

Gateway effects: why the cited evidence does not support their existence for low-risk tobacco products (and what evidence would)

Carl V Phillips

Consumer Advocates for Smoke-free Alternatives Association (CASAA)

9 March 2015

### **Abstract**

It is often claimed that low-risk drugs still create harm because of “gateway effects”, in which they cause the use of a high-risk alternative. Such claims are popular among opponents of tobacco harm reduction, claiming that low-risk tobacco products (e.g., e-cigarettes, smokeless tobacco) cause people to start smoking, sometimes backed by empirical studies that ostensibly support the claim. However, these studies consistently ignore the obvious alternative causal pathways, particularly that observed associations may represent causation in the opposite direction (smoking causes people to seek low-risk alternatives) or confounding (the same individual factors increase the chance of using any tobacco product). Due to these complications, any useful analysis must deal with simultaneity and confounding by common cause. In practice, existing analyses provide almost cartoon examples of drawing naïve causal conclusions from observed associations. The present analysis examines what evidence and research strategies would be needed to empirically detect such a gateway effect, if there were one, explaining key methodological concepts including causation and confounding, examining the logic of the claim, identifying potentially useful data, and debunking common fallacies on both side of the argument, as well as presenting an extended example of proper empirical testing. The analysis demonstrates that none of the empirical studies to date that purport to show a gateway effect from tobacco harm reduction products actually does so. The observations and approaches can be generalized to other cases where observed association of individual characteristics in cross-sectional data can result from one or several causal relationships.

### **Introduction**

An oft-claimed downside of harm reduction strategies is that the lower-risk alternative to the high-risk behavior is a “gateway” to the high-risk behavior. That is, the availability or promotion of the low-risk option causes some people who would not otherwise have done so to adopt the high risk behavior. Various studies have purported to show that such effects exist but – at least for the case of tobacco – none of the evidence actually supports the claim. Rather, they demonstrate a faulty understanding of the authors about what evidence would support the claim.

Most of the analysis presented here applies to any gateway claim, but the presentation focuses on tobacco harm reduction (THR) products, low-risk alternatives to smoking, including smokeless tobacco and e-cigarettes, whose substitution nearly eliminates the health risk from tobacco use. The analysis also generalizes to any situation where cross-sectional data is alleged to support a particular causal claim when competing causal

pathways appear to be much stronger causes of the association. In the THR case, there is clear confounding due to a common cause, which can be thought of as the generic unspecified common cause, typically designated “U” in the methods literature, or specifically in terms of a propensity toward liking the effects of tobacco/nicotine and being willing to resist the social pressures that demand abstinence. There is also clear causation in the opposite direction from the gateway: low-risk tobacco products, especially e-cigarettes, are almost exclusively used as a substitute for smoking, causing an association between their use and the use of cigarettes. Any methodology attempting to identify the effects of one causal pathway among three – and in the THR case, where the effect of interest, if it exists, undoubtedly contributes less to the association than the others – requires a great deal of thought about how to use data.

Gateway claims became common starting in the 1950s, where they were a U.S. Drug War tactic aimed at demonizing cannabis use. Cannabis use is not terribly harmful in itself in terms of biological or social effects, and there was even less evidence suggesting it was harmful at the time. So the claim was that it was harmful because it causes some users to “move on” to “hard drugs”. A similar tactic has been adopted by opponents of THR strategies, in which they claim that, even though any risks from smokeless tobacco and e-cigarettes are so low as to be undetectable, their use causes substantial harm because it causes some people to start smoking. It has even been claimed that e-cigarettes are a gateway to cocaine [1], though that claim is easily debunked [2].

Gateway claims can serve as a refuge for activists who profess concern about health or welfare, but really oppose a particular behavior for some other reason and cannot find any health- or welfare-based objection. The claims serve as a health- or welfare-based rationalization that is sufficiently ominous to influence the thinking of an uncritical target audience due to the common tendency to fixate on the consequences of a possible negative outcome, ignoring its probability and thus net expected value. The consistent lack of quantification in the claims (i.e., *how many* gateway cases have occurred or are predicted) tends to suggest that their only goal is to trigger this cognitive bias.

Opponents of THR invoke the gateway claim in discussions about eliminating the products entirely (which would stop any gateway effect), but also about regulating where the products can be used (which would have no apparent effect on any gateway). They invoke the same claims when the discussion is about whether to encourage the use of low-risk alternatives among current smokers (which could not possibly create a gateway). This suggests the gateway is, to them, merely a rhetorical tactic, not a genuine concern. Indeed, a leading anti-THR strategy is to claim that the low-risk products are much higher risk than they are, which, if believed, tends to *increase*, rather than decrease, any gateway effect because it tells anyone using those products that they might as well smoke [3]. It has recently been argued that propaganda campaigns against e-cigarettes have caused many vapers (users of e-cigarettes) to switch back to smoking [4,5]. These observations suggest that the stated worry about gateway effects is disingenuous.

But even if the gateway claim is primarily used as a rationalization for other goals, there is a legitimate scientific question about the extent to which it occurs. There are certainly

policy-makers and other observers who have been persuaded to believe there are substantial effects that should motivate policy. Thus, it is useful to analyze what empirical evidence is potentially useful in assessing whether such effects exist. Even if the conclusion is that there is little chance of demonstrating an effect, there is value in identifying what research could (and thus what research could not) detect it, because it serves to point out which claims are groundless.

The following analysis explains why most of the evidence that is typically cited in support of gateway claims is invalid, and the evidence is consistent with there being no gateway effect or even with extreme alternative hypotheses. The empirical-based claims that are made only support the hypothesis that people who use one tobacco product are more likely than the average person to use another. They do not support a gateway claim over two other candidate causes of such an association, discussed below. Moreover, since both of the other explanations undoubtedly create that association, the association does not imply there is any gateway effect whatsoever. However, the present analysis also points out that some of the arguments most frequently made against gateway claims are also faulty. It concludes with an assessment of what empirical evidence could be used to support or refute gateway claims and an example of applying such reasoning.

While some of the particulars of this analysis are rather unusual, there are many general lessons about analyzing cross-sectional data. The most obvious is about not confusing causation in one direction with confounding or causation in the other direction, a straightforward point but one that can always benefit from additional teaching examples. More subtle and potentially generalizable lessons can be found in the proposed strategies for dealing with that challenge, in particular how to devise severe hypothesis tests using data that is not as rich as we might prefer.

## **Analysis**

### The logic of the gateway claim

Describing the mechanism by which a cause might lead to an effect is frequently useful. Cause-effect relationships are often established without knowledge of a mechanism when studying the complexities of human biology, and due to the complexities it is always possible to suggest plausible mechanisms. But in a relatively straightforward process of preferences, decisions, and volition, as in the present case, it is possible to characterize the necessary mechanism. Doing so offers some clues about useful empirical research while also emphasizing that the claim in this context is an extraordinary claim that calls for extraordinary evidence.

Gateways certainly can exist. For example, medical opiates appear to sometimes be a gateway to heroin, with some people who never would have considered using heroin acquiring a strong preference for consuming opiates due to medical usage (which is often called addiction or dependence, though such characterizations are immaterial for present purposes), and then discovering that the net benefits of heroin are even greater, considering both the effectiveness and the relative cost of obtaining it. (Note that using the medical opiates is not a *sufficient* cause in this story; there is also the issue of availability and other causes. But it was *a* cause. See below for details of these concepts.)

But for tobacco products there is no analogous story: The barriers to starting smoking are quite low, in contrast with heroin use, so an introduction or “ramping up” effect is not needed to overcome formidable obstacles. Moreover, while some THR alternatives are a bit more difficult to acquire than cigarettes in some places due to bans or anti-THR taxes, they rarely become unavailable to a user who became accustomed to them as can happen with medical opiates. Thus, the most obvious real gateway problem that currently exists has little resemblance to use of tobacco products.

Consider the scenario that is needed to create a gateway case of smoking: Someone chooses abstinence over smoking, though the latter is an easy option. Assume that this individual is among the approximately half of the population that benefits from using tobacco (and thus is even a candidate to start using the products). The choice of abstinence shows she is among the approximately half of that subpopulation who decide that the costs of smoking outweigh the benefits (which is why she does not smoke in spite of those benefits). She is then exposed to a low-risk alternative that she prefers to abstinence. Why, exactly, would she then switch from this product to what had been her least-preferred option, smoking? The logic of this is difficult to conceive and is never explained by those who claim there is such a gateway effect.

It is not impossible, of course. At the moment, smoking delivers nicotine more effectively than existing low-risk alternatives, and includes other psychoactive agents that are not present in extracted nicotine products such as e-cigarettes. It is thus possible that someone who never even tried smoking because of concerns about its health effects could discover how much she would like it by discovering she likes the low-risk alternative. This *might* lead her to conclude that smoking has benefits that are greater than its costs. It is also possible that she had tried smoking and was not enthralled, but acquires a growing taste for nicotine via the use of a low-risk alternative which then makes the better delivery more attractive than it was initially. However, in both cases, the new *net marginal* appeal of smoking would have to be positive. That is, the advantages of smoking over and above those already provided by the low-risk product would have to be greater than the health costs of smoking (net of the trivial health costs of the alternative product), and enough to get her to overcome the inertia to switch products. Even if the change in total appeal promoted smoking over abstinence, a much greater change would be required to elevate it over the low-risk product. It is difficult to imagine this scenario occurring.

It is significant that proponents of the gateway claim never even attempt to present a scenario for why someone would ever make the gateway switch. This is presumably because presenting this scenario would illustrate how improbable and rare the effect would be.

It is possible to *engineer* a gateway from low-risk tobacco products to smoking via public policy, similar to the opiates story. A spike in the punitive smokeless tobacco taxes in Canada caused some nonsmoking smokeless tobacco users to switch to cigarettes, at least occasionally, because they became much cheaper than the smokeless product they

preferred [personal observation]. A ban (either literal or *de facto*) on e-cigarettes, which is still possible in many jurisdictions, could drive some non-smokers to smoke (though almost all would be former smokers who resume smoking, so it is not clear this should be called a gateway). An incentive can be created to switch from a low-risk product to the high-risk product, smoking, but little such incentive exists naturally.

### Causation

As is common when terms from political rhetoric find their way into scientific analysis, the word “gateway” is seldom defined by those who use it. For the gateway claim to match the policy discussion that surrounds it, it must be interpreted as the low-risk product *causing* the use of the high-risk product. This is sometimes stated explicitly and can be inferred in other cases. (The claim that the high-risk behavior is caused to *persist* as a result of the low-risk alternative, but not caused to initiate, has similar implications to the gateway claim. But it is a fundamentally different phenomenon, requiring different research, and so is not included here and does not seem to be properly included in the “gateway” label.)

An alternative potential meaning, “C (the ostensible cause) merely precedes E (the ostensible effect)” is clearly not what is meant, though there is sometimes a bait-and-switch use of such a definition. “C is a convenient step on the way to E, which is the ultimate goal” is the literal interpretation of the metaphor (you pass through a portal because you are trying to arrive at something on the other side of it, so the destination causes the use of the gateway, not the other way around), but clearly this is not the intended interpretation. Similarly, sometimes THR proponents attempt to hijack the term, saying “yes, there is a gateway, but only *away from* smoking.” While it is easy to understand the temptation, this is not a legitimate argument; the word has an accepted meaning and trying to redefine it does not constitute an argument that the phenomenon does not exist.

Though the philosophical nuances of the concept of causation are extensively debated, for practical purposes the statement “C causes E” is defined by the counterfactual statement, “if C is true then E will occur and if C is not true then E will not occur, holding constant everything else that is not caused by C.” Granted the last condition, as phrased, is a bit circular (technically it is an infinite recursion), but it gets the point across for practical purposes. *Counterfactual* refers to the fact that if one of those if-then scenarios is observed then the other cannot be, and so is necessarily counter to the facts. This is why causation can never be observed or proven, only inferred.

The previous paragraph describes causation at the level of a single event or individual. At the population level, a cause is a factor that for *some* individuals has the properties of C in the definition. When causation is inferred from data about a population (as it almost always is in social science), it will not be known which particular cases of E were caused by C and which were not (unless C is a necessary cause – see below). That C is causing some cases of E can be inferred from particular patterns in data. However, as explained below, the simplest observation about the data – a correlation between the variables that

are candidates for C and E – is not in itself sufficient information, and further analysis is required before the claim is established.

Every worldly outcome has multiple causes. Someone's choice to become a smoker might be caused by a friend handing her a first cigarette, but it is also caused by her living long enough to become a smoker and by the historical fact that the proto-tobacco plant evolved and was discovered and cultivated by humans. Indeed, the first of these is the only one that might *not* be a cause of her being a smoker: Someone inclined to smoke would quite likely have become a smoker via some other pathway had that exact event not occurred. But practical discussions of causation would typically only consider the first of these because it is a cause (one of many) whose truth value could be altered.

The friend event might not be a cause because it is the only one on that list that is not a *necessary* cause, which is to say an occurrence in whose absence the outcome would not have happened no matter what else was true. Put another way, if you listed every set of causes that would lead to the outcome if all in the set were true, a necessary cause would appear in every such list. In social sciences, necessary causes are seldom very interesting, though the failure to understand the concept sometimes creates confusion.

A *sufficient* cause is a cause that produces the effect no matter what else is true. So if we created the list of each combination of causes that result in an outcome, one of the entries on the list is the sufficient cause by itself. Single sufficient causes do not exist in the real world, except when they are tautological or definitional. For example, smoking your 100th cigarette causes you to become an ever-smoker if “ever-smoker” is defined as having smoked at least 100 cigarettes. (Strictly speaking, you can concoct a sufficient cause if you take an entry on the list and combine all of its component causes into one long conjunction. If you call that conjunction *a* cause, then it is sufficient. Because of this, each entry on the list is often called *a sufficient set* for causation.) Failure to understand the difference between sufficient cause and cause creates confusion. For example, it is sometimes suggested that “most smokers who have tried e-cigarettes are still smokers” and therefore “e-cigarettes must not cause smoking cessation.” This is wrong because it merely shows that trying an e-cigarette is not *sufficient* for smoking cessation; it is still a cause (for some individuals, and thus at the population level).

A critical observation for the present case is that many effects are *overdetermined*, meaning that all the causes of two or more of the sets in the list are met. This means that if you changed the world so that one set of causes were no longer all true, the outcome would still be caused by another set (assuming the change in the world left that set intact). In the above example, if the subject was not handed her first cigarette by a particular friend on a particular day, there is still a good chance that she would become a smoker because some other set would be completed (e.g., a month later she asked for a puff from a different friend). In common language, each item on each redundant set of causes for an overdetermined effect is often called a cause, though the counterfactual definition says that *none* of them are causes because the effect still would have occurred in their absence, except for items that appear on every one of those redundant sets like

“tobacco exists”. (A version of the classic exposition about overdetermination, the case of the canteen in the desert, can be found at [6].)

This means that if e-cigarette use prompted someone to take up smoking, but he would have started smoking in the absence of e-cigarettes due to some other prompt, then the e-cigarette use did not cause him to become a smoker. This is true in both the logical sense and the relevant practical sense – if someone would have become a smoker anyway, the other product cannot be blamed for making it happen. Of course, it may be useful to study the very different question “what are the proximate triggers to smoking?”, apart from whether they are actually the causes in either sense. But since gateway claims are about creating smokers from would-be nonsmokers, staying with the logical definition of cause is appropriate.

*Confounding*, because of the way it is presented in the context of statistical analyses, is sometimes not understood to be a statement about causation. The association between C and E in a population is confounded if there is a difference in the rate of E between the subpopulation who have C and subpopulation who do not have C *that is not caused by C*. That is, if you could change the world such that no one had C, but all else was the same, the rate of E would still be different between the two subpopulations. A simplistic comparison of the groups, then, would attribute this difference to C even though it was not caused by C.

Notice that the definition of confounding does not make reference to the concept of *confounders*, which refers to either other variables that might be the causes of that difference or, more typically, to other variables that could be put into a statistical analysis that might allow you to sort out the effect of C from the confounding. These are quite different concepts, and the latter – available variables that help control for confounding – are better labeled *deconfounders* [7]. A true confounder would include some characteristic U that is a cause of E and is also a cause of C. There are other more complicated causal roles true confounders also play, but that is the most important one: a *common cause*. For example, a propensity for emotional depression might cause the use of one drug, E, and also cause the use of another, C.

It might be possible to measure U and thus use the measurement as a deconfounder variable, or to measure proxies for U and use them as deconfounders. In practice in public health research, “controlling for confounding” consists of throwing in whatever variables the researcher has that might be (though often, probably are not) proxies for real confounders. While methods have been developed in epidemiology to improve the choices of deconfounder variables (e.g., causal flowcharts), they are almost never employed. Many covariates included in epidemiology models seem about as likely to “adjust” the measure of effect further from the true value as to adjust it closer.

But even when researchers are skilled at identifying optimal deconfounders, it is often the case that no available variables offer good proxies for the postulated confounders. For example, a researcher trying to assess whether something causes smoking may realize that he needs to control for propensity for depression and willingness to defy social

pressure, but the available covariates (age, race, level of education, income, etc.) offer little or no such deconfounding. Seldom do researchers even discuss what hypothesized the confounder variables they are attempting to proxy for, let alone offer any argument that the “control” variables they threw into their model achieve the goal of deconfounding.

Another causal pathway that has similar implications to confounding is *reverse causation*. That is, E is actually causing C. Association of two variables always works in both directions: if E is associated with C, then C is associated with E. Thus, some additional analysis, beyond observing a correlation in the data, is needed to draw conclusions about which direction the causation flows, even in the absence of confounding. For many areas of research, the reverse pathway is so implausible that it may be valid to completely ignore it, but for some questions – particularly including analysis of constellations of behavioral choices – this is not so. (*Association* is most easily interpreted as “knowing the value of one variable for an individual affects your probability of the other”. *Correlation* refers to the same phenomenon, but is a mathematical property of population data rather than a reference to worldly phenomena that is meaningful at the individual level. Note also that associations or correlations can be positive (C true increases the likelihood of E true) or negative (decreases the likelihood), but when used without clarification the words generally refer to the positive.)

Both confounding and reverse causation are plausible explanations for the statistics that are usually cited as supporting the gateway claim. Indeed, because the logic of the gateway claim shows it to be an extraordinary claim, the alternative explanations are more compelling in the absence of evidence that specifically favors the gateway explanation for the association. However, co-occurrence of two or all three of those causal pathways is possible, and thus there could still be a gateway effect even though the other effects were stronger, and they would make it hard to find. Overdetermination further complicates causal inference.

#### Use of similar drugs is inevitably associated

This observation seems so obvious that it should be unnecessary to argue the point. It is difficult to imagine a world in which use of low-risk tobacco products would not be associated with smoking. It would be true if the low-risk product were causing smoking. It would also be true if smoking were causing the use of the low risk product, in particular as part of THR efforts (reverse causation). It is undoubtedly true simply because the same people who like tobacco in one form are more likely to use it in other forms (confounding due to a common cause, the preference for using tobacco).

The closer the similarity of the products, the stronger is the association of their use. Someone who smokes a particular pack of Marlboros will almost certainly smoke another pack of Marlboros, though obviously the fact that the particular first pack was shipped to his local store and he bought it did not cause him to smoke the other. He is also far more likely than the average person to smoke or have smoked a pack of another brand of cigarettes. The probability drops somewhat when the other brand is menthol. The association is smaller, but still far from null for non-cigarette tobacco products. We

would also predict a greater probability of smoking cannabis or using cocaine than would be average for someone of the particular age, cohort, and demographics, since he has demonstrated a greater willingness and desire to use drugs compared to the average person. It should be obvious that none of those other higher probabilities were caused by the fact that the particular pack of Marlboros in question found its way to that consumer's corner store.

This is a cartoon example, and yet, blaming that particular pack for the other behaviors is exactly the logic of most empirically-based gateway claims. Common naïve gateway claims are based on the simple observation that a larger proportion of those who smoke have used a low-risk alternative as compared to those who do not smoke.

For example, Haddock et al. [8] claimed to demonstrate a gateway effect by observing a mere association between smokeless tobacco use and smoking. They also reported a strong association with other risk-taking or socially disparaged behaviors (e.g., motorcycle riding) that were obviously not caused by smokeless tobacco use, but were apparently unaware that this illustrated the fallacy of their conclusion. More recently, Glantz et al. have claimed that e-cigarettes are causing teenagers to smoke, when their data [9,10] actually provided no support for that claim and instead better supported the claim that those who already smoked were seeking a substitute [11].

#### Precedence is not causation

A variation on the claim that mere association indicates a gateway is to suggest that the association must be causal if use of the low-risk product preceded the use of the high-risk product. Some gateway claims are based on merely showing some individuals used the low-risk product before becoming smokers, and then discussing the result as if it supported the much stronger claim of causation. This is clearly invalid, not just because of the change in definition, but because there is no reason to care about non-causal precedence. The classic example is that almost every heroin user consumed milk before ever touching heroin. No one who is not trying to mislead uses the term gateway to refer to mere precedence.

A defensible approach, but still fraught with confounding, is to combine association and precedence. The observation is that nonsmokers who use or have tried a low-risk alternative are more likely to later become smokers than similar nonsmokers who have never used tobacco products. But while this can (partially) address the problem of reverse causation, it does little to reduce confounding. "Similar" is unlikely to be similar enough to eliminate the effects of common causes (absent a serious attempt to measure propensity – see below). In particular, neither having a taste for tobacco use nor being willing to defy the social pressure to avoid it can be measured well using any variable other than measures of tobacco consumption. Association is almost inevitable, given the practical limits of controlling for confounding, even if there is no causation.

Thus, not only does precedence alone not indicate a gateway effect, but observing that earlier use of the low-risk product is associated with later high-risk behavior is also inadequate evidence for a gateway.

However, contrary to a large portion of criticisms of gateway claims, the lack of the “right” temporal ordering does not disprove the gateway claim. That is, just because someone tried a cigarette before ever using the low-risk alternative does not mean that the low-risk alternative is not the cause of later smoking. Trialing smoking in the past, or even a period of regular smoking followed by cessation, does not immunize someone against the possibility of another product causing later smoking. Implicit in the gateway concept is causation of someone becoming *a smoker*, which would not have otherwise happened, not causation of someone trying her first cigarette. The latter could be ruled out by the simplest measures of temporality, but is not the outcome of interest. This is not to say that temporality data is useless; in particular it can be used to identify individuals who are much more likely to be candidates for the gateway, as is done below. But simply declaring that someone who used a cigarette first cannot be a gateway case is incorrect.

#### Testimony is useful, but overdetermination complicates inference

One good method for assessing cause and effect is to ask subjects what they perceive to be a cause. This is not definitive, for obvious reasons, but is potentially quite useful when it is an option (i.e., the objects of study are people and the apparent link between cause and effect is observable by oneself). It is a neglected research option in epidemiology, which seems largely to result from researchers’ failure to acknowledge that they are engaged in social science rather than clinical research [12], although obvious political motivations are often the explanation for ignoring the available testimonial evidence (e.g., [13]). An individual’s testimony about what caused his cancer is worthless because there is no way he can know that. But his testimony about why he decided to change his behavior can be useful scientific evidence.

Unfortunately, even if someone (accurately) communicates a proximate cause of smoking such as “my friends and I were using e-cigarettes all the time, and I wanted to try the real thing”, it does not mean that e-cigarettes were a cause because of overdetermination. It could be that the individual would have also started smoking in a world that lacked e-cigarettes, triggered by some other event. People’s testimony about the proximate trigger for smoking initiation can be reliable, but their assessment about whether the outcome was overdetermined will be much less reliable.

By contrast, someone’s testimony about conscious adoption of THR is quite reliable. Anyone who testifies that their history of product usage includes both smoking and a low-risk alternative, and that they intentionally switched from the former to the latter, is almost certainly correct. One simple survey question, in such a case, is sufficient to rule out a gateway effect in favor of reverse causation, and the failure to ask that question indicates a lack of serious interest in whether there really is a gateway.

Overdetermination is not a problem for these purposes. The overdetermination challenge returns if we try to extend the causal statement to the common stronger claim “...and I would still be smoking were it not for THR” (though other details can help support this claim, and thus the testimonial evidence can still be compelling). But even if the smoker

would have quit via other means, the testimony about the switching pattern is sufficient to rule out the gateway.

Of course, testimonial evidence can seldom be collected in a manner that allows estimation of population-level statistics. Still testimonials can rule out gateway cases, as noted, and can also identify *potential* true gateway cases, though overdetermination would prevent these from being sufficient evidence. It is noteworthy that those who claim that e-cigarettes are a gateway to smoking have failed, to our knowledge, to produce even a single testimonial in which someone reports being such a potential case.

#### Use of population prevalence trend data is not promising

If use of low-risk tobacco products becomes popular in a population, then it is theoretically a testable prediction of the gateway claim that there will be a measurable increase in smoking. That is, if low-risk tobacco product use grows, then it should cause an increase in smoking rates. In the only population (Sweden) where a low-risk tobacco product (snus) has been sufficiently prevalent that such an effect could be seen, Rodu argued that there was no apparent gateway effect [14]. It is not possible to test the prediction during a transition period, where many people are switching from smoking to the alternative (either literally or via cohort replacement); under those circumstances, there could be a gateway effect, but it is masked by the stronger reverse causation trend. However, in Sweden the rate of male snus use stabilized at a high level while smoking rates remained low, which is contrary to what would be predicted by a claim that there is a substantial gateway effect.

By contrast, recent similar claims that e-cigarettes cannot be causing a gateway because smoking continues to decline as e-cigarette use becomes more popular are fallacious. E-cigarette usage prevalence is low compared to smoking (e-cigarette trialing incidence rates among younger people appear to be similar to smoking in populations where e-cigarettes have become popular, but an actual usage practice appears to be remaining much lower). E-cigarette use is dominated by former smokers who intentionally used e-cigarettes to quit, and thus are unlikely candidates for a gateway effect. Thus, when we consider the very few people who adopted e-cigarette use other than for THR, even if half of them were caused to start smoking (presumably far more than even gateway proponents suggest could happen), it would not be detectable. Any signal would be completely swamped by the trends in smoking prevalence and the noise in the estimates alone, to say nothing of the countervailing THR effect.

The gateway hypothesis could be amended to say that a low-risk alternative causes uptake of cigarettes in a population *where cigarettes are currently a lot more popular than the alternative*. This caveat might be seen as cynical expedience, but it is not unreasonable. The claim would be that when people start using any tobacco product, it causes some to drift among multiple tobacco products, and social forces will cause more to drift toward and settle on the most popular product. Indeed, this is the obvious explanation for observations that a larger portion of people who use smokeless tobacco in the USA equilibrate as smokers rather than the other way around, while the converse occurs in Sweden. Any inclination to switch to smoking would be reduced if smoking

were not common in a population. With the addition of this nuance, no population data could ever effectively test the hypothesis because the signal would always be lost in the noise: Once the alternative was sufficiently popular to extract the signal from the noise, smoking would be sufficiently unpopular that it nullified the claim.

Thus, while the Swedish data provides strong evidence against a simplistic gateway claim, no such data will be available for other products or other populations for the foreseeable future, and even then it would not rule out that there is still a gateway while smoking is far more popular.

### Experimental studies

Though not a realistic option, it is useful to consider the implications of a gateway experiment, where abstainers are assigned to use low-risk tobacco products to see if that causes smoking. The advantage of experiments (a.k.a. randomized controlled trials) is that they can eliminate systematic confounding and reverse causation as explanations, and are not affected by population trends. However, the differences between any operationalizable experiment and the real exposure of interest, inevitably introduced by an artificial controlled setting when studying complicated human behaviors, tend to be worse than the problems that are solved. It is often claimed that the reason that randomized trials of tobacco use behavior are not promising is that they would be unethical; while that is largely true, it is not actually the fundamental problem.

The conceptually ideal experiment, always useful to consider when contemplating research design, would be to run two histories of the world, one where the low-risk alternative existed and one where it did not. This would allow observations about how many nonsmokers in the latter world are smokers in the former (with the added bonus of observing how many smokers quit only because of THR). Monitoring corresponding individuals in the parallel worlds would solve the problem of merely analyzing aggregate counts, as in the Swedish example, which shows only the net change from any gateway and any THR effects. From a practical standpoint, many would argue that only the net population effect matters (though there are valid ethical objections to such netting), but it does not answer the specific question about a gateway.

The conceivable real-world experiment, which always falls short of the theoretical ideal in social science, consists of selecting a random collection of nonusers of tobacco products, from the population (age group, etc.) of interest, and conscripting a random sample of them into using a low-risk alternative. If they became smokers more often than those who were not conscripted or were conscripted to a null-treatment arm, it would suggest there is a gateway effect. But conscripted use would obviously be quite different from real-world adoption so even this theoretically possible experiment would tell us little.

Observational studies in epidemiology are typically misinterpreted as if they were an experiment, but the fundamental differences are obvious in this case. People who voluntarily try the low-risk products are not a random sample from all nonusers (i.e., there is important confounding), those who transition from trialing to using are not a

random sample of them (further confounding), and many of them may be trying to substitute for their inclination to smoke, even those who have never been smokers (reverse causation). Using the only available evidence of real-world behavior – which is observational – is much more difficult than interpreting results from an ideal experiment. It can be done but, in this case and more generally, it requires active and honest attention to testing the competing hypothesis, confounding and reverse causation.

### Magnitude matters

Before considering what available evidence would be useful to support or refute the hypothesis, it is important to quantify the hypothesis. If it is merely “there is one person who has been caused to start smoking due to use of the low-risk alternative” then there are no realistic statistics that could address the claim. A testimonial by an individual could support that claim, though the overdetermination problem would call the “I would have otherwise never started smoking” aspect of the testimony into serious doubt unless the individual is an older never-smoker. If the hypothesis is “half of would-be never-smokers who use the alternative product will become smokers,” available statistics, used properly, would address the claim. It would be trivial to observe that more than half (to account for those destined to become smokers anyway) of never-smoker users of the alternative become smokers.

Presumably the hypothesis falls somewhere in between the former inconsequential claim and the latter implausible claim. That “presumably” is critical, because proponents of gateway effect never quantify the claim they are making. This is a serious problem, both for the policy discourse and the epistemology; without quantification, the claim is disingenuous and research is meaningless.

For purposes of this analysis, assume that the hypothesis of interest is both plausible and large enough to matter, that something in the order of 5% of would-be never-smokers who take up the alternative product are caused to smoke.

### How can researchers assess whether there is a gateway?

It is not entirely clear that any systematic study can detect gateway behavior of that magnitude, given how overwhelming the confounding is. If there is any hope, it needs to be based on thoughtful analysis about what constitutes evidence, rather than blind application of first-semester epidemiology methods.

### *The need for testimonial evidence.*

Testimony-based research would be useful for directing quantitative analysis. Given that the proposed gateway behavior defies the welfare-maximizing logic described above, it is difficult to know what data to collect without some research about the nature, motives, and dynamics of apparent gateway cases. Naively hunting for evidence of a gateway effect in quantitative data is likely to be a dead-end because the gross associations that are likely to be discovered do nothing to resolve among competing explanations. But if the specifics of apparent gateway cases were documented, it might be possible to refine the statistical analysis.

For example, it might be that most proposed examples of gateway cases included the story, “I tried a few packs of cigarettes a while ago, but did not really like them; but then I used e-cigarettes every day for six months and found myself thinking that a real cigarette sounded good after all.” If that were the case, the systematic research should focus on those who had previously consumed some quantity of cigarettes but had never become a smoker, and then later used e-cigarettes for a substantial period. It is also possible that such testimonials would reveal potential mediation analyses (statistical analyses based on intermediate variables that are a step along a causal pathway), with testable implications that would not be lost in the noise as the aggregate values are. For example, if it were found that self-identified gateway cases mostly followed a particular pattern of smoking adoption that was generally not common among those starting smoking via other causal pathways, that pattern could be investigated in population statistics. This is a generalizable strategy when attempting to distinguish a weak causal relationship from stronger confounding effects.

Currently such hypotheticals are moot because it appears that no gateway proponent has ever offered the testimony-based *prima facie* evidence that there are *any* instances of gateway behavior. Statistics alone, absent case studies that suggest plausibility, cannot justify such an economically extraordinary claim. Without evidence of even a single case, the claim is not even a defensible hypothesis and more like a religious myth or philosopher’s demon theory, something made up from whole cloth for rhetorical purposes.

#### *Bayesian analysis shows how to avoid garbage-in-garbage-out*

Obvious Bayesian reasoning shows the lack of useful information in the recent studies of e-cigarettes. Merely finding an association between e-cigarette use and smoking does not usefully update any reasonable set of prior beliefs. It should reduce one’s posterior probabilities that there is no association (which should have a very low prior in any case), but alone it should not change the relative probabilities of the competing explanations for the association: confounding, causation in the THR direction, and causation in the gateway direction.

The recent claims can be interpreted as examples of “faith-based Bayesianism”. To take an example I presented previously [11], consider someone whose prior beliefs are that there is a 50% chance that e-cigarettes cause smoking, a 1% chance that e-cigarettes are used for THR, a 1% chance that there is a strong association caused by confounding, with the remaining probability that there is no association. Further assume that there is an 80% probability a study will show an association under the first three hypotheses and a 5% chance under the last. (This characterization is not really proper since magnitude matters for updating the priors, not just this dichotomous characterization that is typical in epidemiology, but set that aside). The posterior probabilities, applying the Bayes Theorem after the observation of association, would be .91, .02, .02, and .05, which appears to be strong support for the first hypothesis. But, of course, the observation increased the probabilities of all of the first three hypotheses by the same ratio, at the expense of the fourth, and so the overwhelming support for the first compared to the

second and third is just an echo of the prior beliefs. This is not valid empirical reasoning, and represents the common critique of simple Bayesian analysis: garbage in, garbage out.

#### *Hypothetico-deductive reasoning*

A good hypothetico-deductive test is based on asking, “what would we expect to observe if this hypothesis were true which we would not observe if it were false?” Such reasoning has recently been discussed in epidemiology under the rubric “negative controls for confounding.” An example of this is observing that the association of better health and getting an influenza vaccination in an elderly population are not stronger during flu season, which is not what we would expect if the association were caused by the vaccine working rather than by confounding [15]. Similarly, if an ostensible cause is strongly associated with an outcome variable that it cannot possibly cause or be caused by, and this persists after “controlling for confounding,” it is apparent that there is uncontrolled residual confounding that may explain the main relationship of interest. The above cited case of motorcycling being associated with smokeless tobacco use provides a good example.

It is often possible to devise hypothetico-deductive tests that can be performed on a particular dataset. Unfortunately there is no easy recipe for identifying them, and they need to come from *ad hoc* scientific reasoning. An example of this appears below.

Another possible application of hypothetico-deductive reasoning is to take advantage of natural experiments. Where there is spatial heterogeneity of low-risk tobacco product use within a similar population, the gateway hypothesis predicts corresponding heterogeneity of smoking adoption rates. The popularity of e-cigarettes among teenagers appears to have involved substantial spatial heterogeneity across schools, due to contagion effects [16] (i.e., it is a social process that is either adopted in a social group or not). For years there were reports of some American high schools repeatedly confiscating e-cigarettes from students, while students at other schools concurrently reported never having seen one. Though overall smoking trends are uninformative for the reasons cited above, the hypothesis makes both positive and negative prediction regarding this heterogeneity which could provide a good test. Something similar might be possible among Norwegian users of smokeless tobacco, where there has been explosive adoption unlike the relatively steady state in Sweden. Anyone genuinely interested in detecting a gateway effect should take advantage of this situation before the heterogeneity disappears. Indeed, it may already be too late to collect such data, but existing recent data might be available and new examples might be found as e-cigarettes become popular in countries where they are not yet.

#### *Propensity scores: mimicking the ideal experiment*

Some hope of finding an answer to that lies in the counterfactual concept, and the notion of the ideal experiment. We want to predict whether someone was destined to become a smoker in a world that lacked a low-risk alternative in lieu of being able to observe that. If those who did use the low-risk alternative became systematically more likely to smoke than would otherwise be predicted, then we could conclude there is a gateway. This suggests the use of propensity scoring or multi-state regression, in which we collect

variables that predict smoking behavior – e.g., parental smoking, SES, educational success, risk-taking behavior – and create a statistical combination of them that has the best predictive value for smoking. While it would never be perfect, it could be a sufficiently good predictor that we could notice if some other variable (i.e., low-risk product use) was substantially changing the likelihood of someone smoking. Apparently only one study ever attempted this for THR products, and found no evidence of a gateway effect [17].

This approach is common in some social sciences, but is largely absent from substance-use epidemiology even though the concept of propensities is a natural fit there. Epidemiology's use of arbitrary and largely untested covariates as "control" variables is an inadequate substitute for propensity scoring or other hypothesis-based statistical models, though it can still be sufficient to show that confounding may explain an uncontrolled result [18].

A proper propensity score is a reasonable predictor of the effect (smoking) using variables that are *mostly* independent of the ostensible cause (low-risk product use). The epidemiology approach to "controlling for confounding" might use the same demographic and general personality covariates, but when used as simple deconfounders they merely serve as weak proxies for the underlying common cause. Meanwhile, more detailed behavioral variables might actually do more harm than good, such as by improperly "controlling for" intermediate steps in complicated causal pathways of interest. The variables may really attenuated measures of the true underlying common cause but the effects of "controlling" may result from other relationships, which often should not be controlled for. By contrast, an effective propensity score provides an attenuated version of the clean ideal experiment (and thus the association will be attenuated, but still present). The propensity score tells us, probabilistically, whether someone would have smoked absent the low-risk product, and the quality of the predictor can be assessed purely in terms of its predictive value. It is not necessary to postulate nearly as much about causal pathways.

Ideally the variables used to create the propensity score are completely independent of the possible causal variable of interest. This is unlikely to be possible in the present context since the main problem is that of common causes. (As an aside, note that this means that the similar strategy, use of instrumental variables to deal with simultaneous causation, is not promising. This is largely moot, however, since the instrumental variables are particularly valuable when analyzing time series data. When examining cross-sectional data, there is little reason to use this method rather than a full-on two stage method.) However, this is not a fatal problem. So long as the variables used to measure propensity cannot plausibly be caused by low-risk product use and are not intermediate steps on a pathway by which low-risk product use causes smoking, the approach is still valid. The addition of temporal data can help determine the direction of causation.

In the case of e-cigarettes, fairly recent data from before e-cigarettes were available could be used to create the score, which at least eliminates the possibility that e-cigarette use

affected any of the variables. A recurring theme here is that researchers interested in honest attempts to detect a gateway effect need to take advantage of a temporary window, before almost-random heterogeneity disappears and propensity data from before the ostensible gateway product became popular is too old. Once this window closes across the populations of interest, it is not clear there is much hope of sorting out the effects of competing causes until such a time that a Swedish-style stability is reached.

#### A checklist for evaluating empirical research about gateway effects

To summarize the above, the following considerations should be applied to the design and interpretation of any empirical study on the gateway effect. Many of these principles generalize to other attempts to distinguish among the three possible causes of an association in cross-sectional data.

1. Is the research based on (a) a clear statement of what “gateway” means, and (b) a hypothesized quantification of what is being tested or claimed? Failing either of those, the effort is only propaganda, not science. Is there any presentation of (c) a theory of why there would ever be gateway behavior, or (d) testimonial evidence suggesting there is any such case and how it came about? Failing both of those, it is unlikely that the statistical analysis will be anything other than a misguided fishing expedition.
2. Does the result or potential result make any effort to discriminate between a gateway effect and causation in the other direction (THR) or confounding? If not, it is completely uninformative about whether a gateway, rather than one of those established sources of association, is occurring.
3. Is the attempt to deal with the inevitable confounding based on some theory of the nature of the confounding (i.e., addressing the counterfactual or hypothetical experiment concept) and is it empirically tested (e.g., by checking the sensitivity to combinations of variables), or does it just consist of throwing in whatever covariates are conveniently available? If the latter, it is probably little better than no attempt to control for confounding, as is likely to be demonstrated by the persistence of negative control associations. If the former, the stated confidence in its success should still be epistemically modest.
4. Do the study results support any prediction that would be true under the hypothesis but unlikely if the hypothesis is false? If the data is highly consistent with the hypothesis being false, it obviously does not provide much support for the hypothesis.
5. Is there any attempt to detect temporal ordering of behaviors? As noted, this is neither necessary nor sufficient for a gateway, but the “right” ordering tends to better support the claim for obvious reasons. In particular, if the data shows that most of the association results from people who are already smokers trying the other product (as is likely for e-cigarettes and other THR products), and if it appears that the association is causal, then the causation is apparently flowing in the other direction.

6. Did the authors report enough different cuts at the data to make clear that their conclusion that the data support a particular hypothesis is apparently not an artifact of the specific models and results that are reported? Researchers can easily check how their results change based on innocuous-seeming details in their models (such as choosing among multiple candidate definitions of smoking, which covariates to control for, which cohorts from a dataset to use, etc.) and choose the set details they like best, leaving the readers (including journal reviewers) unable to detect this was done. This is an overarching problem with epidemiology, not specific to the gateway question [19]. But it is especially critical for highly-politicized hunts for a tiny signal amidst a lot of noise. It is trivial to cherry-pick models and results that tend to support the author's preferred claim. Adoption of methods that mitigate this is clearly beneficial (e.g. [20]).

#### An application of these principles

It is possible to apply these principles to a recent article by Dutra and Glantz (2014) [9], that was widely claimed to show there is a gateway effect. This included claims by the authors in various concurrent and subsequent communications [e.g., 21-24], which suggest this claim was their main interest though they avoided the term "gateway" in the paper itself. (Details of their claims can be found in my previous critique of the original paper [11].) Doing so shows:

- (a) The results failed completely to show what the authors and other commentators claimed, and
- (b) the variables available in the dataset in question were inadequate to answer the gateway question, however
- (c) further analysis of the dataset shows that the claim is even less well supported than the authors' own analysis could be legitimately interpreted.

The original paper used the 2011 and 2012 U.S. National Youth Tobacco Survey to look at the association, among teenage students, between smoking and what they refer to as e-cigarette "use." (This is a controversial and misleading term, as explained in the Appendix.) The analysis was limited to subjects who reported having ever smoked, at least once, and the endpoints consisted of the available measures of smoking, specifically smoking a total of 100 cigarettes and having smoked (at all) in the last 30 days. Note that the original authors made odd use of the available data (e.g., omitting anyone who had never tried a cigarette; using at least one idiosyncratic definition) and only reported the results from a single model without explaining their choice. But the present analysis is not an attempt to reanalyze their entire paper. In order to focus on the above-cited methodological points, what follows is based on replicating their methods, models, and definitions as closely as possible. This should not be seen as an endorsement of them.

The authors observed strong associations between measures of smoking and measures of vaping (their Tables 2 and 3), which serves as the entire basis for the gateway claims

(along with a strong prior assumption that there was a gateway). They claimed emphatically that their results were evidence that e-cigarettes were a gateway to smoking.

Point (a) was presented in detail in my previous critique. In particular, the original paper involved *no* attempt to discriminate between a gateway effect and either confounding (which the authors did not even acknowledge exists, never even using the word) or reverse causation. If the exposure and endpoint variables are reversed, the association remains, as will always be the case (it is illustrated in the Appendix for concreteness). If having tried an e-cigarette predicts being a smoker (the gateway phrasing), then being a smoker predicts having tried an e-cigarette (the THR phrasing). The results support no prediction that would be true under the gateway hypothesis that would not be true if the hypothesis were false (unless the only alternative hypothesis considered is the implausible alternative that there is no association between using one tobacco product and using another).

Moreover, the authors and others who made gateway claims offered no theory of why or under what circumstance a gateway might occur, nor did they give any indication of what magnitude of gateway effect they were testing for or claim to have found. The failure to even try makes clear that they were not making a serious attempt to empirically validate a scientific hypothesis.

Given the lack of acknowledgment of confounding in the original paper, it is no surprise that there was no serious attempt to control for confounding. Moreover (point (b)), the NYTS dataset is simply not rich enough to do so. The original authors controlled for age (which is obviously critical when studying a “have you ever...” phenomena), gender (also important for drug-use behavior), and race (which is rather less useful, though might be a very rough proxy for some cultural phenomena). They control for nothing else (other than which year of data collection a subject was in), and indeed, there is very little else they could have done with this data to assess propensity to use tobacco products. It should be obvious that those few variables do not come close to serving as good propensity estimators or proxies for underlying common causes. Thus there is simply no way this dataset can be used to discriminate a gateway effect from the inevitable confounding.

However, the data can be used more effectively than the original authors attempted. The Appendix shows a replication of the numbers from the original authors to demonstrate that what follows is a simple extension of what they did; the numbers are not quite identical because, as is typical for public health research, the methodology was not reported sufficiently to permit exact replication. Tables showing the results noted in the remainder of this section comprise the rest of the Appendix.

The original authors ignored data that would help predict, albeit roughly, which subjects were smokers before they ever tried e-cigarettes. The dataset is very limited, asking temporal questions only about cigarettes, but since e-cigarettes were a very new phenomenon at the time of the survey, it is safe to infer that almost everyone who tried e-cigarettes had first done so recently.

Reiterating the point that temporality is not decisive (one way or the other) but is useful (since precedence is clearly associated with causation), we can estimate that of those defined as “current smokers” who had tried an e-cigarette, about three quarters (from the 2011 survey) or half (from the 2012 survey) tried their first cigarette before 2009. (This is based on their age at the time of the survey and their reported age of trying their first cigarette, as detailed in the Appendix; we recognize the imperfect correspondence between this and calendar year, but describe it in calendar terms for clarity.) This almost certainly means they tried a cigarette before trying an e-cigarette. Of course, many of those who first tried a cigarette later than 2009 also did so before trying an e-cigarette. While this cannot rule out gateway cases for this subpopulation (as noted, *trying* a cigarette does not immunize you from a later event causing you to become a *smoker*), it seems to provide better support for the reverse-causation (i.e., THR) claim than the gateway claim. We would expect that a substantial portion of those who tried a cigarette earlier, and also met one of the authors’ definitions of becoming a smoker, became a smoker before trying an e-cigarette.

To improve on this simple (arguably, simplistic) cut at the temporality data, there is a more formal hypothetico-deductive test we can perform. This makes use of the novelty of e-cigarettes as described above – i.e., gateway cases would have to have started smoking fairly recently. We could do better if we had temporal data on when subjects first tried an e-cigarette.

If the association of smoking and e-cigarette trialing or use is substantially explained by the gateway effect, then the association with e-cigarette use should increase as we narrow the population of smokers toward those who are more likely to be gateway cases. While long-ago trialers might still be new gateway cases, there is little doubt that those who more recently tried their first cigarette are more likely to be gateway cases because they are more likely to have become smokers recently. If many of them were caused to smoke by their (necessarily) recent e-cigarette use, the association will become stronger when we remove the dilution that results from including those who were more likely to have started smoking earlier. If there is a stronger association among the group that tried a cigarette earlier (in calendar time), and thus almost certainly has more established smokers, it would suggest that THR or some other reverse-causation is more likely: established smokers are the ones who are more often adopting e-cigarettes. If there is little difference, it will tend to support (though obviously not definitively) the hypothesis that confounding dominates both of those explanations. (It is worth noting the epistemically important fact that this test was devised and this paragraph was written before the data was analyzed.)

The simple counts of when subjects first tried a cigarette show earlier smoking among those who had tried e-cigarettes as compared to those who never tried e-cigarettes, which by itself is enough to favor the THR interpretation over the gateway interpretation. We used the original methodology and divided the population by year of first trying a cigarette, which confirms this (details in the Appendix). For parsimony the two waves of the survey are combined, with a covariate for survey wave. I recognize that odds ratios

(ORs), which we use to follow the original analysis, are a terribly misleading statistic for comparing proportions. However, they are adequate for showing whether associations are stronger or weaker between different subpopulations, which is the present goal.

For an endpoint of having ever smoked 100 cigarettes, the OR for ever trying an e-cigarette for those who tried their first cigarette before 2009 is 7.7, versus 5.2 for those who tried their first cigarette more recently. These numbers should not be over-interpreted. The results are similar, though the contrast is weaker, for “current smokers”. We do not see a trend when we step the cutoff through 2007, 2009, and 2010, and the same result does not show up when we look at the “last 30 days” e-cigarettes exposure variable (each of these results appears in the Appendix). Still, this test does show that the available data could be used to test a prediction of the hypothesis in question, and when that is done the results provide somewhat better support for the THR explanation than they do the gateway hypothesis: The association between smoking and trying an e-cigarette tends to be somewhat stronger as we select for long-term smokers and weaker as we select for those who might be gateway cases.

This analysis was easy to perform, and recognizing its potential value in answering the scientific question took only a bit of serious contemplation. Its absence from the original authors’ analysis, along with the more blatant omissions already noted, suggest that the authors were seeking to support the conclusion rather than test the hypothesis. Moreover, journal editors and reviewers still signed off on publishing the paper, demonstrating a lack of critical analysis on their parts.

In addition, this analysis illustrates the ease of cherrypicking results that support a particular claim when authors report only one of many potential results. The results, all considered, support the conclusion that THR better explains the association than does the gateway claim. Had I followed the common practice in epidemiology and public health, however, reporting only the “best” numerical results, it would have misled the reader into believing the support for the conclusion is stronger than it really is. Instead, this truth-seeking method of reporting of the results, showing the results for the eleven other cuts at the data that occurred to me to check, shows that the support for that conclusion is rather weaker than the cherrypicked strongest numbers would have shown. A typical health journal article would have reported only those numbers and would not have even mentioned that the others had been calculated, and would have implied the support for the conclusion was definitive rather than vague and suggestive. It seems likely that such hints of support are the best we are likely to get on this topic, pointing out the need for careful methods, including sufficient reporting to show that authors are not cherrypicking much “clearer” results than the data really supports.

## **Conclusions**

Searching for some signal of a gateway effect amidst overwhelming confounding requires more rigorous methods than are typical in public health epidemiology. This generalizes to any attempt to use cross-sectional data to sort out causation in a particular direction from confounding or reverse causation. When seeking epidemiologic associations where confounding is minimal or relatively simple in its causes, the typical

methods used in the field are still far from optimal, but the empirical results might still be basically useful. That is not the case in this context. While it might never be possible to convincingly demonstrate a gateway effect given the challenges, and statistical analyses have no hope of detecting a tiny effect, there are clearly better and worse ways to pursue the question.

### **Acknowledgments**

Catherine Nissen, CASAA Research Associate, provided research assistance. CVP's work on this paper was done under the auspices of CASAA. CASAA is a consumer NGO dedicated to promoting tobacco harm reduction. Some of the research related to this paper, though not the writing of this paper itself, took place in 2014 when CVP was the recipient of an unrestricted grant from British American Tobacco for support of his research.

## APPENDIX

See main text for context and how these results were used. The first table replicates the 2011 survey wave results from Table 3 of Dutra and Glantz (2014)[9]. As noted in the text, the original authors made various dubious modeling choices, but this is not an attempt to broadly critique their paper. Replicating, to the extent possible, their methodology is sufficient to make the points at hand.

The original authors defined “current smoking” to mean ever smoked 100 cigarettes and smoked (at all) in the last 30 days. The two e-cigarette variables are based on ever using an e-cigarette (at all) and using an e-cigarette (at all) in the last 30 days. Many authors, including Dutra and Glantz refer to these as “used” and “current use.” This is controversial, despite its parallel with the traditional characterization of smoking status, and has been used in political rhetoric to intentionally exaggerate e-cigarette use among teens. Any use in the last 30 days is a decent proxy for a current practice of actually smoking in an adult, since smoking tends to either be an established habit or entirely absent. But this is not so accurate for teenagers and especially for e-cigarettes, which are a novelty that teenagers might try in social settings (and since someone is unlikely to give away their e-cigarette, in contrast with the norm of giving a whole cigarette, so it could easily just be a single puff), or that smokers might use as an occasional convenient substitute. The “use” statistics are presented without such caveats, and the typical reader is likely to interpret “uses” to mean something like “many times per day, every day” even though an affirmative survey answer might refer only to taking a single puff ever, which happened to have been within 30 days. That said, these tables are designed to mimic the original authors’ approach, including their terminology.

The estimated associations are presented odds ratios (ORs; not the proper statistics to use when the endpoint is a proportion, but adequate for all present purposes as noted in the main text), with 95% confidence intervals.

	Ever smoked at least 100 cigarettes	Current Smoking
Ever tried e-cigarette	7.7 [5.5,10.8]	
E-cig use		6.8 [3.9,11.8]
Age	1.3 [1.2,1.4]	1.3 [1.2,1.4]
Black	0.36 [0.2,0.6]	0.36 [0.2,0.5]
Other race	0.72 [0.5,1.0]	0.69 [0.5,0.9]
Male	1.4 [1.1,1.7]	1.4 [1.2,1.8]

The next table shows the obvious result that the association remains if you reverse the ostensible cause and the ostensible effect (i.e., the dependent variable and the independent variable of interest).

	Ever ecig	Current e-cig use
Ever smoked at least 100 cigarettes	7.7 [5.5,10.8]	
Current Smoking		6.7 [3.9,11.5]
Age	0.99 [0.9,1.1]	0.84* [0.8,0.9]
Black	0.40 [0.2,0.7]	0.59 [0.3,1.3]
Other	0.67 [0.5,0.9]	0.76 [0.5,1.2]
Male	1.3 [1.0,1.7]	2.3 [1.5,3.6]

The following counts identify the portion of the “current smokers” study population that tried their first cigarette more and less recently. Note that the description of this is simplified to “before/by 2009” because calendar time is important for the analysis, based on the emerging availability of e-cigarettes. The survey, however, did not ask about calendar time of first trying, but age at first trying. This along with current age of the subject allows a rough calculation of calendar time, but obviously the step functions do not line up perfectly. The definition of “before 2009” is {if [current age]-[age at first trying] >3}, for data collected in 2012, using >2 for data collected in 2011. Someone who is 18 on a given day in 2012 and tried a first cigarette at 14 (defined as “before 2009”) probably did so in 2008 (assuming every distribution in sight is close to flat), but it might have been 2007 or 2009.

	<u>2011 survey wave</u>	
Year tried first cigarette	Never tried e-cigarette	Ever tried e-cigarette
≤ 2008	51% (n=788)	71 (229)
2009-2011	49 (750)	29 (94)
	<u>2012 survey wave</u>	
≤ 2008	33 (412)	44 (386)
2009-2012	67 (835)	56 (486)

The following table follows the original methodology, as described above, with the exception of pooling the two survey waves and adding a covariate for which wave included the record. The population is divided into those who first tried a cigarette before 2009 versus more recently, as well as can be calculated per the above. The purpose of this analysis is explained in the main text.

	Dep. Variable = ever smoked			Dep. Variable = current smoker		
	All	Before 2009	2009 and after	All	Before 2009	2009 and after
Ever tried e-cigarette	6.5 [5.4,7.9]	7.7 [6.0,9.8]	5.2 [4.1,6.6]	6.3 [5.2,7.5]	6.7 [5.3,8.4]	5.2 [4.0,6.7]
Male	1.3 [1.2,1.4]	1.3 [1.2,1.4]	1.2 [1.1,1.3]	1.3 [1.2,1.4]	1.3 [1.2,1.4]	1.2 [1.1,1.3]
Age	1.4 [1.2,1.6]	1.4 [1.2,1.7]	1.2 [1.0,1.6]	1.4 [1.2,1.6]	1.4 [1.1,1.6]	1.3 [1.0,1.6]
Black (ref=White)	0.38 [0.3,0.5]	0.28 [0.2,0.4]	0.36 [0.2,0.6]	0.43 [0.3,0.6]	0.35 [0.2,0.5]	0.36 [0.2,0.6]
Other race	0.72 [0.6,0.9]	0.53 [0.4,0.7]	0.73 [0.6,0.9]	0.76 [0.6,0.9]	0.61 [0.5,0.8]	0.73 [0.6,0.9]
Year 2012	0.58 [0.5,0.7]	0.64 [0.5,0.8]	0.89 [0.7,1.2]	0.58 [0.5,0.7]	0.63 [0.5,0.8]	0.87 [0.6,1.2]
Observations	11183	4,396	6,787	11183	4,396	6,787

The following two tables repeat the previous table’s approach, substituting “before 2007” and “before 2010” for “before 2009” (the former two being defined analogously to the latter).

	Dep. Variable: DG ever smoked2			Dep. Variable: DG Current smoked3		
	Pool	Before 2007	2007 and after	Pool	Before 2007	2007 and after
Ever tried e-cigarette4	6.538 [5.4,7.9]	8.672 [6.4,11.7]	6.031 [4.9,7.4]	6.250 [5.2,7.5]	6.873 [5.2,9.1]	5.975 [4.8,7.4]
Male	1.311 [1.2,1.4]	1.276 [1.2,1.4]	1.248 [1.2,1.3]	1.298 [1.2,1.4]	1.222 [1.1,1.3]	1.253 [1.2,1.3]
Age	1.432 [1.2,1.6]	1.635 [1.3,2.1]	1.283 [1.1,1.5]	1.413 [1.2,1.6]	1.617 [1.3,2.1]	1.262 [1.1,1.5]
Ref. White						
Black	0.378 [0.3,0.5]	0.254 [0.2,0.4]	0.367 [0.2,0.6]	0.427 [0.3,0.6]	0.346 [0.2,0.5]	0.394 [0.3,0.6]
Other	0.723	0.628	0.616	0.759	0.806	0.600

Ref. Year 2011	[0.6,0.9]	[0.5,0.8]	[0.5,0.7]	[0.6,0.9]	[0.6,1.1]	[0.5,0.7]
Year 2012	0.582	0.573	0.690	0.577	0.555	0.686
	[0.5,0.7]	[0.4,0.8]	[0.5,0.9]	[0.5,0.7]	[0.4,0.7]	[0.5,0.9]
Observations	11183	2,281	8902	11183	2,281	8902

1DG experimenters: Ever tried a cigarette one or two puffs  
2 DG ever smoked: smoked at least 100 cigarettes in lifetime  
3 DG current smokers: 100+ cig and smoked at least one day during the past 30 days  
4 Ever tried e-cigarette: Ever tried e-cigarettes  
All models applied sample weights; 95% confidence intervals in brackets

	Dep. Variable: DG ever smoked2			Dep. Variable: DG Current smoked3		
	Pool	Before 2010	2010 and after	Pool	Before 2010	2010 and after
Ever tried e-cigarette4	6.538	6.630	5.486	6.250	6.048	5.415
	[5.4,7.9]	[5.3,8.3]	[4.2,7.1]	[5.2,7.5]	[4.9,7.4]	[4.1,7.2]
Male	1.311	1.319	1.088	1.298	1.306	1.086
	[1.2,1.4]	[1.3,1.4]	[1.0,1.2]	[1.2,1.4]	[1.2,1.4]	[1.0,1.2]
Age	1.432	1.402	1.330	1.413	1.368	1.353
	[1.2,1.6]	[1.2,1.7]	[1.0,1.8]	[1.2,1.6]	[1.1,1.6]	[1.0,1.8]
Ref. White						
Black	0.378	0.296	0.368	0.427	0.359	0.362
	[0.3,0.5]	[0.2,0.4]	[0.2,0.6]	[0.3,0.6]	[0.3,0.5]	[0.2,0.7]
Other	0.723	0.569	0.808	0.759	0.627	0.804
	[0.6,0.9]	[0.5,0.7]	[0.6,1.1]	[0.6,0.9]	[0.5,0.8]	[0.6,1.1]
Ref. Year 2011						
Year 2012	0.582	0.691	1.076	0.577	0.675	1.018
	[0.5,0.7]	[0.5,0.9]	[0.7,1.6]	[0.5,0.7]	[0.5,0.9]	[0.7,1.5]
Observations	11183	6,040	5143	11183	6,040	5143

1DG experimenters: Ever tried a cigarette one or two puffs  
2 DG ever smoked: smoked at least 100 cigarettes in lifetime  
3 DG current smokers: 100+ cig and smoked at least one day during the past 30 days  
4 Ever tried e-cigarette: Ever tried e-cigarettes  
All models applied sample weights; 95% confidence intervals in brackets

The following three tables repeat the above three tables, for recently using an e-cigarette rather than ever trying.

	Dep. Variable: DG ever smoked2			Dep. Variable: DG Current smoked3		
	Pool	Before 2009	2009 and after	Pool	Before 2009	2009 and after
Current e-cigarette users4	7.461	7.218	7.359	7.731	7.380	7.576
	[5.6,10.0]	[5.1,10.3]	[5.1,10.6]	[5.8,10.3]	[5.4,10.1]	[5.3,10.9]
Male	1.343	1.343	1.194	1.334	1.326	1.194
	[1.3,1.4]	[1.3,1.4]	[1.1,1.3]	[1.3,1.4]	[1.2,1.4]	[1.1,1.3]

Age	1.448 [1.3,1.7]	1.433 [1.2,1.7]	1.270 [1.0,1.6]	1.427 [1.2,1.6]	1.352 [1.1,1.6]	1.318 [1.0,1.7]
Ref. White						
Black	0.300 [0.2,0.4]	0.208 [0.1,0.3]	0.311 [0.2,0.5]	0.339 [0.3,0.4]	0.261 [0.2,0.4]	0.312 [0.2,0.5]
Other	0.636 [0.5,0.8]	0.455 [0.4,0.6]	0.676 [0.5,0.9]	0.666 [0.6,0.8]	0.510 [0.4,0.6]	0.674 [0.5,0.9]
Ref. Year 2011						
Year 2012	0.758 [0.6,0.9]	0.824 [0.7,1.0]	1.149 [0.9,1.5]	0.741 [0.6,0.9]	0.790 [0.6,1.0]	1.119 [0.8,1.5]
Observations	11183	4,396	6,787	11183	4,396	6,787

1DG experimenters: Ever tried a cigarette one or two puffs

2 DG ever smoked: smoked at least 100 cigarettes in lifetime

3 DG current smokers: 100+ cig and smoked at least one day during the past 30 days

4 Current e-cigarette users: used e-cigarettes at least one time during the past 30 days

All models applied sample weights; 95% confidence intervals in brackets

	Dep. Variable: DG ever smoked2			Dep. Variable: DG Current smoked3		
	Pool	Before 2007	2007 and after	Pool	Before 2007	2007 and after
Current e-cigarette users4	7.461 [5.6,10.0]	9.717 [5.8,16.2]	6.574 [4.7,9.1]	7.731 [5.8,10.3]	9.040 [5.7,14.3]	6.941 [5.0,9.6]
Male	1.343 [1.3,1.4]	1.324 [1.2,1.4]	1.277 [1.2,1.4]	1.334 [1.3,1.4]	1.272 [1.2,1.4]	1.285 [1.2,1.4]
Age	1.448 [1.3,1.7]	1.592 [1.3,2.0]	1.316 [1.1,1.6]	1.427 [1.2,1.6]	1.571 [1.2,2.0]	1.293 [1.1,1.6]
Ref. White						
Black	0.300 [0.2,0.4]	0.194 [0.1,0.3]	0.296 [0.2,0.5]	0.339 [0.3,0.4]	0.263 [0.2,0.4]	0.317 [0.2,0.5]
Other	0.636 [0.5,0.8]	0.525 [0.4,0.7]	0.556 [0.5,0.7]	0.666 [0.6,0.8]	0.664 [0.5,0.9]	0.541 [0.4,0.7]
Ref. Year 2011						
Year 2012	0.758 [0.6,0.9]	0.722 [0.5,1.0]	0.906 [0.7,1.2]	0.741 [0.6,0.9]	0.672 [0.5,0.9]	0.896 [0.7,1.2]
Observations	11183	2,281	8902	11183	2,281	8902

1DG experimenters: Ever tried a cigarette one or two puffs

2 DG ever smoked: smoked at least 100 cigarettes in lifetime

3 DG current smokers: 100+ cig and smoked at least one day during the past 30 days

4 Current e-cigarette users: used e-cigarettes at least one time during the past 30 days

All models applied sample weights; 95% confidence intervals in brackets

	Dep. Variable: DG ever smoked <sup>2</sup>			Dep. Variable: DG Current smoked <sup>3</sup>		
	Pool	Before 2010	2010 and after	Pool	Before 2010	2010 and after
Current e-cigarette users <sup>4</sup>	7.461 [5.6,10.0]	7.212 [5.1,10.3]	7.165 [4.9,10.6]	7.731 [5.8,10.3]	7.442 [5.3,10.4]	7.260 [4.9,10.7]
Male	1.343 [1.3,1.4]	1.349 [1.3,1.4]	1.119 [1.0,1.2]	1.334 [1.3,1.4]	1.340 [1.3,1.4]	1.117 [1.0,1.2]
Age	1.448 [1.3,1.7]	1.398 [1.2,1.6]	1.362 [1.0,1.8]	1.427 [1.2,1.6]	1.359 [1.2,1.6]	1.384 [1.0,1.9]
Ref. White						
Black	0.300 [0.2,0.4]	0.234 [0.2,0.3]	0.297 [0.2,0.5]	0.339 [0.3,0.4]	0.281 [0.2,0.4]	0.294 [0.2,0.5]
Other	0.636 [0.5,0.8]	0.492 [0.4,0.6]	0.743 [0.6,1.0]	0.666 [0.6,0.8]	0.536 [0.4,0.7]	0.741 [0.5,1.0]
Ref. Year 2011						
Year 2012	0.758 [0.6,0.9]	0.884 [0.7,1.1]	1.409 [1.0,2.0]	0.741 [0.6,0.9]	0.848 [0.7,1.1]	1.328 [0.9,1.9]
Observations	11183	6,040	5143	11183	6,040	5143

1 DG experimenters: Ever tried a cigarette one or two puffs

2 DG ever smoked: smoked at least 100 cigarettes in lifetime

3 DG current smokers: 100+ cig and smoked at least one day during the past 30 days

4 Current e-cigarette users: used e-cigarettes at least one time during the past 30 days

All models applied sample weights; 95% confidence intervals in brackets

## Reference List

1. E-Cigarettes May Promote Illicit Drug Use and Addiction. *Columbia University Medical Centre* **2014**. <http://newsroom.cumc.columbia.edu/blog/2014/09/03/e-cigarettes-may-promote-illicit-drug-use-addiction/>
2. Phillips, C. V. CASAA response to new study that claims ecigs are a gateway to cocaine. *Anti-THR lies and related topics* **2013**. <http://antithrlies.com/2014/09/03/casaa-response-to-new-study-that-claims-ecigs-are-a-gateway-to-cocaine/>
3. Phillips, C. V.; Wang, C.; Guenzel, B. You might as well smoke; the misleading and harmful public message about smokeless tobacco. *BMC Public Health* **2005**, 5 (31).
4. E-cigarette sales in Spain drop by 70 per cent. *thinkSPAIN* **2014**. <http://www.thinkspain.com/news-spain/24345/e-cigarette-sales-in-spain-drop-by-70-per-cent>
5. Blackwell, T. Public health officials inadvertently bolstering tobacco market with crackdown on e-cigarettes, analysts say. *National Post*, **Aug 15, 2014**. <http://news.nationalpost.com/2014/08/15/public-health-officials-inadvertently-bolstering-tobacco-market-with-crackdown-on-e-cigarettes-analysts-say/>
6. Causation - Overdetermination And Preemption. *JRank: Science Encyclopedia*. <http://science.jrank.org/pages/8555/Causation-OVERDETERMINATION-PREEMPTION.html>
7. Dawid, A. P. Influence Diagrams for Causal Modelling and Inference. *International Statistical Review* **2002**, 70 (2), 161-189.
8. Haddock CK; Weg MV; DeBon M; Klesges RC; Talcott GW; Lando H; Peterson A Evidence that smokeless tobacco use is a gateway for smoking initiation in young adult males. *Preventive Medicine* **2001**, 32 (3), 262-267.
9. Dutra LM; Glantz SA Electronic Cigarettes and Conventional Cigarette Use Among US Adolescents: A Cross-sectional Study. *JAMA Pediatrics* **2014**, 168 (7), 610-617.
10. 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. *Journal of Adolescent Health* **2013**, 54 (6), 684-690.
11. Phillips, C. V. Working paper: Reports of an e-cigarette gateway to cigarettes by Glantz et al.: the study results provide no support for the conclusions. *EP-ology* **2014**. <http://ep-ology.com/2014/05/17/working-paper-reports-of-an-e->

[cigarette-gateway-to-cigarettes-by-glantz-et-al-the-study-results-provide-no-support-for-the-conclusions/](#)

12. Phillips, C. V. Dear Public Health: the public despises you, so you are probably doing it wrong. *Anti-THR lies and related topics* **2014**.  
<http://antithrlies.com/2014/09/21/dear-public-health-the-public-despises-you-so-you-are-probably-doing-it-wrong/>
13. Phillips, C. V. Properly Interpreting the Epidemiologic Evidence About the Health Effects of Industrial Wind Turbines on Nearby Residents. *Bulletin of Science, Technology & Society* **2011**, 31 (4).
14. Rodu, B.; Nasic, S.; Cole, P. Tobacco use among Swedish schoolchildren. *Tob. Control* **2005**, 14 (6), 405-408.
15. Lipsitch, M.; Tchetgen, E. T.; Cohen, T. Negative Controls: A Tool for Detecting Confounding and Bias in Observational Studies. *Epidemiology* **2010**, 21 (3), 383-388.
16. Phillips, C. V. Agent-based model of THR adoption (and basic case for THR from City Health 2012). *EP-ology* **2012**. <http://ep-ology.blogspot.ca/2012/10/agent-based-model-of-thr-adoption-and.html>
17. Timberlake, D. S.; Huh, J.; Lakon, C. M. Use of propensity score matching in evaluating smokeless tobacco as a gateway to smoking. *Nicotine Tob. Res.* **2009**, 11 (4), 455-462.
18. O'Connor, R. J.; Flaherty, B. P.; Quinio, E. B.; Kozlowski, L. T. Regular smokeless tobacco use is not a reliable predictor of smoking onset when psychosocial predictors are included in the model. *Nicotine Tob. Res.* **2003**, 5 (4), 535-543.
19. Phillips, C. V. Publication bias in situ. *BMC Medical Research Methodology* **2004**, 4 (20).
20. Heavner, K. K.; Phillips, C. V.; Burstyn, I.; Hare, W. Dichotomization:  $2 \times 2 (\times 2 \times 2 \times 2 \dots)$  categories: infinite possibilities. *BMC Medical Research Methodology* **2010**, 10 (59).
21. Fernandez, E. E-Cigarettes: Gateway to Nicotine Addiction for U.S. Teens, Says UCSF Study. *University of California San Francisco* **2014**.  
<http://www.ucsf.edu/news/2014/03/112316/e-cigarettes-gateway-nicotine-addiction-us-teens-says-ucsf-study>
22. Glantz, S. A. US kids who use e-cigs more likely to progress from experimentation to established smoking, less likely to quit, smoke more. *University of California: Centre for Tobacco Control Research and Education* **2014**. <https://tobacco.ucsf.edu/us-kids-who-use-e-cigs-more-likely-progress-experimentation-established-smoking-less-likely-quit-smo>

23. Glantz, S. A. FDA should not assume a "continuum of risk" until there is proof that it exists for ecigs as they are actually used. *University of California: Centre for Tobacco Control Research and Education* **2014**. [https://tobacco.ucsf.edu/fda-should-not-assume-continuum-risk-until-there-proof-it-exists-ecigs-they-are-actually-used#\\_ftnref20](https://tobacco.ucsf.edu/fda-should-not-assume-continuum-risk-until-there-proof-it-exists-ecigs-they-are-actually-used#_ftnref20)
24. Tavernise, S. Young Using E-Cigarettes Smoke Too, Study Finds. *The New York Times*, **Mar 6, 2014**. <http://www.nytimes.com/2014/03/07/health/young-users-of-e-cigarettes-less-likely-to-quit-smoking-study-finds.html>