NBER WORKING PAPER SERIES

E-CIGARETTES AND RESPIRATORY DISEASE: A REPLICATION, EXTENSION, AND FUTURE DIRECTIONS

Donald S. Kenkel
Alan D. Mathios
Hua Wang

Working Paper 27507
http://www.nber.org/papers/w27507

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2020

This working paper reports the authors' research conducted as Cornell faculty and academic staff. Cornell University, through its Office of Sponsored Programs, has received funding from the Foundation for a Smoke-Free World for a different project titled "An Economic Study of Risk Perceptions and Consumer Demand for Harm Reduction Products." Kenkel is the Principal Investigator and Wang is a research associate on that project. Kenkel has also served as a paid consultant for the Foundation for a Smoke-Free World. The Foundation for a Smoke-Free World is an independent, nonprofit organization committed to reducing deaths and diseases caused by smoking. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

E-Cigarettes and Respiratory Disease: A Replication, Extension, and Future Directions
Donald S. Kenkel, Alan D. Mathios, and Hua Wang
NBER Working Paper No. 27507
July 2020
JEL No. I12

ABSTRACT

Electronic cigarettes show potential to reduce the harms from smoking combustible tobacco, but there is uncertainty about the long-term health consequences. We replicate and extend the study by Bhatta and Glantz (20192), which reports longitudinal statistical associations between e-cigarette use and long-term respiratory disease. We are able to closely replicate their results. When we use a more flexible empirical specification, among respondents who had never smoked combustible tobacco, we find no evidence that current or former e-cigarette use is associated with respiratory disease. The statistical associations between e-cigarette use and respiratory disease are driven by e-cigarette users who are also current or former smokers of combustible tobacco. A striking feature of the data is that almost all e-cigarette users were either current or former smokers of combustible tobacco. We then discuss the potential for future applied econometric research to credibly identify the causal effects of e-cigarette use on health. Challenges include the potential selection biases that stem from the complex set of consumer choices to initiate and quit smoking combustible tobacco, use of e-cigarettes, and dual use of both products. We suggest using a variety of identification strategies to uncover the causal effects that use a variety of econometric methods.

Donald S. Kenkel
Department of Policy Analysis & Management
College of Human Ecology
Cornell University
Martha Van Rensselaer Hall
Ithaca, NY 14853-4401
and NBER
dsk10@cornell.edu

Alan D. Mathios
182 MVR Hall
Cornell University
Ithaca NY 14853
adm5@cornell.edu

Hua Wang
Department of Policy Analysis & Management
MVR Hall Cornell University
Ithaca NY 14853
hw227@cornell.edu

1. Introduction

After they came onto the U.S. market around 2007, the use of electronic cigarettes has steadily increased. In data from the 2018 U.S. National Health Interview, 3.2 percent of adults were current users of e-cigarettes (Villarroel, Cha, and Vahratian 2020). Current e-cigarette use was highest among former combustible cigarette smokers who had quit within the past year, at just over 25 percent, and among former smokers who had quit within the past one to four years, at 17 percent. Continuing a decades-long trend, the fraction of U.S. adults who smoke combustible cigarettes declined from 20 percent in 2007 to 14 percent in 2018 (Thorne, Malarcher, Maurice and Caraballo 2008, Creamer, Wang, Babb, et al. 2019).

To the extent the growth in e-cigarette use contributed to the decline in adult smoking, e-cigarettes might be a promising harm-reduction strategy. The consensus report by the National Academy of Sciences, Engineering, and Medicine (NASEM) concluded that: "There is *conclusive* evidence that completely substituting e-cigarettes for combustible tobacco cigarettes reduces users' exposure to numerous toxicants and carcinogens present in combustible tobacco cigarettes." (NASEM 2018).[1] However, the policy statements on e-cigarettes of the American College of Physicians (Crowley, 2015), American Heart Association (Bhatnagar et al., 2014), and American Lung Association (2015) emphasize uncertainty about the long-term health consequences of e-cigarette use.

In this paper we replicate and extend the analysis of a recent study by Bhatta and Glantz (2019, hereafter B & G) of the association between e-cigarette use and long-term respiratory disease. B & G analyzed observational data from the first three waves of the Population Assessment of Tobacco and Health (PATH) Study. Based on statistically significant longitudinal associations between former and current e-cigarette use and respiratory disease, B & G conclude that: "Use of e-cigarettes is an independent risk factor for respiratory disease in addition to combustible tobacco smoking." Major news media reported the results, including NBC (2019), Reuters (2019), and National Public Radio (NPR, 2019). For example, NPR reported that the study "found that people who used *only* e-cigarettes had about a 30% increased risk of developing lung disease, compared with people who didn't use any nicotine products." (NPR 2019, emphasis in the original). The accompanying press release (Alvarez 2019) and news media reports interpreted the estimated associations as showing that e-cigarettes are "harmful on their own" (Glantz, quoted in Alvarez 2019).

Shortly before the publication of the B & G study, in the late summer and fall of 2019 the U.S. experienced an outbreak of acute lung injury that the CDC termed e-cigarette, or vaping, associated lung injury (EVALI). As of the final CDC reporting date February 18, 2020, there were a total of 2,807 hospitalized EVALI cases or deaths reported nationwide, including 68 deaths (CDC 2020). The CDC eventually linked the EVALI outbreak to illegally manufactured

[1] Conclusion 18-1, italics in original to indicate that the evidence meets the standard for conclusive defined as: "There are many supportive findings from good-quality controlled studies (including randomized and non-randomized controlled trials) with no credible opposing findings. A firm conclusion can be made, and the limitations to the evidence, including chance, bias, and confounding factors, can be ruled out with reasonable confidence." *Conclusive* is the highest evidence standard defined in the NASEM report.

e-cigarettes that contain THC (the psychoactive component of marijuana) and vitamin E acetate; EVALI was not found to be linked to commercially available e-cigarettes that contain nicotine.[2] B & G and our replication examine the association between nicotine-containing e-cigarettes and long-term respiratory diseases such as chronic obstructive pulmonary disease (COPD). We do not study EVALI.

When we replicate the empirical analysis of B & G, we estimate statistical associations that are very similar to their published results. We extend the analysis by adopting a more flexible empirical specification. Our regression specification uses a set of eight indicator variables (and a reference category) based on the exhaustive set of nine possible combinations (3 by 3 design with never, current, or former smoker of cigarettes being interacted with never, current, or former user of e-cigarettes). Among respondents who had never smoked combustible tobacco, we find no evidence that current or former e-cigarette use is associated with respiratory disease. The statistical associations that B & G find between e-cigarette use and respiratory disease are driven by e-cigarette users who are also current or former smokers of combustible tobacco. Compared to only smoking combustible tobacco, dual use of e-cigarettes and combustible tobacco is associated with a about the same risk of respiratory disease.

A striking feature of the PATH data analyzed by B & G is that almost all e-cigarette users were either current or former smokers of combustible tobacco. In the longitudinal analysis sample with 17,601 observations, there were only 12 current e-cigarette users who had never smoked combustible tobacco. None of the 12 respondents had incident (new) respiratory disease. The number of respondents who only used e-cigarettes is simply not large enough to draw meaningful conclusions about the independent association between e-cigarette use and respiratory disease. More recent data sets will face similar limitations, although to a lesser extent. For example, in the 2018 National Health Interview Survey data the prevalence of current e-cigarette use among people who had never smoked cigarettes was 1.1 percent (Villarroel, Cha, and Vahratian 2020).

In section 2 we present our replication of the main results of B & G. In section 3 we present the results of our more flexible specification. Section 4 uses the framework of applied econometrics to discuss future directions to address the empirically challenging problem of identifying the causal effects of e-cigarette use on health outcomes. Section 5 provides a brief conclusion.

2. Replication of the Main Results of Bhatta and Glantz

In our replication exercise we try to use the same data, variable definitions, and empirical methods used by B & G. The data are from waves 1 -3 of the PATH Study and span the years 2013-2016. B & G conduct two sets of analyses; the first set examines cross-sectional associations at wave 1, and the second set examines longitudinal associations between wave 1 and waves 2 or 3. B & G emphasize the longitudinal analysis; as discussed in section 4 below, in public health research longitudinal associations are often seen as more compelling evidence about causality.

[2] See Dave et al. (2020) for more discussion of the EVALI outbreak and CDC announcements.

B & G report a wave 1 sample size of 32,320 respondents (their Table 3), but in the text note that 5.3 percent of the sample was dropped due to missing data. This leaves an apparent sample size of 30,607 observations for the cross-sectional analysis. We find that the full sample for the wave 1 adult file in PATH is 32,320 respondents, but when we drop observations with missing data our sample size for the cross-sectional analysis is 28,717.

In the cross-sectional analysis, the dependent variable that measures respiratory disease takes a value of one for respondents who self-report at wave 1 that they had ever been told that they had COPD, chronic bronchitis, emphysema, or asthma. B & G report that by this measure the prevalence of lung disease in wave 1 was 15.1 percent; in the Cornell sample the prevalence is also 15.1 percent.

In terms of independent variables, B & G define four indicator variables to describe wave 1 tobacco product use – former e-cigarette use, current e-cigarette use, former smoker of combustible tobacco, and current smoker of combustible tobacco. The reference category consists of respondents who never used either e-cigarettes or combustible tobacco. Combustible tobacco includes cigarettes, cigars, cigarillos, pipe tobacco, and hookahs. Additional control variables include wave 1 demographics and self-reported health conditions. Demographics measure age, BMI, sex, race/ethnicity, and poverty level. The health condition variables indicate if the respondent self-reports being told that they had high blood pressure, high cholesterol, or diabetes. As can be seen in Table 1, the means of the independent variables reported by B & G are very close to the means in the Cornell sample.

The second set of analyses examine longitudinal associations between incident (new) respiratory disease at wave 2 or wave 3 and e-cigarette use and smoking combustible tobacco at wave 1. The sample for this analysis excludes respondents who reported respiratory disease at wave 1. After dropping these respondents and accounting for attrition between waves, the sample size is 19,475 as reported by B & G (their Table 3); after also dropping observations with missing information our sample size for the longitudinal analysis is 17,601. The rate of incident (new) respiratory disease between wave 1 and wave 2 or 3 is 4.5 percent in the Cornell analysis sample; in B & G Appendix Table 2 the weighted sample mean for this variable is also 4.5 percent.

Following B & G, we estimate logistic regressions and report adjusted odd ratios (AORs). We also present the results from linear models of the probability of respiratory illness as a function of e-cigarette use, combustible smoking, and the other independent variables. The AORs from the logistic regressions can be difficult to interpret because they are ratios of ratios. The coefficients from the linear models are directly interpretable as showing the difference in the probability of respiratory illness associated with a unit change in the independent variable. The linear probability model is widely used in applied econometrics (Angrist and Pischke 2009). Although the linearity assumption cannot be valid throughout the range of the dependent variable, in most cases linearity is a reasonable approximation and yields estimated marginal effects that very similar to the marginal effects implied by non-linear models such as probit and logit. We followed B & G and use the PATH sample weights. We used Stata to estimate all models; B & G used R.

Table 2 presents our replication of the main results of the B & G study. Our models control for the other independent variables included in the B & G models. Our results very closely replicate the estimated AORs and associated 95 percent confidence intervals. In the cross-sectional analysis, the linear probability model results show that the estimated AORs around 1.3 for former and current e-cigarette use correspond to a four-percentage point increase in the probability of respiratory disease. In the longitudinal analysis, the linear probability model results show that the estimated AORs around 1.3 correspond to a one-percentage point increase in the probability of incident respiratory disease.

3. Extension to a More Flexible Empirical Specification

In this section we report results from models that extend the B & G analysis by adopting a more flexible empirical specification. The B & G specification uses four estimated parameters (and a reference category) to model the relationships between respiratory disease and e-cigarette use and smoking combustible tobacco. Our more flexible specification uses eight estimated parameters (and the same reference category – never used e-cigarettes or smoked combustible tobacco). We create eight indicator variables for each of the following categories: former e-cigarette user/never combustible smoker, current e-cigarette user/never combustible smoker, former combustible smoker/ never e-cigarette user, current combustible smoker/ never e-cigarette user/, former e-cigarette user/former combustible smoker, former e-cigarette user/current combustible smoker, current e-cigarette user/former combustible smoker, and current e-cigarette user/current combustible smoker.

The added flexibility allows our specification to isolate the relationship between respiratory disease and e-cigarette use among those respondents who only use e-cigarettes. Our specification includes all possible interaction terms between the tobacco use indicator variables in the B & G specification. For example, our specification allows the relationship between respiratory disease and current e-cigarette use to be different between respondents who were never, former, and current smokers of combustible products.

The results of the more flexible specification are reported in Table 3. We find no statistically or practically significant positive associations between respiratory disease and e-cigarette use. In the cross-sectional sample the estimated AORs are close to one for the former e-cigarette user/never combustible smoker category and the current e-cigarette user/never combustible smoker category. In the longitudinal sample, those who were former e-cigarette users and who had never smoked combustibles had reduced odds of respiratory disease, i.e. the AOR is less than one, but the AOR is not statistically significantly different from one. The statistical software (Stata) dropped observations of the current e-cigarette users/never combustible smokers from the model because the indicator for this category perfectly predicts the outcome – all 12 respondents in this category did not have incident respiratory diseases. In three of the four linear probability models reported in Table 3, the estimated coefficients are small and not statistically significant different from zero. In the remaining linear model, current e-cigarette users who never smoked combustible tobacco have a statistically significant lower probability of respiratory disease. We do not see this as an informative negative association because the negative coefficient reflects the outcomes of 12 respondents.

Table 3 highlights a striking feature of the PATH data – there are very few respondents who never smoked combustible tobacco and who are former or current users of e-cigarettes. In the cross-sectional sample of 28,717 observations, of those who never smoked combustible tobacco there are only 87 former e-cigarette users and 19 current e-cigarette users. In the longitudinal sample, there are only 51 such former e-cigarettes users and 12 such current e-cigarette users. The small numbers of respondents who only use or used e-cigarettes result in large confidence intervals around the estimated AORs. The small numbers in the PATH data reflect wave 1 e-cigarette use patterns, i.e. current and past use measured in 2013. Given the upward trend in e-cigarette use over time, more recent data sets will include higher proportions of respondents who never smoked combustible products and who are former or current e-cigarette users. In the most recent data on adolescents, e-cigarette use is more common than smoking cigarettes (Wang et al. 2019). Future data might eventually provide sufficient statistical power to estimate health outcomes associated with the use of e-cigarettes only. However, the numbers of such respondents in the PATH data are simply not large enough to draw meaningful conclusions about the independent association between respiratory disease and e-cigarette use.

In addition to the independent association between respiratory disease and e-cigarette use, B & G also analyze the dual use of e-cigarettes and combustible tobacco products. They interpret their results as showing increased risk from dual use; in the press release Glantz is quoted as stating that dual users "are actually worse off than tobacco smokers" (Alvarez 2019). In the longitudinal sample, the estimated AORs from our specification provide some support for this claim. For current combustible smokers who never used e-cigarettes, the estimated AOR is 2.84, compared to 3.07 for current smokers who were former e-cigarette users and 3.23 for current smokers who were current e-cigarette users. However, as can be seen in Figure 1, the 95 percent confidence intervals around these point estimates substantially overlap. It should also be kept in mind that the rate of incident respiratory disease between wave 1 and wave 2 or 3 is low, at 4.5 percent. The differences in the estimated AORs thus reflect very small differences in the absolute levels of the probability of incident respiratory disease. In the results from the linear probability models, the estimated coefficients (rounded to two digits after the decimal point) are almost exactly the same for current smokers of combustible regardless of whether they are never, former, or current e-cigarette users. On balance, we do not see the results in Table 3 as providing strong evidence that dual use of e-cigarettes and combustible tobacco is associated with a higher risk of respiratory disease than smoking combustibles alone.

4. Future Directions: Identifying the Causal Effects of E-Cigarettes on Health

Private decisions about e-cigarette use and public policy decisions about e-cigarette regulation require evidence on the causal effects of e-cigarette use on health outcomes. In contrast, the study by B & G and our replication and extension are examples of risk-factor epidemiology and estimate statistical associations, not causal effects. Identifying statistical risk factors is a potentially useful step in public health research. For example, after public health research initially identified factors including social class, geography, and ethnicity as risk factors for neural tube birth defects, subsequent animal studies and clinical trials established periconceptional folic acid deficiency to be the causal mechanism behind the statistical

associations (Susser and Susser 1996). However, it may be challenging to follow up risk factor studies of e-cigarette use with research that will establish convincing evidence on causal health effects, or the lack thereof.[3]

We believe that the framework of applied econometrics provides a useful framework to address the challenges of using observational data to estimate the causal effects of e-cigarette use on health. The key challenge is that in observational data, people are not randomly assigned into e-cigarette use. Instead, they self-select into e-cigarette use. The high levels of e-cigarette use among adult former smokers suggest that many adults self-select into e-cigarette use as a way to stop smoking combustible tobacco. Because health concerns often drive smoking cessation, this creates potential reverse causality if health problems cause people to select into e-cigarette use.

Public health research often uses longitudinal data to address reverse causality and other sources of selection bias. To explore reverse causality, B & G (their Table 4) estimate the longitudinal associations between e-cigarette use at wave 2 or 3 and respiratory disease at wave 1. For this analysis, the sample is limited to people who at wave 1 had never used e-cigarettes. B & G find evidence consistent with reverse causality; wave 1 respiratory disease is associated with statistically significantly higher odds of e-cigarette use at wave 2 or 3 (AOR = 1.44). As explained above, in their main longitudinal analysis of incident respiratory disease, B & G drop individuals with wave 1 respiratory disease. They argue that: "The longitudinal design allows much stronger conclusions about causality than in earlier cross-sectional studies…."

However, the use of longitudinal data is not sufficient to eliminate self-selection bias in observational studies. Remaining unobservable heterogeneity, also called unobservable confounders, can still result in biased estimates of causal effects. For example, B & G were only able to drop respondents who reported that they had ever been told by a health professional that they had a chronic respiratory disease. Some respondents with diagnosed respiratory disease might not recall being told.[4] Other respondents might never have been told by a health professional that they had respiratory disease, yet still experienced respiratory symptoms. Such respondents are likely to self-select into e-cigarette use as a way to quit smoking combustible tobacco. Because they are more likely to report respiratory disease in wave 2 or 3 due to their

[3] Laboratory research establishes that the vapor from e-cigarettes contains fewer and much lower levels of the combustion-generated toxicants in tobacco smoke that are linked to cancer, chronic lung disease, and cardiovascular disease (Hajek et al 2014). Yet the gold standard evidence from human clinical trials is lacking because of ethical concerns about randomly assigning subjects to use e-cigarettes. Future research on the health effects of e-cigarette use will probably continue to rely on observational data.

[4] Although we do not observe PATH respondents before wave 1, we can compare self-reports of disease at wave 1 and wave 2. For example, of the 1002 respondents with reported chronic bronchitis in wave 1, 398 of them (forty percent) do not report in wave 2 having been told within the past 12 months that they had chronic bronchitis. By definition, chronic bronchitis is a long-term respiratory disease that should be expected to persist across waves. The 398 respondents who no longer report chronic bronchitis in wave 2 might not have discussed their disease with a health professional within the past year, or they might have forgotten the diagnosis. Either way, the comparison of wave 1 and wave 2 self-reports suggests that using the wave 2 measure would result in substantial misclassification of which respondents have respiratory disease. If there are also substantial rates of misclassification in the wave 1-based measures, using measures of respiratory disease at wave 1 will not control for important sources of reverse causality.

pre-existing symptoms, longitudinal associations between newly reported respiratory disease and e-cigarette use reflects selection bias as well as any true causal effect.

Applied econometrics provides a systematic framework for future research on the causal health effects of e-cigarette use. A productive line of applied econometrics emphasizes the importance of finding quasi-experimental variation that credibly identifies the causal effect of an independent or treatment variable on the outcome of interest. Possible examples include plausibly exogeneous supply-side shocks or public policies that increase the cost of obtaining e-cigarettes. The resulting variation in e-cigarette use provides a quasi-experiment where some people are more-or-less randomly assigned (by the exogenous shifters) to different levels of e-cigarette use. Quasi-experimental methods in econometrics to identify causal effects include event studies, differences-in-differences, instrumental variables, and the regression discontinuity design (Angrist and Pischke 2009).

To date, empirical economic research on the determinants of e-cigarette use is thin (DeCicca, Kenkel, and Lovenheim 2020). As a result, it is not clear whether within the U.S. there are strong supply-side shocks or enough variation in public policies to provide the quasi-experimental variation required to identify causal health effects. Across countries, public policies towards e-cigarettes range from complete bans, to the regulatory approach in the U.S., to Britain's approach which actively promotes e-cigarettes for smoking cessation. A careful quasi-experimental approach to cross-country data is a potential strategy to study causal health effects, keeping in mind the threats to external validity that might make it inappropriate to extrapolate across countries.

Another key feature of the applied econometrics framework is that it is guided by economic theory. Quasi-experimental studies are broadly guided by the insights from the economic approach to human behavior, while structural studies estimate econometric models that are tightly linked to well-specified theoretical models. Darden (2017) estimates a structural model of lifetime cigarette consumption. In each period, the smoker is assumed to decide about her current smoking to maximize lifetime utility, given her information at that point in time. Medical exams provide the consumer with new information about future health risks. Consistent with the model, the empirical analysis finds that major health shocks – the onset of cancer or heart disease – and recent changes in biomarkers, e.g. blood pressure and cholesterol levels, significantly increase the probability of quitting. Darden, Gilleskie and Strumpf (2018) use the same data to estimate the causal effect of smoking on life expectancy.

Regardless of whether future econometric research adopts the structural approach, the studies by Darden (2017) and Darden, Gilleskie and Strumpf (2018) provide a starting point that highlights the potentially complex pattern of lifetime choices of e-cigarette use and combustible tobacco smoking. Given the complexity, a productive strategy might be to conduct separate analyses of initiation and cessation (DeCicca, Kenkel, and Mathios 2008). Teens self-select into initiating e-cigarette use for much different reasons than adult smokers who select into e-cigarette use as a way to quit smoking combustible tobacco.

To sum up, future applied econometric research into the health effects of e-cigarettes could usefully use quasi-experimental research designs, perhaps combined with more structural econometric analysis. Instead of econometric analysis of observational data, another approach would be to conduct randomized clinical trials (RCTs). While longitudinal RCT's can be very helpful in identifying causality, they often raise significant ethical concerns, especially in the area of tobacco use. Moreover, RCT's have their own drawbacks that relate to self-selection of who participates, attrition within longitudinal trials, and other related issues. RCTs are also expensive to implement. We suggest using a variety of identification strategies to uncover the causal effects that use a variety of econometric methods for identification such as instrumental variables, difference-in-difference models, and regression discontinuity analysis. Taken together, these approaches can help assess the causal relationships within this complex environment.

5. Conclusion

Harm reduction is an established part of U.S. public health policy, for example to reduce the harms from substance abuse and sexually transmitted disease. The development of modern e-cigarettes has renewed interest in tobacco harm reduction. In his proposal for safer cigarettes published in the British Medical Journal, Russell (1976) observed that, "People smoke for the nicotine but die from the tar." Tar is the by-product of combustion, so Russell's observation explains the harm-reduction potential of non-combusted tobacco products including smokeless tobacco, e-cigarettes, and heat-not-burn products. After reviewing scientific evidence submitted by Swedish Match USA Inc., in 2019 the U.S. Food and Drug Administration authorized Swedish Match to market eight snus smokeless tobacco products through the modified risk tobacco product pathway. The manufacturer is allowed to market the products with the claim "Using General Snus instead of cigarettes puts you at a lower risk of mouth cancer, heart disease, lung cancer, stroke, emphysema, and chronic bronchitis."

Of course, the harm-reduction potential of e-cigarettes depend on whether use actually reduces harm. B & G report longitudinal statistical associations between e-cigarette use and respiratory disease. We are able to closely replicate their results, but when we use a more flexible empirical specification, we find no evidence that current or former e-cigarette use is associated with respiratory disease. Replication is a vital part of the scientific method. Our study adds to a body of research that is building the evidence base for e-cigarette regulatory policy.

Looking to the future, we believe that a promising approach is to use the methods of modern applied econometrics to credibly identify the causal effects of e-cigarette use on health. The original B & G study and our replication and extension are examples of risk-factor epidemiology. Even when epidemiologic studies find robust evidence for risk factors, the statistical associations should not be interpreted as evidence of causation. Unfortunately, this distinction is often obscured in media reports and sometimes in the epidemiology studies. By the same token, the lack of statistical associations in our extension should be cautiously interpreted. The key empirical challenge is that consumers are not randomly assigned but instead self-select into e-cigarette use. If consumers self-select into e-cigarettes as a way to quit smoking, the epidemiologic estimates are likely biased away from zero towards finding spurious associations

with health problems. To provide more definitive evidence, we suggest using a variety of identification strategies to uncover the causal effects that use of a variety of econometric methods for identification such as instrumental variables, difference-in-difference models, and regression discontinuity analysis.

References

Alvarez, Jason (2019). "E-Cigarettes Significantly Raise Risk of Chronic Lung Disease, First Long-Term Study Finds." University of California, San Francisco. https://www.ucsf.edu/news/2019/12/416216/e-cigarettes-significantly-raise-risk-chronic-lung-disease-first-long-term

American Lung Association (2015). American Lung Association statement on e-cigarettes. Retrieved from http.//www.lung.org/stop-smoking/tobacco-control-advocacy/federal/e-cigarettes.html.

Angrist, Joshua D., and Jörn-Steffen Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.

Bhatnagar, A., Whitsel, L. P., Ribisl, K. M., Bullen, C., Chaloupka, F., Piano, R. M. R., & Benowitz, N. (2014). Electronic cigarettes: A policy statement from the American Heart Association. *Circulation*, 130, 1418–1436.

Bhatta, DN & SA Glantz (2019). "Association of E-Cigarette Use With Respiratory Disease Among Adults: A Longitudinal Analysis." *American Journal of Preventive Medicine*.

CDC (2020). "Outbreak of Lung Injury Associated with the Use of E-Cigarette, or Vaping, Products." https://www.cdc.gov/tobacco/basic_information/e-cigarettes/severe-lung-disease.html

Crowley, R. A.; Health Public Policy Committee of the American College of Physicians. (2015). Electronic nicotine delivery systems: Executive summary of a policy position paper from the American college of physicians. *Annals of Internal Medicine*, 162, 583–584.

Creamer MR, Wang TW, Babb S, et al. (2019) "Tobacco Product Use and Cessation Indicators Among Adults — United States, 2018." *Morbidity and Mortality Weekly Report* 68:1013–1019. DOI: http://dx.doi.org/10.15585/mmwr.mm6845a2

Darden, Michael (2017). "Smoking, Expectations, and Health: A Dynamic Stochastic Model of Lifetime Smoking Behavior." *Journal of Political Economy* 125 (5): 1465–1522.

Darden, Michael, Donna B. Gilleskie, and Koleman Strumpf (2018). "Smoking and Mortality: New Evidence from a Long Panel." *International Economic Review* 59 (3): 1571-619.

Dave, Dhaval, Daniel Dench, Donald Kenkel, Alan Mathios, and Hua Wang (2020). "News that Takes Your Breath Away: Risk Perceptions During an Outbreak of Vaping-related Lung Injuries." *Journal of Risk & Uncertainty*, forthcoming.

DeCicca, Philip, Don Kenkel, and Michael Lovenheim (2020). "The Economics of Tobacco Regulation: A Comprehensive Review." NBER Working Paper 26923.

DeCicca, Philip, Don Kenkel, and Alan Mathios (2008). "Cigarette Taxes and the Transition from Youth to Adult Smoking: Smoking Initiation, Cessation, and Participation." *Journal of Health Economics* 27 (4): 904–17.

Hajek, P., Etter, J.-F., Benowitz, N., Eisenberg, T., & McRobbie, H. (2014). Electronic cigarettes: Review of use, content, safety, effects on smokers and potential for harm and benefit. *Addiction* 109: 1801–1810.

National Academies of Sciences, Engineering, and Medicine. 2018. *Public Health Consequences of E-Cigarettes*. Washington, DC: The National Academies Press. ttps://doi.org/10.17226/24952.

NBC (2019). "E-cigarettes Linked to Lung Problems, First Long-term Study on Vaping Finds." https://www.nbcnews.com/health/vaping/e-cigarettes-linked-lung-problems-first-long-term-study-vaping-n1101641

NPR (2019). "Vaping Nicotine Linked to Increased Risk of Chronic Lung Disease." https://www.npr.org/sections/health-shots/2019/12/16/788540191/vaping-nicotine-linked-to-increased-risk-of-chronic-lung-disease

Reuters (2019). "Vaping Increases the Risk of Lung Disease by a Third: U.S. Study Finds." https://www.reuters.com/article/us-health-vaping-lungs-idUSKBN1YK260

Russell, M. A. H. (1976). "Low-tar medium-nicotine cigarettes: A new approach to safer smoking." *British Medical Journal*, 1, 1430–1433.

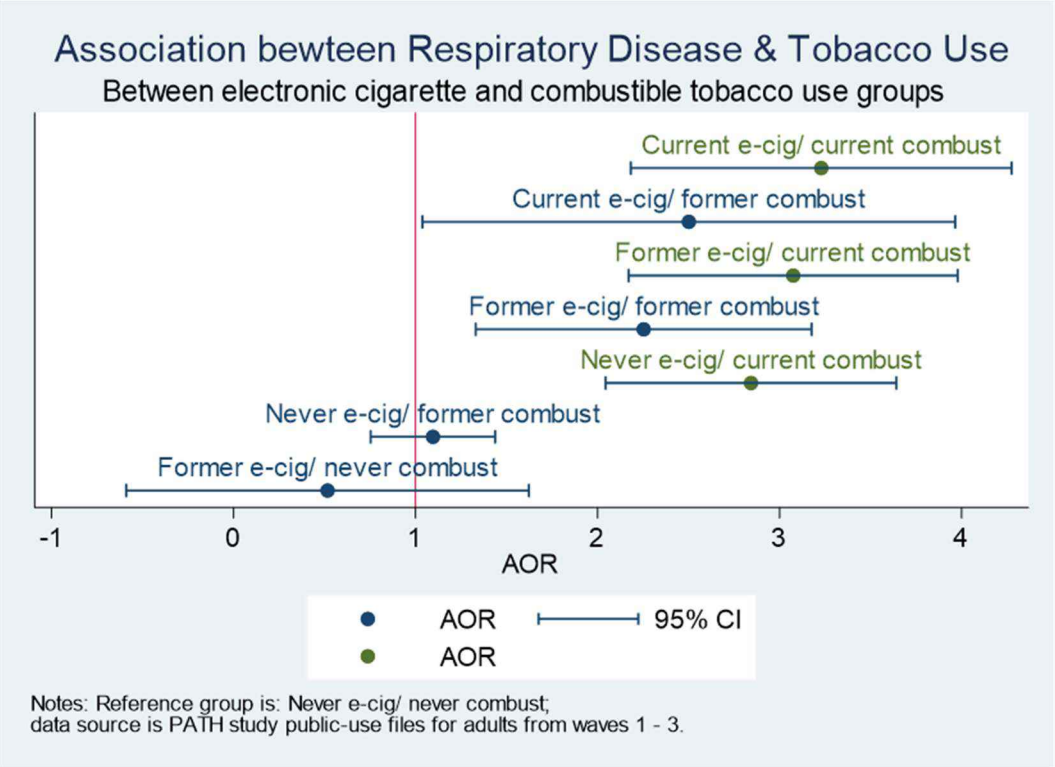
Susser, Mervyn and Ezra Susser (1996). "Choosing a Future for Epidemiology: I, Eras and Paradigms." *American Journal of Public Health* 86 (5): 668-673.

Thorne, SL, A Malarcher, E Maurice, R Caraballo (2008). "Cigarette Smoking Among Adults --- United States, 2007." *Morbidity and Mortality Weekly Report* 57(45);1221-1226.

Villarroel MA, Cha AE, Vahratian A. (2020). "Electronic cigarette use among U.S. adults, 2018." NCHS Data Brief, no 365. Hyattsville, MD: National Center for Health Statistics.

Wang TW, Gentzke AS, Creamer MR, et al. (2019). "Tobacco Product Use and Associated Factors Among Middle and High School Students —United States, 2019." MMWR Surveill Summ 68(No. SS-12):1–22.

Figure 1. Longitudinal Associations between Wave 1 and Wave 2 or 3N



As noted in the text, the logistic model does not provide an estimate of the AOR for the category current e-cig/never combust. None of the respondents in this category had an incident respiratory condition.

Table 1: Summary Statistics

	Mean B & G	Mean Cornell replication
Respiratory disease at Wave 1	0.151	0.151
New respiratory disease at Wave 2 or 3	0.045	0.0448
E-cigarette user		
Never	0.823	0.817
Former	0.122	0.126
Current	0.055	0.0573
Combustible tobacco smoker		
Never	0.286	0.275
Former	0.454	0.458
Current	0.260	0.267
E-cigarette and Combustible tobacco use		
Never e-cig user/never comb smoker		0.273
Former e-cig user/never comb smoker		0.00143
Current e-cig user/never comb smoker		0.000274
Former comb smoker/never e-cig user		0.416
Current comb smoker/ never e-cig user		0.128
Former comb smoker/former e-cig user		0.0343
Current comb smoker/former e-cig user		0.0899
Current e-cig/ former comb smoker		0.00778
Current e-cig/current comb smoker		0.0492
Age group		
18 to 24 years old	0.131	0.129
25 to 34 years old	0.177	0.186
35 to 44 years old	0.165	0.172
45 to 54 years old	0179	0.181
55 to 64 years old	0.166	0.165
65 to 74 years old	0.111	0.105
75 years old or older	0.071	0.0620
Female	0.481	0.510
Race		
White	0.779	0.779
Black	0.123	0.124
Other	0.098	0.0967
At or above poverty	0.748	0.752
BMI	0.280	28.11
High blood pressure	0.278	0.275
High cholesterol	0.230	0.228
Diabetes mellitus	0.140	0.141
Observations	32,320	28,717

Table 2. Replications

	Cross-Sectional Associations at Wave 1			Longitudinal Associations between Wave 1 and Wave 2 or 3		
Variables	AOR (95% CI) **B & G**	AOR (95% CI) **Cornell replication**	Coeff. (95% CI) **Cornell replication LPM**	AOR (95% CI) **B & G**	AOR (95% CI) **Cornell replication**	Coeff. (95% CI) **Cornell replication LPM**
E-cigarette user						
Never	Ref	Ref	Ref	Ref	Ref	Ref
Former	1.34***	1.36***	0.04***	1.31**	1.28*	0.01*
	[1.23,1.46]	[1.24,1.49]	[0.03,0.05]	[1.07,1.60]	[1.05,1.56]	[0.00,0.02]
Current	1.32***	1.34***	0.04***	1.29*	1.30*	0.01
	[1.17,1.49]	[1.20,1.50]	[0.02,0.06]	[1.03,1.61]	[1.03,1.65]	[-0.00,0.03]
Combustible tobacco smoker						
Never	Ref	Ref	Ref	Ref	Ref	Ref
Former	1.29***	1.31***	0.03***	1.16	1.17	0.01
	[1.14,1.47]	[1.14,1.50]	[0.01,0.04]	[0.87,1.57]	[0.86,1.57]	[-0.00,0.02]
Current	1.61***	1.64***	0.06***	2.56***	2.62***	0.04***
	[1.42,1.82]	[1.43,1.87]	[0.04,0.07]	[1.92,3.41]	[1.98,3.46]	[0.03,0.05]
Sample Size	32,320	28,717	28,717	19,475	17,601	17,601

LPM = linear probability model.

Table 3. Extensions

	Cross-Sectional Associations at Wave 1			Longitudinal Associations between Wave 1 and Wave 2 or 3		
Variables	AOR (95% CI)	Coeff. (95% CI) LPM	N	AOR (95% CI)	Coeff. (95% CI) LPM	N
E-cigarette user (never combustible tobacco smoker)						
Never	Ref	Ref	4,128	Ref	Ref	2,705
Former	0.94	-0.01	87	0.52	-0.01	51
	[0.45,1.99]	[-0.09,0.07]		[0.06,4.40]	[-0.06,0.03]	
Current	1.04	0.00	19		-0.03***	12
	[0.29,3.69]	[-0.12,0.12]			[-0.04,-0.01]	
Combustible tobacco smoker (never e-cigarette user)						
Never	Ref	Ref	4,128	Ref	Ref	2,705
Former	1.30***	0.03***	7,536	1.10	0.00	4,830
	[1.13,1.50]	[0.01,0.04]		[0.80,1.50]	[-0.01,0.01]	
Current	1.64***	0.06***	6,655	2.84***	0.05***	4,001
	[1.43,1.88]	[0.04,0.07]		[2.15,3.77]	[0.04,0.06]	
Dual User						
Former e-cig/ former combust	1.83***	0.07***	1,800	2.25***	0.03***	1,089
	[1.53,2.20]	[0.05,0.10]		[1.50,3.39]	[0.01,0.05]	
Former e-cig/ current combust	2.21***	0.10***	5,214	3.07***	0.05***	3,020
	[1.92,2.54]	[0.08,0.12]		[2.29,4.13]	[0.04,0.06]	
Current e-cig/ former combust	1.70***	0.06**	413	2.50**	0.04*	256
	[1.26,2.29]	[0.02,0.10]		[1.39,4.49]	[0.01,0.07]	
Current e-cig/ current combust	2.21***	0.10***	2,865	3.23***	0.05***	1,637
	[1.89,2.58]	[0.08,0.12]		[2.34,4.46]	[0.04,0.07]	
Sample Size	28,717	28,717	28,717	17,589	17,601	17,601

LPM = linear probability model