrey, SM2 5DA, UK

**v2** First published: 07 Mar 2019, 8:264 (<https://doi.org/10.12688/f1000research.18354.1>)  
 Latest published: 04 Dec 2019, 8:264 (<https://doi.org/10.12688/f1000research.18354.2>)

**Abstract**

**Background:** A recent meta-analysis of nine cohort studies in youths reported that baseline ever e-cigarette use strongly predicted cigarette smoking initiation in the next 6-18 months, with an adjusted odds ratio (OR) of 3.62 (95% confidence interval 2.42-5.41). A recent e-cigarette review agreed there was substantial evidence for this “gateway effect”. As the number of confounders considered in the studies was limited we investigated whether the effect might have resulted from inadequate adjustment, using Waves 1 and 2 of the US PATH study.

**Methods:** Our main analyses considered Wave 1 never cigarette smokers who, at Wave 2, had data on smoking initiation. We constructed a propensity score for ever e-cigarette use from Wave 1 variables, using this to predict ever cigarette smoking. Sensitivity analyses accounted for other tobacco product use, linked current e-cigarette use to subsequent current smoking, or used propensity scores for ever smoking or ever tobacco product use as predictors. We also considered predictors using data from both waves, attempting to reduce residual confounding from misclassified responses.

**Results:** Adjustment for propensity dramatically reduced the unadjusted OR of 5.70 (4.33-7.50) to 2.48 (1.85-3.31), 2.47 (1.79-3.42) or 1.85 (1.35-2.53), whether adjustment was made as quintiles, as a continuous variable or for the individual variables. Additional adjustment for other tobacco products reduced this last OR to 1.59 (1.14-2.20). Sensitivity analyses confirmed adjustment removed most of the gateway effect. Control for residual confounding also reduced the association.

**Conclusions:** We found that confounding is a major factor, explaining most of the observed gateway effect. However, our analyses are limited by small numbers of new smokers considered and the possibility of over-adjustment if taking up e-cigarettes affects some predictor variables. Further analyses are intended using Wave 3 data to try to minimize these problems, and clarify the extent of any true gateway effect.

**Keywords**

Cigarettes, Confounding, E-cigarettes, Gateway effects, Modelling, Propensity score

**Open Peer Review**

Reviewer Status

	Invited Reviewers	
	1	2
<b>REVISED</b> version 2 published 04 Dec 2019		 report
version 1 published 07 Mar 2019	 report	 report

- Raymond Niaura**, NYU College of Global Public Health, New York City, USA  
**Shu Xu**, NYU College of Global Public Health, New York City, USA
- Reiner Hanewinkel**, Institute for Therapy and Health Research (IFT-Nord), Kiel, Germany

Any reports and responses or comments on the article can be found at the end of the article.

**Corresponding author:** Peter Lee ([PeterLee@pnlee.co.uk](mailto:PeterLee@pnlee.co.uk))

**Author roles:** **Lee P:** Conceptualization, Funding Acquisition, Methodology, Project Administration, Supervision, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Fry J:** Data Curation, Formal Analysis, Investigation, Methodology, Software, Validation, Writing – Review & Editing

**Competing interests:** Peter Lee, director of P N Lee Statistics and Computing Ltd, is an independent consultant to a number of tobacco companies. John Fry is a former employee of Peter Lee's company.

**Grant information:** Financial support was provided by Philip Morris Products SA, through Project Agreement No. 19 with P N Lee Statistics and Computing Ltd.

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**Copyright:** © 2019 Lee P and Fry J. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Lee P and Fry J. **Investigating gateway effects using the PATH study [version 2; peer review: 2 approved]** F1000Research 2019, 8:264 (<https://doi.org/10.12688/f1000research.18354.2>)

**First published:** 07 Mar 2019, 8:264 (<https://doi.org/10.12688/f1000research.18354.1>)

**REVISED Amendments from Version 1**

The abstract final sentence now reads "Further analyses are intended using Wave 3 data to try to minimize these problems and clarify the extent of any gateway effect". Minor changes also made earlier in the abstract.

A new paragraph in the introduction "There are two potential contributors ..... where vaping has increased" cites two new references, one mentioned by Dr Hanewinkel, the other another paper discussing general considerations.

The final sentence of the first paragraph now ends "... there is concern that vaping may encourage youths to start smoking who would otherwise not have done so, this possibility being the focus of our paper".

The next paragraph now starts "Recent analyses (Soneji *et al.*, 2017) based on ....".

The next paragraph now starts "A recent review of e-cigarettes ... considered this to provide "statistical evidence"....".

The final sentence of the next paragraph now starts "In order to gain better insight into the magnitude of any true gateway effect, information from ....".

The discussion now starts "We have described analyses aimed at deriving further insight into the magnitude of any true "gateway effect" by attempting to control better for confounding factors linked to initiation of smoking. We used a propensity score approach, which is intended to ...".

The third paragraph now ends "We are currently conducting additional work to try to obtain more precise answers by also using Wave 3 data". This, and some later changes, updates the position from saying we were planning to do these analyses.

The seventh paragraph of the discussion starting "While our analysis .." has been completely rewritten.

The second sentence in the conclusions section of the abstract has been amended to start "Indeed, it is not completely clear whether vaping actually increases subsequent uptake of cigarette smoking ....".

Two new references are included.

**Any further responses from the reviewers can be found at the end of the article**

**Introduction**

In youths, use of e-cigarettes ("vaping") and cigarette smoking are strongly associated, as shown in, e.g. Canada (Aleyan *et al.*, 2018), France (Dautzenberg *et al.*, 2016), Great Britain (Eastwood *et al.*, 2015), Korea (Lee *et al.*, 2014) and Poland (Goniewicz *et al.*, 2014), as well as the U.S. (e.g. Cooper *et al.*, 2016; Dutra & Glantz, 2014; Wills *et al.*, 2017). Since vaping significantly reduces exposure to harmful constituents compared to smoking (National Academies of Sciences Engineering and Medicine, 2018), one might expect risks from vaping to be substantially lower (Nutt *et al.*, 2014). While the benefits of introducing e-cigarettes would seem clear for smokers switching to e-cigarettes who would have continued smoking otherwise, for established smokers who are helped to quit, and for individuals who would otherwise have started smoking who start vaping instead, there are possible downsides. While risk increases may be modest for smokers intending to quit who vape instead, and for smokers who vape but retain their usual cigarette consumption, there is concern that vaping may encourage youths to start smoking who would otherwise not have done so, this possibility being the focus of our paper.

There are two potential contributors to any observed association between vaping and the subsequent initiation of smoking. One is "common liability", with youths who choose to vape already possessing attributes which make them more likely to smoke, and the other is a true causal effect of vaping, the so-called "gateway effect". Obtaining evidence to determine the extent to which any observed association is actually due to a true gateway effect, and not confounded by common liability is not straightforward, and may need to be addressed not only, as here, by detailed study of data from a prospective cohort study, but by looking at trends in smoking prevalence in countries where vaping has increased (Etter, 2018; Lee *et al.*, 2018).

Recent analyses (Soneji *et al.*, 2017) based on nine U.S. cohort studies in young people have linked previous vaping to subsequent initiation of smoking. This publication reported that among baseline never-smokers, ever vaping at baseline strongly predicted initiation in the next six to 18 months, with an odds ratio (OR) of 3.62 (95% confidence interval (CI) 2.42-5.41) after adjustment for various predictors of initiation. Similarly baseline past 30-day vaping also predicted subsequent 30-day cigarette use (OR 4.25, 95% CI 2.52-7.37).

A recent review of e-cigarettes (National Academies of Sciences Engineering and Medicine, 2018) considered this to provide "substantial evidence" of a gateway effect, noting the "wide range of covariates" that the relevant studies had considered, and thought it "unlikely" that confounding entirely accounts for the association, as reductions in the association following adjustment were not consistently observed.

While some studies (Barrington-Trimis *et al.*, 2016; Primack *et al.*, 2015; Primack *et al.*, 2016) do report that the association increases following adjustment, many more (Best *et al.*, 2018; Conner *et al.*, 2018; Hammond *et al.*, 2017; Hornik *et al.*, 2016; Leventhal *et al.*, 2015; Loukas *et al.*, 2018; Lozano *et al.*, 2017; Miech *et al.*, 2017; Spindle *et al.*, 2017; Unger *et al.*, 2016; Watkins *et al.*, 2018; Wills *et al.*, 2017) report a decrease. Furthermore, though adjusted associations are usually statistically significant, adjustment is often limited. Factors never considered include, for example, school performance, parental smoking and peer attitudes to smoking. In order to gain better insight into the magnitude of any true gateway effect, information from a large cohort study which collected data on very many factors would therefore clearly be useful, as would gaining some insight into the extent of bias resulting from misclassification of such variables.

Here we report detailed analyses of the gateway effect based on Wave 1 (2013–2014) and Wave 2 (2014–2015) of the Population Assessment of Tobacco and Health (PATH) study (Berry *et al.*, 2019b; Hyland *et al.*, 2017), a longitudinal cohort study in the U.S. supported by federal funds. The databases provide extensive information on tobacco product use and on many other factors possibly linked to smoking initiation. At each Wave, data are separately collected for youths aged 12–17 and adults aged 18+, and our analyses, which concern smoking

initiation by youths, use the youth data of each Wave, plus Wave 2 data for adults previously youths at Wave 1.

## Methods

### The main analysis

The main analysis considers those who had never smoked cigarettes by Wave 1 and who, at Wave 2, had information available on whether initiation of cigarette smoking had occurred. Use of other tobacco products is not considered. For a youth to be considered, data should be available on each of five demographics (age, sex, Hispanic origin, race, region) and on vaping.

The analyses, which relate ever having vaped by Wave 1 to initiation of cigarette smoking by Wave 2, after adjustment for factors linked to e-cigarette use recorded at Wave 1, was conducted in two steps.

**Step 1.** In step 1, Wave 1 data were used to develop a propensity score for e-cigarette use based on the five demographic variables and on 60 smoking predictor variables selected from a much longer list. As described more fully in the Extended Data, we ignored questions only asked in a population subset, only really relevant to smokers, or of dubious relevance. Also, where many related questions were asked, attention was limited to those seemingly more likely to be relevant.

We used a logistic regression model where the propensity for vaping ( $P_i$ ) for a youth  $i$  ( $i = 1$  to  $n$ ) was linked to various smoking predictors  $x_{ij}$  ( $j = 1$  to  $m$ ) by

$$\log_e\left(\frac{P_i}{1-P_i}\right) = a_0 + \sum_{j=1}^m a_j x_{ij}$$

For a given set of predictors, we refer to the value of the term on the left as the propensity score.

All logistic regression analyses were weighted by the person-level weights provided on the PATH database, with the weights normalized to sum to 1.

Introducing all 65 variables simultaneously into the model would have involved two problems. First, as the analysis required individuals with complete data on all variables, substantial information may be lost. Second, analyses including very many variables sometimes fail to solve. We therefore introduced variables in stages, using groups of conceptually-related variables, with missing values likely to be on the same individuals.

At stage 1 the variables were divided into groups numbered 1–11. In each group, an analysis was carried out for each variable individually, followed by a forward stepwise approach with the most significant variable introduced first, then the next, until no further variables in the group significant at  $p < 0.01$  could be introduced. At stage 2 significant variables from the first stage were divided into three groups (A, B and C), and a forward stepwise approach again used to identify significant variables.

Finally, at stage 3, the stepwise approach was applied to the stage 2 significant variables to generate a final list of variables

for the propensity score, which was then re-calculated based on youths with complete data on all these variables.

At each stage, the analyses involved all participants with complete data for each variable considered in the group being analysed.

**Step 2.** Step 2 involved the outcome of interest, initiation of cigarette smoking between Wave 1 and Wave 2. The first analysis was a weighted logistic regression analysis to determine the unadjusted OR and 95% CI for the association of ever vaping at Wave 1 with subsequent initiation of cigarette smoking. The second analysis was similar but adjusted for propensity by dividing the youths into five quintiles of the propensity score, the separate OR (95% CI) values for each stratum being then combined to form an overall propensity-adjusted estimate. Propensity-adjusted analyses were also conducted using the score as a continuous variable, and also using the variables making up the score individually rather than combined.

### Sensitivity analyses

Five sets of sensitivity analyses were conducted linking vaping to initiation of cigarette smoking, along the lines of the main analysis. For each set, ORs (95% CIs) were again calculated with no adjustment for propensity, adjustment as quintiles, adjustment as a continuous variable, and adjustment for the variables making up the score.

**Sensitivity analysis 1** Youths who, by Wave 1, had ever used any other tobacco product (i.e. than cigarettes or e-cigarettes) were excluded. As there were considerably fewer significant predictor variables from the stage 1 analyses, the stage 2 analyses were omitted.

**Sensitivity analysis 2** Here ever users of other tobacco products by Wave 1 were not excluded but use of other products was included as an extra predictor. The stage 1 and 2 analyses were not repeated. Rather the final model used was one that included the same final set of variables plus that for ever using other tobacco products.

**Sensitivity analysis 3** Whereas the main analysis and sensitivity analyses 1 and 2 linked a propensity score for ever vaping by Wave 1 to ever cigarette smoking by Wave 2, sensitivity analysis 3 linked a propensity score for current e-cigarette use at Wave 1 to current cigarette smoking at Wave 2, last 30-day use being considered current. Again, as few significant variables emanated from the stage 1 analyses, the stage 2 analyses were omitted.

**Sensitivity analysis 4** Whereas all the analyses described above relate to a propensity score for vaping, sensitivity analysis 4 was essentially the same as for the main analysis but based on a propensity score for ever cigarette smoking.

**Sensitivity analysis 5** Sensitivity analysis 5 was also like the main analysis, but the propensity score was based on ever use of any tobacco product.

For sensitivity analyses 4 and 5, the full three-stage process described in Step 1 was used to determine the variables included in the propensity score.

### Analyses investigating residual confounding

In the main analysis, the propensity score for e-cigarettes used data provided by youths at Wave 1. Although there was no gold standard to validate reported answers, it seemed possible that more accurate predictors could be based on data from Wave 1 and 2 combined. For those variables forming the propensity score we investigated whether there was further useful information available in Wave 2 and, if so, created a revised variable. How this was done is detailed in the Results section, the procedure depending on which variables were selected for the propensity score. Once the revised predictor variables were created, the main analyses were rerun using a modified propensity score.

### Software

Relevant data were transferred for analysis to a ROELEE database, and analysed using the ROELEE program (Release 59, Build 49). SAS version 9.4 (SAS Institute Inc, 2017) was also used to check some results from ROELEE and generate results when ROELEE failed to converge. The GLM Package and the Step function from the R Program (<https://www.r-project.org>) could be used to run all the analyses.

## Results

### Main analysis

The propensity score for ever vaping by Wave 1 was developed using five demographic variables and 60 other predictors (Table 1). Each variable, except for variable 52, depended on a question involving a few possible answers (see Table 1 footnotes) with the regression analyses estimating a coefficient per level. Exceptionally, variables 10–13, where numbers of youths were very small for some levels, were treated as continuous variables in analysis.

Stage 1 in developing the propensity score involved separate regression analyses within each of the 11 groups. As Table 1 shows, 38 variables were eliminated from consideration at that stage, with 27 retained for stage 2, 8 considered in group A, 10 in group B and 9 in group C. Following eliminating 9 more variables at stage 2, 18 variables entered stage 3 with 6 more eliminated. After rerunning the regression analysis based on 10,671 youths with data on all 12 predictors, rather than 10,361 with data on 18 predictors, the final model was as shown in Table 2.

Ever e-cigarette use was independently associated with older age, male sex, use of alcohol and prescription drugs, social networking, and preferring exciting and unpredictable friends. It was also associated with cohabitants using tobacco, parents or guardians not being very upset if they found the youth using tobacco, agreeing that some tobacco products are safer than others, the youth being curious about smoking, and the youth thinking they will smoke a cigarette in the next year. Note that, for these last two variables, the grading system ascribed lower scores for greater curiosity or greater likelihood to smoke cigarettes in the next year so the fitted ORs were <1. The results for the variable regarding enjoying using tobacco was less straightforward to interpret as very few youths strongly agreed

they would enjoy it. However, those who strongly disagreed that they thought that they would enjoy using tobacco had a clearly lower odds of ever e-cigarette use than those who simply agreed or disagreed.

As Table 3 shows, the unadjusted OR for the association of vaping by Wave 1 with cigarette smoking initiation by Wave 2 was 5.702 (95% CI 4.334-7.502). The OR was markedly reduced by adjustment for the propensity score, whether as quintiles (2.476, 1.852-3.310), as a continuous variable (2.474, 1.791-3.419), or for the 12 variables making up the score (1.847, 1.347-2.533). Table 3 also shows the effects of introducing the variables successively. With one minor exception, introducing each variable reduced the OR, the largest reductions relating to the first four variables considered, which in combination already reduced the OR to 2.185 (1.608-2.969).

### Sensitivity analyses

Table 4 summarizes the sensitivity analysis results, comparing them with those from the main analysis. While the number of significant variables included varies between analyses, all show that adjustment markedly reduces the unadjusted association, reducing ORs of over 5 to less than 3. The effect of adjustment was always greater when made for each of the individual variables making up the score. Relating ever vaping by Wave 1 to initiation of cigarette smoking by Wave 2 (main analysis, sensitivity analyses 1 and 2), the lowest adjusted ratio of 1.586 (1.194-2.198) is seen in sensitivity analysis 2, where adjustment is made for 13 predictor variables, including smoking of other products. Here, the adjustment explains 87.5% of the unadjusted association (as estimated from the ratio of the excess ORs, i.e. OR – 1). Sensitivity analysis 3, which concerns current (last 30 day) rather than ever use of both products also produced similar results, though the estimates are more variable due to the very few new cigarette smokers among e-cigarette users. Sensitivity analysis 4, where the score was based on variables linked to Wave 1 cigarette smoking rather than vaping also gave similar results, as did sensitivity analysis 5, where the score was based on variables linked to use of any tobacco product.

### Residual confounding

The propensity score used in the main analyses was revised using modified versions of the 12 predictor variables. The age range of 12–14 or 15–17 at Wave 1 was modified to be 12–13, 14, 15–16, 17 depending on whether 12–14 year-olds at Wave 1 were 15 at Wave 2, and whether 15–17 year-olds at Wave 1 were adults at Wave 2. Gender was unchanged, being consistent between waves. For three ever use variables (alcohol; prescription drugs; social networking) non-users at Wave 1 were now considered users if use was reported at Wave 2. For four variables where questions were asked at both waves (reaction if parent or guardian found using tobacco; think would enjoy tobacco; relative safety of tobacco products; cohabitant uses tobacco) the level most associated with vaping use was used (i.e. maximum for reaction and minimum for the other three). These questions were only asked of youths, so if the participant became an adult at Wave 2, the Wave 1 response was used. For the other three variables

**Table 1. Wave 1 predictor variables used with details of which stage they were eliminated from consideration.** For each of the 65 Wave 1 predictor variables considered, Table 1 gives details of their type (graded or continuous) and of the stage at which they were eliminated from consideration for the final propensity score.

Group	Variables	Levels or continuous <sup>a</sup>	Eliminated at stage 1 <sup>b</sup>	Eliminated at stage 2 <sup>c</sup>	Eliminated at stage 3 <sup>d</sup>
1	<b>Demographic variables</b>				
1	Age range	2	No → A	No → F	No → P
2	Gender	2	No → A	No → F	No → P
3	Hispanic origin	2	Yes		
4	Race	3	Yes		
5	Census region	4	Yes		
2	<b>Education, internalizing disorders</b>				
6	Highest grade or year at school completed by parent	5	Yes		
	Last time significant problems with:				
7	(a) Feeling very trapped, lonely, sad, blue, depression	4	No → A	No → F	Yes
8	(b) Sleep trouble – such as bad dreams, sleeping restlessly	4	Yes		
9	(c) Becoming very distressed when something reminded of past	4	No → A	Yes	
3	<b>Susceptibility to smoking</b>				
10	Ever been curious about smoking a cigarette	4	No → B	No → F	No → P
11	Think you will smoke a cigarette in the next year	4	No → B	No → F	No → P
12	Think you will try a cigarette soon	4	Yes		
13	Would smoke a cigarette if one your best friends offered you one	4	Yes		
4	<b>Externalizing disorders</b>				
	Last time 2+ times				
14	(a) Lied or conned to get things you wanted or avoid doing something	4	No → B	Yes	
15	(b) Had a hard time paying attention at school, work or home	4	No → B	Yes	
16	(c) Had a hard time listening to instructions at school, work or home	4	Yes		
17	(d) Were a bully or threatened other people	4	Yes		
18	(e) Started physical fights with other people	4	No → B	Yes	
19	(f) Felt restless or the need to run around or climb things	4	Yes		
20	(g) Gave answers before the other person finished asking the question	4	Yes		
5	<b>Attitudes to tobacco</b>				
	Agree/disagree				
21	(a) I think I would enjoy using tobacco	4	No → B	No → F	No → P
22	(b) Using tobacco would be energizing	4	Yes		

Group	Variables	Levels or continuous <sup>a</sup>	Eliminated at stage 1 <sup>b</sup>	Eliminated at stage 2 <sup>c</sup>	Eliminated at stage 3 <sup>d</sup>
	23 (c) Using tobacco would help me reduce or handle stress	4	No → B	Yes	
	24 (d) Using tobacco would help me calm down when I am angry	4	Yes		
	25 (e) Using tobacco would help me control my weight	4	Yes		
	26 (f) Using tobacco would help me feel more comfortable at parties	4	Yes		
	27 How much you think people harm themselves when they smoke cigarettes every day	4	Yes		
	28 How much you think people harm themselves when they smoke cigarettes some but not every day	4	Yes		
	29 How long do you think someone has to smoke cigarettes before it harms their health	6	Yes		
	30 Agree/disagree: some tobacco products are safer than others	5	No → B	No → F	No → P
	31 Seen a tobacco sweepstakes ad in past 6 months	2	Yes		
6	<b>Substance use</b>				
	32 Ever used alcohol at all	2	No → B	No → F	No → P
	33 Ever used prescription drug not prescribed to you: Ritalin or Adderall	2	No → B	No → F	No → P
	34 Ever used prescription drug not prescribed to you: Painkillers, sedatives or tranquilizers	2	Yes		
	35 Ever used substance: Cocaine or crack	2	Yes		
	36 Ever used substance: Stimulants like methamphetamine or speed	2	Yes		
	37 Ever used substance: Any other drugs like heroin, inhalants, solvents or hallucinogens	2	Yes		
7	<b>Risk taking variables</b>				
	Agree/disagree:				
	38 (a) Like to do frightening things	5	Yes		
	39 (b) Like new and exciting experiences even if I have to break the rules	5	No → C	No → F	Yes
	40 (c) Prefer friends who are exciting and unpredictable	5	No → C	No → F	No → P
8	<b>School performance, Health Behaviour Overall, Anxiety</b>				
	41 Youth's grade performance in school in past 12 months	10	No → C	No → F	Yes
	42 How often youth missed school due to illness in past 12 months	5	Yes		
	43 Youth's overall health status (as reported by parent or guardian)	5	Yes		
	44 Last time problems with: feeling very anxious, nervous, tense, scared, panicked	4	No → C	Yes	

Group	Variables	Levels or continuous <sup>a</sup>	Eliminated at stage 1 <sup>b</sup>	Eliminated at stage 2 <sup>c</sup>	Eliminated at stage 3 <sup>d</sup>
9	<b>Accessibility to tobacco, Susceptibility to social influences</b>				
	45 How easy you think it is for people your age to buy tobacco products in a store	4	No → C	Yes	
	46 Hours spent watching TV on a typical day	5	Yes		
	47 How often do you use the internet	7	Yes		
	48 Has a Facebook, Google Plus, MySpace, Twitter or other social networking	2	No → C	No → F	No → P
10	<b>Family-related variables</b>				
	49 Parent has a spouse or partner that lives in household	2	Yes		
	50 Cigarettes or tobacco might be available to youth at parent or guardian's home	2	Yes		
	51 Anyone who lives with you now uses tobacco	3	No → C	No → F	No → P
	52 Number of hours in past 7 days you were in close contact with others when they were smoking	C	Yes		
	53 Parent or guardian marital status	3	Yes		
	54 Youth has a curfew or set time to be home on school nights	2	Yes		
	55 Youth has a curfew or set time to be home on weekend nights	2	Yes		
	56 Rules about using tobacco products that are burned inside home	3	Yes		
	57 Statement that best describes the rules about using combustible tobacco	3	No → C	Yes	
	58 Parent or guardian talk with you, even once about not using any type of tobacco	2	Yes		
	59 Reaction of parent/guardian found you using tobacco	3	No → C	No → F	No → P
11	<b>Pocket money allowance, Movie influence</b>				
	60 Money received in total during an average week	9	No → A	No → F	Yes
	61 Number of times seen Movie 1	4	Yes		
	62 Number of times seen Movie 2	4	No → A	No → F	Yes
	63 Number of times seen Movie 3	4	No → A	Yes	
	64 Number of times seen Movie 4	4	No → A	No → F	Yes
	65 Number of times seen Movie 5	4	Yes		

<sup>a</sup>For graded variables, the number of levels is shown in the table, and the levels they represent is shown below. Continuous variables are indicated by the letter C. Exceptionally variables 10–13 were treated as a continuous variable with levels 1 to 4.

<sup>b</sup>Those not eliminated at stage 1 went into stage 2 analyses A, B or C as indicated.

<sup>c</sup>Those variables not eliminated at stage 2 went into stage 3 (final = F) analysis.

<sup>d</sup>Those variables not eliminated at stage 3 were included in the propensity score (P).



Grading systems used:	Variable(s)
12–14; 15–17	1
Male; Female	2
Hispanic; Non-Hispanic	3
White alone; Black alone, Other	4
North east; Mid west; South; West	5
Less than high school; Higher school graduate or equivalent; Some college (no degree) or associate degree; Bachelor's degree, Advance degree	6
Past month, 2–12 months; Over a year; Never	7–9, 14–20, 44
Very curious; Somewhat curious; A little curious; Not at all	10
Definitely yes; Probably yes; Probably not; Definitely not	11–13
Strongly agree; Agree; Disagree; Strongly agree	21–26
No harm; Little harm; Some harm; A lot or harm	27–28
It will never happen; Less than a year, 1 year; 5 years; 10 years; 20 years or above	29
Strongly agree; Agree; Neither agree or disagree; Disagree; Strongly disagree	30, 38–40
Yes; No	31–37, 48–50, 54, 55, 58
Mostly A's; A's or B's; Mostly B's; B's or C's; Mostly C's; C's or D's; Mostly D's; D's or F's; Mostly F's; School is ungraded	4
Never; Rarely; Sometimes; Often; Very often	42
Excellent; Very good; Good; Fair; Poor	43
Very easy; Somewhat easy; Somewhat difficult; Very difficult	45
None; less than 1 hour; 1 or 2 hours; 3 to 4 hours; More than 4 hours	46
Several times a day; About once a day; 3–5 times a week; 1–2 days a week; Every few weeks; Less often; Don't have regular internet access	47
Cigarettes; Cigars; Cigarillos or filtered cigars; Smokeless or other tobacco user; No one living in the home uses tobacco	51
Married; Widowed; Divorced or separated; Never married	53
It is not allowed anywhere or at any time inside my home; It is allowed in some places or at some times inside my home; It is allowed anywhere and at any time inside my home	56, 57
Be very upset; Not be too upset; Have no reaction	59
None; Less than \$1; \$1 to \$5; \$6 to \$10; \$11 to \$20; \$21 to \$50; \$51 to \$100; \$101-\$150; \$151 or more	60
Never, Once, Twice; 3 or more times	61–65

(curiosity about smoking; think will smoke a cigarette; prefer exciting and unpredictable friends) a corresponding question was not asked at Wave 2, so the Wave 1 value was used.

As the subjects included in this analysis were as for the main analysis, the unadjusted OR remained 5.702 (95% CI 4.334–7.502). After adjusting for propensity score as quintiles, the OR value reduced to 2.390 (1.791–3.188), somewhat lower than the 2.476 (1.852–3.310) in the main analysis, with no misclassification adjustment. After adjusting as a continuous variable, the OR reduced to 2.262 (1.625–3.150), again somewhat lower than the 2.474 (1.791–3.519) in the main analysis. Adjusting for the

12 variables making up the score reduced the OR to 1.772 (1.264–2.484), somewhat lower than the 1.847 (1.347–2.533) in the main analysis.

## Discussion

We have described analyses aimed at gaining further insight into the magnitude of any true “gateway effect” by attempting to control better for confounding factors linked to initiation of smoking. We used a propensity score approach, which is intended to remove confounding in the analysis of outcomes by balancing exposure groups on potential confounders, the score being developed prior to, so independently of, the analysis of

**Table 2. Final model relating 12 predictor variables to ever e-cigarette use at Wave 1.** The table shows the final model used for the propensity score. It is based on those 12 of the 18 predictor variables which entered stage 3 and is based on data from those 10,671 youth with data on all the 12 predictors.

Variable <sup>a</sup>	Levels	N	OR (95% CI)
Age range	12–14	5738	1.000 (base)
	15–17	4933	1.949 (1.557-2.440)
Ever used alcohol at all	Yes	3441	1.000 (base)
	No	7230	0.401 (0.319-0.504)
Ever been curious about smoking a cigarette			0.639 (0.556-0.734) <sup>b</sup>
Think you will smoke a cigarette in the next year			0.486 (0.393-0.600) <sup>c</sup>
Agree/disagree: Prefer friends who are exciting and unpredictable	Strongly agree	716	1.000 (base)
	Agree	2578	0.838 (0.602-1.168)
	Neither agree nor disagree	3818	0.604 (0.432-0.846)
	Disagree	2062	0.345 (0.220-0.542)
Reaction if parent/guardian found you using tobacco	Strongly disagree	1497	0.708 (0.439-1.143)
	Be very upset	10255	1.000 (base)
	Not be too upset	314	1.986 (0.383-2.853)
	Have no reaction	102	1.717 (0.798-3.693)
Gender	Male	5395	1.000 (base)
	Female	5216	0.660 (0.535-0.813)
Agree/disagree; I think I would enjoy using tobacco	Strongly agree	25	1.000 (base)
	Agree	164	6.720 (1.317-34.28)
	Disagree	2136	5.002 (1.011-24.74)
Agree/disagree: some products are safer than others	Strongly disagree	8346	3.798 (0.760-18.98)
	Strongly agree	332	1.000 (base)
	Agree	2621	0.514 (0.358-0.738)
Ever used prescription drug not prescribed to you: Ritalin or Adderall	Neither agree nor disagree	2192	0.515 (0.347-0.764)
	Disagree	2317	0.348 (0.225-0.538)
	Strongly disagree	3209	0.354 (0.230-0.544)
Ever used prescription drug not prescribed to you: Ritalin or Adderall	Yes	132	1.000 (base)
	No	10539	0.323 (0.198-0.528)
Has a Facebook, Google Plus, MySpace, Twitter or other social networking	Yes	8909	1.000 (base)
	No	1762	0.468 (0.299-0.733)
Anyone who lives with you now use Tobacco	Cigarettes, cigars,	2929	1.000 (base)
	cigarillos, filtered cigars		
	Smokeless or other tobacco	456	1.389 (0.924-2.088)
	No one living in the home uses tobacco	7286	0.741 (0.594-0.925)

<sup>a</sup>The variables are shown in order of their inclusion into the model

<sup>b</sup>Odds ratio are per unit of the graded variable which represents decreasing curiosity

<sup>c</sup>Odds ratio are per unit of the graded variable which represents decreasing likelihood

**Table 3. Predicting Wave 2 ever smoking from Wave 1 ever e-cigarette use – effects of confounder adjustment.** Table 3 first shows the unadjusted OR, and then the OR adjusted for the propensity score in two ways. Finally, it shows the results of a stepwise regression based on successively including the most significant adjustment variables.

Adjustment variables	OR (95% CI)
None	5.702 (4.334-7.502)
Propensity score as quintiles	2.476 (1.852-3.310)
Propensity score as continuous variable	2.474 (1.791-3.419)
Age range	4.806 (3.637-6.351)
+ Ever used alcohol at all	3.799 (2.855-5.055)
+ Ever been curious about smoking a cigarette	2.852 (2.123-3.831)
+ Think you will smoke a cigarette in the next year	2.185 (1.608-2.969)
+ Agree/disagree: Prefer friends who are exciting and unpredictable	2.111 (1.552-2.869)
+ Reaction if parent/guardian found you using tobacco	2.025 (1.489-2.756)
+ Gender	2.028 (1.489-2.761)
+ Agree/disagree; I think I would enjoy using tobacco	1.939 (1.420-2.648)
+ Agree/disagree: some products are safer than others	1.925 (1.406-2.635)
+ Ever used prescription drug not prescribed to you: Ritalin or Adderall	1.865 (1.360-2.557)
+ Has a Facebook, Google Plus, MySpace, Twitter or other social networking	1.852 (1.350-2.539)
+ Anyone who lives with you now use tobacco	1.847 (1.347-2.533)

outcomes. This approach attempts to transpose observational data into what would have been obtained from a randomized trial, using groups balanced on baseline covariates. The main analysis, which aims to balance potential confounders across vapers and non-vapers at Wave 1 (the comparison groups in the analysis where cigarette smoking is the outcome) is strictly designed for this approach. An alternative approach addressed using sensitivity analysis 5, views the propensity outcome more broadly, considering use of any nicotine-containing product as indicative of an inclination to initiate cigarette smoking. The difficulty in the propensity score approach, as with use of observational data generally, is to ensure that all relevant variables are considered in the score, and to account for possible inaccuracies in the variables included.

The main and five sensitivity analyses summarized in Table 4 all show that adjustment for propensity determined at Wave 1 markedly weakens the gateway effect, the association between vaping by Wave 1 and subsequent initiation of smoking. This was true whether propensity was based on variables associated with vaping (the main analyses), cigarette smoking (sensitivity analysis 4) or any tobacco product use (sensitivity analysis 5). Sensitivity analyses 1–3 also demonstrated this marked reduction, whether users of other products at Wave 1 were excluded or included, whether or not adjustment was made for such use, or whether analyses were based on ever or current use. There was no consistent difference between results adjusted for propensity

as quintiles or as a continuous variable, but adjustment for the individual variables making up the score produced lower adjusted ORs. The proportion of the unadjusted excess OR (i.e.  $OR - 1$ ) explained by adjustment for the individual variables was at least 75.4% in the main and sensitivity analyses, with a maximum of 87.5% for sensitivity analysis 2, where other product use was adjusted for, as well as the 12 main analysis predictor variables.

Our analyses have limitations. One is the small numbers of new smokers considered, never exceeding 71 and as low as six in sensitivity analysis 3. We are currently conducting additional work to try to obtain more precise answers by also using Wave 3 data.

A second issue is the possibility of over-adjustment. One can argue that vaping by Wave 1 may have affected some answers given then. For example, taking up e-cigarettes may make youths more curious about cigarettes, or more likely to think they will smoke or enjoy them. Using Wave 3 data, we are also conducting further analyses relating initiation of cigarette smoking at Wave 3 to vaping at Wave 2, restricting attention to those who, at Wave 1, had never vaped, and using propensity indicators recorded at Wave 1.

It is possible that more complete propensity adjustment might have explained more of the “gateway effect”. Our analyses

**Table 4. Comparison of ORs (95% CIs) from main and sensitivity analysis.** For the main analysis Table 4 shows the number of new smokers considered, as well as the OR (95% CI) for the relationship between ever e-cigarette use at Wave 1 and ever smoking initiation by Wave 2 with no adjustment for propensity to smoke, or with adjustment in three ways. Similar results are also shown for the five sensitivity analyses.

Analysis	Number of new smokers	Adjustment for propensity			
		None	Quintiles	Continuous	As variables
Main – relating ever e-cigarette use at Wave 1 to ever smoking initiation by Wave 2	421 (71 in e-cig users)	5.702 (4.334-7.502)	2.476 (1.852-3.310)	2.474 (1.791-3.519)	1.847 (1.347-2.533) <sup>a</sup>
Sensitivity 1 – as main but exclude ever smokers of other products at Wave 1	333 (37 in e-cig users)	5.128 (3.570-7.366)	2.298 (1.577-3.349)	2.297 (1.496-3.529)	1.659 (1.098-2.508) <sup>b,c</sup>
Sensitivity 2 – as main but include variable for other product use at Wave 1	421 (71 in e-cig users)	5.702 (4.334-7.502)	2.324 (1.737-3.110)	2.086 (1.487-2.925)	1.586 (1.144-2.198) <sup>d</sup>
Sensitivity 3 – as main but linking current e-cigarettes at Wave 1 to current smoking at Wave 2	149 (6 in e-cig users)	6.320 (2.915-13.700)	2.860 (1.292-6.332)	2.682 (1.053-6.834)	1.741 (0.743-4.080) <sup>e</sup>
Sensitivity 4 – as main but based on a propensity score for ever cigarette smoking	440 (71 in e-cig users)	5.442 (4.140-7.153)	2.286 (1.700-3.073)	2.569 (1.883-3.506)	2.096 (1.532-2.868) <sup>c,f</sup>
Sensitivity 5 – as main but based on propensity score for ever any tobacco product	413 (69 in e-cig users)	5.661 (4.289-7.472)	2.430 (1.813-3.257)	2.217 (1.608-3.057)	1.822 (1.321-2.513) <sup>c,g</sup>

<sup>a</sup>12 variables shown in Table 2

<sup>b</sup>10 variables (variables 1, 10, 11, 23, 32, 44, 48, 57, 59 and 62 from Table 1)

<sup>c</sup>Estimated using SAS version 9.4 as failed to converge using ROELEEE

<sup>d</sup>12 variables shown in Table 2 plus ever smoked other products

<sup>e</sup>7 variables (variables 1, 10, 14, 19, 30 and 33 from Table 1 plus ever smoked other products)

<sup>f</sup>18 variables (variables 1, 6, 17, 21, 23, 24, 32, 33, 41, 42, 45, 48, 50, 51, 52, 59 and 62 from Table 1 plus ever smoked other products)

<sup>g</sup>16 variables (variables 1, 10, 11, 21, 24, 30, 32, 39, 41, 44, 48, 51, 59, 60, 62 and 64 from Table 1)

only included variables showing an effect significant at  $p < 0.01$ . Weakening significance to  $p < 0.05$  or  $p < 0.1$  might have included more variables in the propensity score and explained even more of the association.

As is well documented, inaccurately determining confounding variables may limit the ability to fully adjust. Thus, many years ago (Tzonou *et al.*, 1986), it was demonstrated that “even misclassification rates as low as 10% can prevent adequate control of confounding” and other publications highlight the residual confounding problem (Ahlbom & Steineck, 1992; Fewell *et al.*, 2007; Greenland, 1980; Phillips & Davey Smith, 1994; Savitz & Baron, 1989). Proper adjustment for residual confounding requires a gold standard to compare reported answers with, but such data were unavailable in the PATH study. However, some insight was given by using predictor variables derived from answers given at both Wave 1 and Wave 2. Thus, if a youth reported alcohol use at one wave and not the other, it is possible this was not reported at one wave, and a predictor based on ever reported use may better predict smoking initiation. Such analyses usually weakened the adjusted association of prior vaping with subsequent smoking initiation, but only slightly. This may, however, reflect methodological limitations rather than lack of serious residual confounding.

While our analyses make it clear that most of the observed relationship between vaping and subsequent initiation of smoking results from confounding, the significant association seen even after extensive adjustment for confounders does seem to be consistent with there being some true gateway effect. However concerns about incomplete adjustment for confounding remain and our results do not unequivocally demonstrate that any true effect exists. More reliable information emerging from the further analyses we are currently conducting using Wave 3, should provide a better insight into the magnitude of any true gateway effect.

Another gateway analysis based on PATH Waves 1 and 2 has recently been published (Watkins *et al.*, 2018). This differed from ours in that their “unadjusted” model already included all non-cigarette tobacco products as indicators of cigarette initiation, and their “adjusted” models added only a restricted list of predefined variables, rather than using models to include more relevant variables. While their adjusted models did include some established determinants of cigarette use (sensation seeking; alcohol use; living with a tobacco user; and variables regarding health warnings and advertising), their adjusted ORs were higher than ours. Thus, whereas the variables we included in our main analysis reduced the OR from 5.702 to 1.847, their similar analysis reduced it only from 3.50 to 2.53. Consequently, although they recognized uncontrolled confounding may exist, they dubiously considered vaping was “independently associated with cigarette smoking one year later”.

In considering whether a true important gateway effect exists, one should note the lack of any increase in the US in cigarette

smoking prevalence following the rise in vaping (Levy *et al.*, 2019), and the fact that, in the PATH study, considerably more (279 vs. 79) Wave 1 cigarette only smokers took up e-cigarettes by Wave 2, than Wave 1 e-cigarette only users who took up smoking. Despite any possible gateway effect, introducing e-cigarettes may have reduced overall youth smoking prevalence.

## Conclusions

The results presented, based on Waves 1 and 2, strongly suggest that reported estimates of the gateway effect (Soneji *et al.*, 2017) are much too high. Indeed, it is not completely clear whether vaping actually increases subsequent uptake of cigarette smoking if potential confounding effects were to be fully accounted for.

## Addendum

At the time this paper was being finalised, an analysis was published (Berry *et al.*, 2019a) investigating gateway effects based on data from Waves 1, 2 and 3 of the PATH study. The authors reported that prior e-cigarette use was associated with increases in the odds of ever and current cigarette use by, respectively, 4.09 (95% CI 2.97-5.63) and 2.75 (1.60-4.73). Though noting that they could not rule out the possibility of residual confounding, they concluded that their findings supported a gateway effect. We will examine this claim in detail based on the results of our ongoing analyses using the data from all three Waves.

## Data availability

### Underlying data

National Addiction & HIV Data Archive Program: Population Assessment of Tobacco and Health (PATH) Study [United States] Public-Use Files (ICPSR 36498). <https://doi.org/10.3886/ICPSR36498.v8> (United States Department of Health and Human Services, 2018).

Data are available under the Terms of Use as set out by ICPSR, which can be accessed when users start the process of downloading the data.

### Extended data

Open Science Framework: Investigating gateway effects using the PATH study <https://doi.org/10.17605/OSF.IO/QCFZR> (Lee, 2019).

This project contains the following extended data files:

- Gateway PATH\_F1000 Research\_Supplementary file.docx

Data are available under the terms of the [Creative Commons Zero “No rights reserved” data waiver](#) (CC0 1.0 Public domain dedication).

## Acknowledgements

We thank Mrs Y Cooper and Mrs D Morris for typing the various drafts of this report, and Philip Morris for financial support.

## References

- Ahlbom A, Steineck G: **Aspects of misclassification of confounding factors.** *Am J Ind Med.* 1992; **21**(1): 107–112.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Aleyan S, Cole A, Qian W, *et al.*: **Risky business: a longitudinal study examining cigarette smoking initiation among susceptible and non-susceptible e-cigarette users in Canada.** *BMJ Open.* 2018; **8**(5): e021080.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Barrington-Trimis JL, Urman R, Berhane K, *et al.*: **E-Cigarettes and Future Cigarette Use.** *Pediatrics.* 2016; **138**(1): pii: e20160379.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Berry KM, Fetterman JL, Benjamin EJ, *et al.*: **Association of Electronic Cigarette Use With Subsequent Initiation of Tobacco Cigarettes in US Youths.** *JAMA Netw Open.* 2019a; **2**(2): e187794.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Berry KM, Reynolds LM, Collins JM, *et al.*: **E-cigarette initiation and associated changes in smoking cessation and reduction: the Population Assessment of Tobacco and Health Study, 2013-2015.** *Tob Control.* 2019b; **28**(1): 42–49.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Best C, Haseen F, Currie D, *et al.*: **Relationship between trying an electronic cigarette and subsequent cigarette experimentation in Scottish adolescents: a cohort study.** *Tob Control.* 2018; **27**(4): 373–378.  
[Publisher Full Text](#)
- Conner M, Grogan S, Simms-Ellis R, *et al.*: **Do electronic cigarettes increase cigarette smoking in UK adolescents? Evidence from a 12-month prospective study.** *Tob Control.* 2018; **27**(4): 365–372.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Cooper M, Case KR, Loukas A, *et al.*: **E-cigarette Dual Users, Exclusive Users and Perceptions of Tobacco Products.** *Am J Health Behav.* 2016; **40**(1): 108–116.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Dautzenberg B, de Souza Moura MA, Rieu N, *et al.*: **L'e-cigarette bouleverse les autres consommations des adolescents parisiens (2012 à 2014) (The e-cigarette disrupts other consumptions in Parisian teenagers (2012-2014)).** *Rev Mal Respir.* 2016; **33**(3): 225–34.  
[Publisher Full Text](#)
- Dutra LM, Glantz SA: **Electronic cigarettes and conventional cigarette use among U.S. adolescents: a cross-sectional study.** *JAMA Pediatr.* 2014; **168**(7): 610–7.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Eastwood B, Dockrell MJ, Arnott D, *et al.*: **Electronic cigarette use in young people in Great Britain 2013-2014.** *Public Health.* 2015; **129**(9): 1150–6.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Etter JF: **Gateway effects and electronic cigarettes.** *Addiction.* 2018; **113**(10): 1776–1783.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Fewell Z, Davey Smith G, Sterne JA: **The impact of residual and unmeasured confounding in epidemiologic studies: a simulation study.** *Am J Epidemiol.* 2007; **166**(6): 646–55.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Goniewicz ML, Gawron M, Nadolska J, *et al.*: **Rise in electronic cigarette use among adolescents in Poland.** *J Adolesc Health.* 2014; **55**(5): 713–5.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Greenland S: **The effect of misclassification in the presence of covariates.** *Am J Epidemiol.* 1980; **112**(4): 564–9.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Hammond D, Reid JL, Cole AG, *et al.*: **Electronic cigarette use and smoking initiation among youth: a longitudinal cohort study.** *CMAJ.* 2017; **189**(43): E1328–e1336.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hornik RC, Gibson L, Lerman C: **Prediction of cigarette use from six month prior electronic and combustible cigarette use for a U.S. national sample of 13-25 year olds [abstract POS5-30].** Society for Research on Nicotine & Tobacco. 2016 Rapid Response Posters. 2016.  
[Reference Source](#)
- Hyland A, Ambrose BK, Conway KP, *et al.*: **Design and methods of the Population Assessment of Tobacco and Health (PATH) Study.** *Tob Control.* 2017; **26**(4): 371–378.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Lee P: **Investigating gateway effects using the PATH study.** *OSF.* 2019.  
<http://www.doi.org/10.17605/OSF.IO/QCFZR>
- Lee PN, Coombs KJ, Afolalu EF: **Considerations related to vaping as a possible gateway into cigarette smoking: an analytical review [version 3; peer review: 2 approved].** *F1000Res.* 2018; **7**: 1915.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Lee S, Grana RA, Glantz SA: **Electronic cigarette use among Korean adolescents: a cross-sectional study of market penetration, dual use, and relationship to quit attempts and former smoking.** *J Adolesc Health.* 2014; **54**(6): 684–90.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Leventhal AM, Strong DR, Kirkpatrick MG, *et al.*: **Association of Electronic Cigarette Use With Initiation of Combustible Tobacco Product Smoking in Early Adolescence.** *JAMA.* 2015; **314**(7): 700–707.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Levy DT, Warner KE, Cummings KM, *et al.*: **Examining the relationship of vaping to smoking initiation among US youth and young adults: a reality check.** *Tob Control.* 2019; **28**(6): 629–635.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Loukas A, Marti CN, Cooper M, *et al.*: **Exclusive e-cigarette use predicts cigarette initiation among college students.** *Addict Behav.* 2018; **76**: 343–347.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Lozano P, Barrientos-Gutierrez I, Arillo-Santillan E, *et al.*: **A longitudinal study of electronic cigarette use and onset of conventional cigarette smoking and marijuana use among Mexican adolescents.** *Drug Alcohol Depend.* 2017; **180**: 427–430.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Miech R, Patrick ME, O'Malley PM, *et al.*: **E-cigarette use as a predictor of cigarette smoking: results from a 1-year follow-up of a national sample of 12th grade students.** *Tob Control.* 2017; **26**(e2): e106–e111.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- National Academies of Sciences Engineering and Medicine: **Public health consequences of e-cigarettes.** The National Academies Press, Washington DC. 2018.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Nutt DJ, Phillips LD, Balfour D, *et al.*: **Estimating the harms of nicotine-containing products using the MCDA approach.** *Eur Addict Res.* 2014; **20**(5): 218–225.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Phillips AN, Davey Smith G: **Cigarette smoking as a potential cause of cervical cancer: has confounding been controlled?** *Int J Epidemiol.* 1994; **23**(1): 42–49.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Primack BA, Freedman-Doan P, Sidani JE, *et al.*: **Sustained Waterpipe Tobacco Smoking and Trends Over Time.** *Am J Prev Med.* 2015; **49**(6): 859–67.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Primack BA, Shensa A, Sidani JE, *et al.*: **Initiation of cigarette smoking after e-cigarette use: a nationally representative study.** *Ann Behav Med.* 2016; **50** (Suppl 1): S68.
- Savitz DA, Barón AE: **Estimating and correcting for confounder misclassification.** *Am J Epidemiol.* 1989; **129**(5): 1062–1071. Erratum appears in *American Journal of Epidemiology* 1989; **130**: 1260.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Soneji S, Barrington-Trimis JL, Wills TA, *et al.*: **Association Between Initial Use of e-Cigarettes and Subsequent Cigarette Smoking Among Adolescents and Young Adults: A Systematic Review and Meta-analysis.** *JAMA Pediatr.* 2017; **171**(8): 788–797.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Spindle TR, Hiler MM, Cooke ME, *et al.*: **Electronic cigarette use and uptake of cigarette smoking: A longitudinal examination of U.S. college students.** *Addict Behav.* 2017; **67**: 66–72.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Tzonou A, Kaldor J, Smith PG, *et al.*: **Misclassification in case-control studies with two dichotomous risk factors.** *Rev Epidemiol Sante Publique.* 1986; **34**(1): 10–17.  
[PubMed Abstract](#)
- Unger JB, Soto DW, Leventhal A: **E-cigarette use and subsequent cigarette and marijuana use among Hispanic young adults.** *Drug Alcohol Depend.* 2016; **163**: 261–4.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- United States Department of Health and Human Services. National Institutes of Health. National Institute on Drug Abuse, United States Department of Health and Human Services. Food and Drug Administration. Center for Tobacco Products: **Population Assessment of Tobacco and Health (PATH) Study [United States] Public-Use Files.** Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2018.  
[Publisher Full Text](#)
- Watkins SL, Glantz SA, Chaffee BW: **Association of Noncigarette Tobacco Product Use With Future Cigarette Smoking Among Youth in the Population Assessment of Tobacco and Health (PATH) Study, 2013-2015.** *JAMA Pediatr.* 2018; **172**(2): 181–7.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Wills TA, Knight R, Sargent JD, *et al.*: **Longitudinal study of e-cigarette use and onset of cigarette smoking among high school students in Hawaii.** *Tob Control.* 2017; **26**(1): 34–39.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

# Open Peer Review

Current Peer Review Status:  

---

## Version 2

Reviewer Report 08 January 2020

<https://doi.org/10.5256/f1000research.23710.r57453>

© 2020 Hanewinkel R. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



### Reiner Hanewinkel

Institute for Therapy and Health Research (IFT-Nord), Kiel, Germany

I do not fully agree with the arguments raised by the authors of the paper, but that's normal in science.

The data are presented in a way, that every reader can make up his own mind about the conclusions made by the authors. For that reason, I approve the paper.

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** I am trained as Psychologist.

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

---

## Version 1

Reviewer Report 08 November 2019

<https://doi.org/10.5256/f1000research.20078.r55826>

© 2019 Hanewinkel R. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



### Reiner Hanewinkel

Institute for Therapy and Health Research (IFT-Nord), Kiel, Germany

“On 19 December 2018, Altria announced an offer of US\$12.8billion for a 35% share of Juul Labs. The deal combines Altria (formerly Philip Morris), the company with the largest share of the US cigarette

market, with the large and rapidly growing vaping product (aka e-cigarette) company, Juul Labs. The acquisition price was based on a US\$38billion valuation, which was more than twice Juul Labs' August valuation, and a surprise to many investors on Wall Street.”<sup>1</sup>

We have to have this information in mind, because the manuscript submitted to F1000Research by Peter Lee and John Fry is financed by Philip Morris. My personal opinion is that Philip Morris has an interest to show that the so called gateway effect is not existent.

I am not an expert in the propensity score approach, but what I understand is, that the authors tried everything to show, that the gateway effect does not exist. It seems trivial to me, that adjustments reduce an OR. Even after adjustment for a broad range of confounders, there is still a significant and clinical meaningful association between e-cigarette use at baseline and smoking initiation in the observational period.

My main point is that I do not agree with the conclusion and discussion of the results. Instead of saying that confounding is a major factor explaining most of the observed gateway effect, the authors should emphasize that even after controlling for a large number of confounders, the gateway effect is still present.

I am counting nearly two dozens of cohort studies, which investigated the gateway effect. The authors only cite a portion of these studies. In detail, young people have been recruited in the US and Canada<sup>2-15</sup>, UK<sup>16-18</sup>, Finland<sup>19</sup>, Mexico<sup>20</sup>, the Netherlands<sup>21</sup>, Romania<sup>22</sup>, Taiwan<sup>23</sup>, and Germany<sup>24</sup>. More than 100,000 young people have been recruited for these studies with observational periods of up to 24 months. Outcome variables range from one cigarette to daily smoking. All these studies found a significant association between e-cigarette use at baseline and uptake of smoking in the observational period. Applying the well-known Bradford-Hill criteria, it seems to me that the following criteria are fulfilled: (1) Strength (effect size), (2) Consistency (reproducibility), (3) Temporality, and (4) Plausibility.

In the Introduction the authors should describe the two opponent explanations for the observed association between e-cigarette use and the later uptake of smoking: the so called “common liability” theory vs. the “gateway” theory<sup>25</sup>. Another important theoretical paper should also be discussed<sup>26</sup>.

If anything, the results reported are in line with the gateway theory and not with the common liability theory.

## References

1. Levy DT, Sweanor D, Sanchez-Romero LM, O'Connor R, Goniewicz ML, Borland R: Altria-Juul Labs deal: why did it occur and what does it mean for the US nicotine delivery product market. *Tob Control*. 2019. [PubMed Abstract](#) | [Publisher Full Text](#)
2. Barrington-Trimis JL, Urman R, Berhane K, Unger JB, Cruz TB, Pentz MA, Samet JM, Leventhal AM, McConnell R: E-Cigarettes and Future Cigarette Use. *Pediatrics*. 2018; **138** (1). [PubMed Abstract](#) | [Publisher Full Text](#)
3. Leventhal AM, Strong DR, Kirkpatrick MG, Unger JB, Sussman S, Riggs NR, Stone MD, Khoddam R, Samet JM, Audrain-McGovern J: Association of Electronic Cigarette Use With Initiation of Combustible Tobacco Product Smoking in Early Adolescence. *JAMA*. 2015; **314** (7): 700-7 [PubMed Abstract](#) | [Publisher Full Text](#)
4. Loukas A, Marti CN, Cooper M, Pasch KE, Perry CL: Exclusive e-cigarette use predicts cigarette initiation among college students. *Addict Behav*. 2018; **76**: 343-347 [PubMed Abstract](#) | [Publisher Full Text](#)
5. Miech R, Patrick ME, O'Malley PM, Johnston LD: E-cigarette use as a predictor of cigarette smoking:



- results from a 1-year follow-up of a national sample of 12th grade students. *Tob Control*. **26** (e2): e106-e111 [PubMed Abstract](#) | [Publisher Full Text](#)
6. Spindle TR, Hiler MM, Cooke ME, Eissenberg T, Kendler KS, Dick DM: Electronic cigarette use and uptake of cigarette smoking: A longitudinal examination of U.S. college students. *Addict Behav*. **67**: 66-72 [PubMed Abstract](#) | [Publisher Full Text](#)
7. Unger JB, Soto DW, Leventhal A: E-cigarette use and subsequent cigarette and marijuana use among Hispanic young adults. *Drug Alcohol Depend*. 2016; **163**: 261-4 [PubMed Abstract](#) | [Publisher Full Text](#)
8. Wills TA, Knight R, Sargent JD, Gibbons FX, Pagano I, Williams RJ: Longitudinal study of e-cigarette use and onset of cigarette smoking among high school students in Hawaii. *Tob Control*. **26** (1): 34-39 [PubMed Abstract](#) | [Publisher Full Text](#)
9. Leventhal A, Stone M, Andrabi N, Barrington-Trimis J, Strong D, Sussman S, Audrain-McGovern J: Association of e-Cigarette Vaping and Progression to Heavier Patterns of Cigarette Smoking. *JAMA*. 2016; **316** (18). [Publisher Full Text](#)
10. Hammond D, Reid JL, Cole AG, Leatherdale ST: Electronic cigarette use and smoking initiation among youth: a longitudinal cohort study. *CMAJ*. 2017; **189** (43): E1328-E1336 [PubMed Abstract](#) | [Publisher Full Text](#)
11. Bold KW, Kong G, Camenga DR, Simon P, Cavallo DA, Morean ME, Krishnan-Sarin S: Trajectories of E-Cigarette and Conventional Cigarette Use Among Youth. *Pediatrics*. **141** (1). [PubMed Abstract](#) | [Publisher Full Text](#)
12. Primack BA, Shensa A, Sidani JE, Hoffman BL, Soneji S, Sargent JD, Hoffman RM, Fine MJ: Initiation of Traditional Cigarette Smoking after Electronic Cigarette Use Among Tobacco-Naïve US Young Adults. *Am J Med*. **131** (4): 443.e1-443.e9 [PubMed Abstract](#) | [Publisher Full Text](#)
13. Primack BA, Soneji S, Stoolmiller M, Fine MJ, Sargent JD: Progression to Traditional Cigarette Smoking After Electronic Cigarette Use Among US Adolescents and Young Adults. *JAMA Pediatr*. 2015; **169** (11): 1018-23 [PubMed Abstract](#) | [Publisher Full Text](#)
14. Watkins S, Glantz S, Chaffee B: Association of Noncigarette Tobacco Product Use With Future Cigarette Smoking Among Youth in the Population Assessment of Tobacco and Health (PATH) Study, 2013-2015. *JAMA Pediatrics*. 2018; **172** (2). [Publisher Full Text](#)
15. McMillen R, Klein JD, Wilson K, Winickoff JP, Tanski S: E-Cigarette Use and Future Cigarette Initiation Among Never Smokers and Relapse Among Former Smokers in the PATH Study. *Public Health Rep*. **134** (5): 528-536 [PubMed Abstract](#) | [Publisher Full Text](#)
16. Best C, Haseen F, Currie D, Ozakinci G, MacKintosh AM, Stead M, Eadie D, MacGregor A, Pearce J, Amos A, Frank J, Haw S: Relationship between trying an electronic cigarette and subsequent cigarette experimentation in Scottish adolescents: a cohort study. *Tob Control*. 2017. [PubMed Abstract](#) | [Publisher Full Text](#)
17. Conner M, Grogan S, Simms-Ellis R, Flett K, Sykes-Muskett B, Cowap L, Lawton R, Armitage CJ, Meads D, Torgerson C, West R, Siddiqi K: Do electronic cigarettes increase cigarette smoking in UK adolescents? Evidence from a 12-month prospective study. *Tob Control*. 2017. [PubMed Abstract](#) | [Publisher Full Text](#)
18. East K, Hitchman SC, Bakolis I, Williams S, Cheeseman H, Arnott D, McNeill A: The Association Between Smoking and Electronic Cigarette Use in a Cohort of Young People. *J Adolesc Health*. **62** (5): 539-547 [PubMed Abstract](#) | [Publisher Full Text](#)
19. Kinnunen JM, Ollila H, Minkkinen J, Lindfors PL, Timberlake DS, Rimpelä AH: Nicotine matters in predicting subsequent smoking after e-cigarette experimentation: A longitudinal study among Finnish adolescents. *Drug Alcohol Depend*. 2019; **201**: 182-187 [PubMed Abstract](#) | [Publisher Full Text](#)
20. Lozano P, Barrientos-Gutierrez I, Arillo-Santillan E, Morello P, Mejia R, Sargent J, Thrasher J: A longitudinal study of electronic cigarette use and onset of conventional cigarette smoking and marijuana use among Mexican adolescents. *Drug and Alcohol Dependence*. 2017; **180**: 427-430 [Publisher Full Text](#)

21. Treur JL, Rozema AD, Mathijssen JJP, van Oers H, Vink JM: E-cigarette and waterpipe use in two adolescent cohorts: cross-sectional and longitudinal associations with conventional cigarette smoking. *Eur J Epidemiol.* **33** (3): 323-334 [PubMed Abstract](#) | [Publisher Full Text](#)
22. Pénczes M, Foley KL, Nădășan V, Paulik E, Ábrám Z, Urbán R: Bidirectional associations of e-cigarette, conventional cigarette and waterpipe experimentation among adolescents: A cross-lagged model. *Addict Behav.* **80**: 59-64 [PubMed Abstract](#) | [Publisher Full Text](#)
23. Chien Y, Gao W, Sanna M, Chen P, Chen Y, Glantz S, Chiou H: Electronic Cigarette Use and Smoking Initiation in Taiwan: Evidence from the First Prospective Study in Asia. *International Journal of Environmental Research and Public Health.* 2019; **16** (7). [Publisher Full Text](#)
24. Morgenstern M, Nies A, Goecke M, Hanewinkel R: E-Cigarettes and the Use of Conventional Cigarettes. *Deutsches Aerzteblatt Online.* 2018. [Publisher Full Text](#)
25. Etter JF: Gateway effects and electronic cigarettes. *Addiction.* **113** (10): 1776-1783 [PubMed Abstract](#) | [Publisher Full Text](#)
26. Schneider S, Diehl K: Vaping as a Catalyst for Smoking? An Initial Model on the Initiation of Electronic Cigarette Use and the Transition to Tobacco Smoking Among Adolescents. *Nicotine Tob Res.* 2016; **18** (5): 647-53 [PubMed Abstract](#) | [Publisher Full Text](#)

**Is the work clearly and accurately presented and does it cite the current literature?**

Partly

**Is the study design appropriate and is the work technically sound?**

Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**

Partly

**If applicable, is the statistical analysis and its interpretation appropriate?**

I cannot comment. A qualified statistician is required.

**Are all the source data underlying the results available to ensure full reproducibility?**

Yes

**Are the conclusions drawn adequately supported by the results?**

No

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** I am trained as Psychologist

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.**

Author Response 27 Nov 2019

**Peter Lee**, P.N.Lee Statistics and Computing Ltd, Sutton, UK

We thank Dr. Hanewinkel for his comments. However, it seems unfortunate that he suggests our interpretation may have been affected by the funding source. As experienced medical

statisticians/epidemiologists our interest is in unbiased assessment of data, not in trying to reach conclusions favourable to any sponsor. The objective of our paper was to provide insight into how much the observed gateway effect could be explained by adjustment for confounding factors, in order to provide insight into the possible magnitude of any true effect. That adjustment could reduce an unadjusted OR from 5.70 to 1.59 should be interpreted as providing doubt as to whether there is any real effect, despite the 1.59 being statistically significant. Our discussion, and also our companion paper in F1000Research<sup>[1]</sup>, already highlight the potential problems of interpretation due to residual confounding.

Dr. Hanewinkel states that “it seems trivial” to him “that adjustments reduce an OR”. However, theoretically adjustment can increase or decrease an OR. He cites numerous cohort studies that reported a significant gateway effect after adjustment, and we have cited all but the most recently published ones in our present paper or in its companion<sup>[1]</sup>. We note that many of these papers have also shown a large reduction in the OR after allowing for confounding variables. The existence of a significant association following adjustment does not necessarily mean that the adjusted association is real, for the reasons we have already discussed.

We have now extended the introduction as suggested to describe the “common liability” and the “gateway” theory, making it clear that both can contribute to the observed association between e-cigarette use and subsequent initiation of smoking.

Our abstract already made clear that a significant association remained after adjustment, and that its interpretation is not straightforward. We have now modified the conclusion in the abstract to point out that further work is needed to clarify the extent of any true gateway effect that may exist.

We have also extended the discussion section of the paper. We feel that our results adequately support our conclusion that “confounding is a major factor, explaining most [though not all] of the observed gateway effect”. Further analysis of later waves of the PATH study could well help to establish whether the residual effect is or is not real.

#### Reference

1 Lee PN, Coombs KJ and Afolalu EF. Considerations related to vaping as a possible gateway into cigarette smoking: an analytical review [version 3; peer review: 2 approved]. *F1000Research* 2018; 7: (<https://doi.org/10.12688/f1000research.16928.3>) [DOI: [doi.org/10.12688/f1000research.16928.3](https://doi.org/10.12688/f1000research.16928.3)]

**Competing Interests:** The authors are long term consultants to the tobacco industry.

Reviewer Report 12 June 2019

<https://doi.org/10.5256/f1000research.20078.r48433>

© 2019 Niaura R et al. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Raymond Niaura**

Department of Social and Behavioral Sciences, NYU College of Global Public Health, New York City, NY,

USA

**Shu Xu**

Department of Biostatistics, NYU College of Global Public Health, New York City, NY, USA

The authors aimed to evaluate the “gateway effect” of baseline e-cigarette use predicting cigarette smoking initiation among youths, addressing the inadequate adjustment of confounders considered in the previous published studies. The first two waves of PATH data from the youth sample were used for this purpose.

The strength of this study is that the authors used multiple approaches to evaluate this gateway effect, and performed five sets of sensitivity analysis. Propensity scores for e-cigarette use at wave 1 were created. Propensity score based approaches are preferred to statistically control for numerous confounding variables; however, this approach was not completely adequate as it appears in this study for several reasons. A few major concerns are below.

First, due to the availability of data at the time of analyses, the confounding variables were measured concurrently with e-cigarette exposure or afterwards (i.e., analyses investigating residual confounding). Considering the Wave 3 PATH data have been available since fall 2018, the authors are encouraged to revise the manuscript using data from the first 3 waves. They state they will do so but they could probably access the data at this point.

Second, the authors aimed to address the inadequate adjustment of confounding in previous studies, however, this limitation still exists in the current study. As a result of Step 1, only a set of 13 covariates were included of the main analysis, sensitivity analyses, and residual confounding analysis. Moreover, it is unclear whether the balancing of covariates was achieved between the (e-cigarette) exposure and non-exposure groups.

A few details regarding each approach need to be provided. Specifically, it was unclear what the authors meant by saying “the first analysis was a weighted logistic regression analysis.” How was a weighting variable created? Was this what the authors called “no adjustment of propensity” afterwards (in the first paragraph of the Sensitivity Analysis section, and in Table 4)? The results from the “adjustment as quintiles” and “adjustment as a continuous variable” approaches typically agree, and the effect of e-cigarette exposure would be biased if only a subset of confounding variables were considered in analyses. Moreover, the “using the variables making up the score individually” approach was actually a traditional covariate approach, which assumed a linear association between the outcome and each covariate. Authors need to justify whether this assumption was satisfied or not before they concluded the findings of the study.

Third, in analyses investigating residual confounding, authors used some of the covariates from Wave 2, which might be influenced by the exposure of “e-cigarette” at Wave 1. In this case, the partial coefficient of e-cigarette use on outcome is actually the controlled direct effect in a mediation analysis. This coefficient only showed the partial effect of e-cigarette use on outcome, which was an underestimated total effect of exposure. In this set of analysis, the confounding was over-adjusted.

Fourth, all analyses only involved participants with complete data for each variable considered in the group being analyzed. Missing complete at random was actually assumed in all analyses. Authors need to report the sample sizes of the original data and the working data, and justify whether/why this

assumption was valid in this study.

In addition, a few minor issues are below.

1. Methods: Some information in the last paragraph of the Introduction section need to be relocated in the Methods section, under the title of “sample and data collection”.
2. Software: what is the ROELEE program? Please briefly introduce this software. Under what circumstance did ROELEE fail to converge (and SAS did not)?
3. Supplementary document: the very long list of variables was not helpful to understand how confounding variables were selected. It seems like the authors have already selected (without details on criteria) 60 smoking predictor variables out of 200+. Please provide sample questions as needed.
4. Sensitivity analysis: report sample size of each sensitivity analysis in Table 4.

**Is the work clearly and accurately presented and does it cite the current literature?**

Yes

**Is the study design appropriate and is the work technically sound?**

Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**

Partly

**If applicable, is the statistical analysis and its interpretation appropriate?**

Partly

**Are all the source data underlying the results available to ensure full reproducibility?**

Yes

**Are the conclusions drawn adequately supported by the results?**

Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Tobacco regulatory science.

**We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Author Response 27 Jun 2019

**Peter Lee**, P.N.Lee Statistics and Computing Ltd, Sutton, UK

Response by the first author, Peter Lee.

In his first major point, Dr Niaura expressed concern about the possibility of confounding arising if vaping by Wave 1 had affected some answers given then. We noted this in our discussion saying that further analyses are planned relating initiation of cigarette smoking at Wave 3 to vaping at Wave 2, restricting attention to those who, at Wave 1, had never vaped, and using propensity indicators recorded at Wave 1. We have had access to the Wave 3 data for some time, and in due course will publish a further paper. We do not propose to revise the current manuscript.

As regards to the second concern, our final analyses were based on adjustment for a limited number of confounding variables. However, this list was based on a much longer list, with no further variables adding significantly to the final model, the much longer list (shown in Supplementary File 1) being itself derived from a review of the literature. The fact that the model enormously reduced the association between vaping and subsequent initiation of cigarette smoking was the main point we wished to make. It is possible other models may explain more of the unadjusted association but that will be addressed in the paper currently being prepared.

Dr. Niuara refers to reporting whether the balancing of covariates was achieved between the e-cigarette exposed and non-exposed groups. Presumably, he is referring to balancing after propensity adjustment. We intend to report on this in the planned new paper.

Dr. Niaura asked how the weighting variable was created. This variable is supplied directly to users of the PATH study and allows the user to generate results representative of the U.S. population.

It is stated that “the effect of e-cigarette exposure would be biased if only a subset of confounding variables were considered in analyses”. As noted above we did take into account a very large number of potential confounders in our analyses. It would be expected that adjustment for some variables would not in fact have a material effect when other variables were already included in the analyses, i.e. they would not be true confounders.

As regards linear associations it should be noted that most of the variables were yes/no variables. We will consider non-linear relationships in our planned analyses using Wave 3.

It is well known that adjustment for inaccurately measured confounding variables may be incomplete. In the absence of a gold standard we attempted to gain insight into the problem of looking at variation in answers to corresponding questions at different Waves. It is true that this may cause problems if the answers are affected by e-cigarette use, but this could only be addressed by analyses involving more than three Waves, relating e-cigarette use at Wave 3 to initiation of smoking by Wave 4, with adjustment for questions asked at Waves 1 and 2 in non-tobacco users.

We will report fuller details regarding missing data in our planned analyses based on Waves 1 to 3. As regards the other points we do not propose to modify the current paper, preferring to take all the points into account in the paper on Waves 1 to 3.

For information, the ROELEE program is one my company started to develop over 30 years ago, initially to get output in a user-friendly form not available using programs like SAS or SPSS. It has undergone numerous updates over the years and is fully tested. It is available from ROELEE Statistics Ltd. While, for logistic regression analyses, it gives the same results as SAS, but in a more convenient form, the version of ROELEE we used at the time sometimes failed to converge when there are large numbers of predictor variables, whereas SAS did not. We are currently modifying the code of ROELEE to avoid this non-convergence problem.

As regards the sample sizes used in the 6 sets of analyses in Table 4, these were as follows: Main, S2 and S5 9423; S1 8900; S3 9907; S4 9354.

**Competing Interests:** None additional to those disclosed in the paper itself.

---

---

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact [research@f1000.com](mailto:research@f1000.com)

**F1000Research**