*J Natl Cancer Inst Monogr* (2022) 2022(59): lgac007

https://doi.org/10.1093/jncimonographs/lgac007
Monograph

MONOGRAPH

# How Data Security Concerns Can Hinder Natural Experiment Research: Background and Potential Solutions

Michael F. Pesko, PhD [1,2,*]

[1]Department of Economics, Andrew Young School of Policy Studies, Georgia State University, Atlanta, GA, USA and [2]Institute of Labor Economics, Bonn, Germany

*Correspondence to: Michael F. Pesko, PhD, Department of Economics, Andrew Young School of Policy Studies, Georgia State University, 55 Park Place, Suite 657, Atlanta, GA 30303, USA (e-mail: mpesko@gsu.edu).

## Abstract

Health economists conducting cancer-related research often use geocoded data to analyze natural experiments generated by policy changes. These natural experiments can provide causal interpretation under certain conditions. Despite public health benefit of this rigorous natural experiment methodology, data providers are often reluctant to provide geocoded data because of confidentiality concerns. This paper provides an example of the value of natural experiments from e-cigarette research and shows how this research was hindered by security concerns. Although the tension between data access and security will not be resolved overnight, this paper offers 3 recommendations: 1) provide public access to aggregated data at area levels (eg, state) where possible; 2) approve projects with enough time to allow for publication in journals with lengthy peer-review times; and 3) improve communication and transparency between data providers and the research community. The Foundations for Evidence Based Policymaking Act of 2018 also presents a unique opportunity for improving the ability of researchers to use geocoded data for natural experiment research without compromising data security.

## Introduction

Natural experiments can provide important insights for cancer-related research. Natural experiments rely on variation in treatment exposure that may be unrelated to other factors associated with the outcomes to mimic the randomization used in randomized control trials to determine causal effects (1). One of the first known natural experiments was John Snow's analysis of the effect of disabling a pump containing contaminated water in London in 1854. He found that cases of cholera fell for people using this pump compared with the control group of individuals using other pumps (2). The 2021 Noble Prize in Economics was awarded to 3 economists who "pointed out cause and effect can be drawn from natural experiments. Their research has substantially improved our ability to answer key causal questions, which has been of great benefit to society" (3).

Natural experiments are regularly used for cancer-related research. For example, several recent papers use state-identifying information from the Behavioral Risk Factor Surveillance System to study effects of the timing in which states expanded Medicaid (if at all) on outcomes of insurance coverage, access to care, preventive care use (including cancer screenings), health behaviors (including tobacco use), and self-reported health for a variety of newly eligible Medicaid populations (4,5). Another recent paper uses Behavioral Risk Factor Surveillance System and National Health Interview Survey (NHIS) data to study the effect of disenrolling people from Medicaid in Tennessee on mammography and breast cancer exams, among other outcomes (6), and still other papers study the effect of expanding Medicaid on prescription fills for breast cancer hormonal therapies (7) and smoking cessation medications (8) and on cancer mortality directly (9).

**Received:** September 28, 2021; **Accepted:** February 24, 2022

All of these natural experiments require the use of location of residence information (eg, geocodes) to identify individuals affected by the policy change. A concern is that geographical information could be combined with demographic information to identify specific individuals and result in the loss of subject confidentiality. So, a tension exists between restricting geographical information and, at the same time, giving society as a whole the benefit of research capable of detecting causal effects to understand, for example, key relationships and how well (or not) costly government programs and regulations are working (10,11).

This paper first makes a case for the value that natural experiments have in overcoming confounding and returning causal estimates; in particular, drawing an example from e-cigarette research. Second, it discusses potential solutions that would help improve researchers' ability to conduct natural experiments without compromising data security.

## Relevance to Cancer Health Economic Research

According to the National Cancer Institute, cancer health economic research is the application of health economics theory and models to cancer prevention and screening, diagnosis, treatment, survivorship, and end-of-life care (12). Natural experiment-style research is regularly used to evaluate policies designed to improve cancer prevention and treatment (4-9), so safeguarding and improving researchers' abilities to evaluate natural experiments is an important avenue to expand cancer health economic research. Specifically related to modeling cancer prevention, which is one of the tenants of cancer health economic research, this paper provides an example from tobacco research. Tobacco research is an important area of research on cancer prevention because, for example, cigarette smoking is estimated to cause 29% of all cancer deaths in the United States (13). Therefore, improving the ability of modeling to understand factors that affect cigarette use is an example of cancer health economic research on prevention.

## Example of Using Natural Experiments to Overcome Confounding

The Food and Drug Administration has expressed an interest in understanding "e-cigarettes initiation [and] transitions to other tobacco products" (14). A longitudinal cohort study design in which people's tobacco product use is tracked over time is one approach for examining the relationship between e-cigarette use in one period and cigarette use in another period (eg, transitions). A generalized version of this empirical approach appears in the following equation:

$$cig\ use_{is(t+1)} = \propto + ecig\ use_{ist}\beta_1 + X_{ist} + \pi_s + \omega_t + u_{ist}, \quad (1)$$

where $i$ is for individual, $t$ is for time, and $s$ is for state.[1] $X_{ist}$ is a vector of time-varying demographic and policy variables. Fixed effects for state and time are included in the model; $u_{ist}$ is the error term.

In this equation, $\beta_1$ is capturing 3 things:

1) The causal effect of e-cigarette use on subsequent cigarette use: If e-cigarettes are "a 1-way street to traditional smoking" (15), this would clearly be problematic from a public health perspective. If, however, e-cigarettes are preventing initiation into combustible tobacco use and/or are causing more smoking cessation than smoking initiation, this would be evidence of a public health benefit of e-cigarettes.
2) Time-invariant confounding across individuals: These are variables specific to individuals that do not vary over time and could be correlated with both e-cigarette use and subsequent cigarette use but are not controlled for in the regression model. Examples of this include genetics and individual preferences to test the waters with less risky products (e-cigarettes) before transitioning to a possible a priori preferred choice of cigarettes.[2]
3) Time-varying confounding within individuals: These are variables specific to individuals that vary over time and are not controlled for in the regression model. This could include stressful life events (eg, death or divorce), learning new information about the harms of tobacco products, and so forth.

If item 2 and item 3 are large, $\beta_1$ may not be a close approximation of the causal effect or in the right direction. Although propensity-score matching is often used in these types of studies, it is only able to reduce confounding to the extent that the observables that are used in propensity score matching are correlated with unobservable sources of confounding, which will not cover many sources of potential bias in items 2 and 3.

In a world without human subjects concerns, $ecig\ use_{ist}$ in equation 1 could be randomly assigned to determine what effect this assignment has on subsequent cigarette use, which would give the causal effect of e-cigarette use on subsequent cigarette use. Because this is not feasible for youth on human subjects grounds, the next best case is to use a natural experiment research design leveraging e-cigarette policy variation as an "exogenous shock" to e-cigarette use. In short, e-cigarette policies may make it less likely that youth use e-cigarettes. Researchers analyzing natural experiments then attempt to leverage a small component of exogenous change in whether or not individuals use e-cigarettes (forced on them by the policy) to observe what impact this then has on subsequent cigarette use. These points have been made previously in other gateway research (16,17). The natural experiment equation replaces individual-level e-cigarette use, which is subject to confounding from individual-level selection, with plausibly exogenous policy variation[3]:

$$cig\ use_{is(t+1)} = \propto + ecig\ policy_{st}\tilde{\beta}_1 + X_{ist} + \pi_s + \omega_t + u_{ist}. \quad (2)$$

This is now a 2-way fixed effect (state and time) difference-in-differences model.[4]

1 Longitudinal cohort studies for a single state do not contain state controls. In cases in which multiple states are used but state is not available, the inability to control for state fixed effects and state-level policy changes represents potential sources of confounding.

2 Though these models are rarely used in this type of research, this sort of confounding could be resolved by using individual fixed effects in equation 1 rather than individual characteristics.

3 This is considered a reduced form model, though another approach would be to use an instrumental variable model in which e-cigarette use in equation 1 is instrumented with the e-cigarette policy variation.

4 A key identifying assumption in difference-in-differences models is that the treatment states have similar trends to the control states in the absence of treatment. If this can be

In the case of understanding the effect of e-cigarettes on cigarette use, findings differ considerably depending on the design used. Five early studies documented a strong association between e-cigarette use and subsequent cigarette use among tobacco-naive youth and young adults using longitudinal cohort studies (18-22). A Surgeon General report used these studies to issue a major conclusion that "e-cigarette use is strongly associated with the use of other tobacco products among youth and young adults, including combustible tobacco products" (23).[5] If this association is causal, then rising youth e-cigarette use rates would translate to rising youth cigarette use rates.

Two natural experiment difference-in-differences studies were published at the same time as the longitudinal cohort studies using policy variation from e-cigarette minimum legal purchase age laws (24,25). These papers reached the opposite conclusion by finding e-cigarettes were same-period economic substitutes, with one specifically exploring (and finding evidence for) e-cigarettes being intertemporal economic substitutes as well.[6,7]

The Food and Drug Administration's Population Assessment of Tobacco and Health data, a national longitudinal cohort study, could have helped resolve this disagreement in the literature between the findings of longitudinal cohort studies and natural experiments. But geocoded linkages were not allowed until February 1, 2019, and so natural experiments were not possible until then.[8] Longitudinal cohort studies were allowed, and these studies found evidence supporting other longitudinal cohort studies that e-cigarette use was associated with subsequent combustible use (34-37).

As several years have elapsed since these studies and the 2016 Surgeon General report, the forecasting abilities of these 2 empirical approaches can now be compared to see which was more accurate in forecasting objective real-world data. Youth cigarette use fell much more sharply than predicted during the decade. The United States established Healthy People 2020 goals that, among other things, called for reducing youth cigarette use from 19.5% to 16.0% by 2019 (38). By 2019, the youth cigarette use rate was at a remarkable 6.0% (record lows) thus beating Healthy People 2020's ambitious goal of a 3.5 percentage point reduction (from 19.5% to 16.0%) over the decade by 386%. Natural experiments predicted the correct direction of the relationship between e-cigarette use and cigarette use, whereas longitudinal cohort studies predicted the opposite.

In the absence of geocoded dating owning to security concerns, methodologically weaker longitudinal cohort studies proliferated during the past decade, and unobserved confounding from this study design appears to have resulted in a flawed Surgeon General major conclusion as well as lawmakers passing e-cigarette regulation that many studies find slowed the decline in cigarette use that would have otherwise occurred (24-33). In conclusion, data access restrictions can result in an imbalanced literature that magnifies errors, which in turn can contribute to suboptimal policy.

## Potential Solutions

The above example shows that natural experiments can provide more accurate estimates of important policy-relevant questions, but in at least some cases, government-collected data are not collected or made available to researchers in a way that encourages or allows natural experiment research methods.

One change afoot is the Foundations for Evidence Based Policymaking Act of 2018. Although there are provisions of this law that should expand data access to researchers such as centralized processing of federal data requests,[9] there are also aspects of the law that may reduce data access by, for example, requiring that agencies conduct comprehensive risk analyses that take into account the increasing potential for inappropriate re-identification (39). The act also requires each data provider to have a chief data officer.

I offer a few recommendations based on my experience as a user of government-collected survey data to conduct tobacco control and cancer prevention and early detection research that I believe will improve the ability of researchers to conduct natural experiment research without compromising data security. These recommendations appear to be supported in spirit by the Foundations for Evidence Based Policymaking Act of 2018 and

shown (such as through the use of an event study design), this provides a credible method to explore this relationship in a way that avoids bias from individual-level selection.

5 Although not stated in the language of the "Major Conclusions," the 5 referenced studies only examined the relationship for tobacco-naive youth rather than considering possible beneficial effects as well of e-cigarettes on youth smoking cessation. Unless there is reason that one transition path (initiation of combustible tobacco) is more important than another transition path (cessation of combustible tobacco), it is unclear why these studies stratify the populations based on tobacco use or not rather than simply studying the average effect across both users and nonusers.

6 Both of these studies used aggregated state-level data rather than individual-level data, controlling for demographics at the state level. This could, in theory, reduce precision and cause bias if relevant state-level demographic variables are not included, but no such bias was detected when one of the analyses was redone using individual-level data (24,26).

7 Many other studies have since also provided additional evidence from natural experiments that e-cigarettes are economic substitutes, hence, overall displacing smoking (26-33).

8 The Population Assessment of Tobacco and Health study is a national longitudinal study of tobacco use and how it affects the health of people in the United States and is one of the first large tobacco research efforts undertaken by the National Institutes of Health and the Food and Drug Administration since Congress gave Food and Drug Administration authority to regulate tobacco products in 2009. The Population Assessment of Tobacco and Health data collection began in September 2013, and data collection is ongoing today. Data are available to researchers with approximately a 1-year lag through the National Addiction & HIV Data Archive Program system, which is a secure enclave accessible from the researcher's work computer with a full range of statistical programs. For researchers to receive access, the National Addiction & HIV Data Archive Program requires a project description, institutional review board approval, and data use agreement. All output is subject to disclosure review (eg, an analyst views all output for confidentiality concerns before releasing to the researcher). This model is also used by the National Center for Health Statistics, though it additionally requires the researcher to use a computer in one of their designated facilities rather than allowing researchers to access the enclave from their work computer.

9 For example, see https://www.ResearchDataGov.org.

would be valuable to implement in the process of operationalizing this law.

## Recommendation 1: Provide Public Access to Aggregated Data at the Smallest Levels of Area and Time Possible Without Compromising Data Security

A number of data providers provide public use aggregate data at the geocode level that make natural experiments possible without compromising data security. For example, the Centers for Disease Control and Prevention (CDC) WONDER system[10] provides an online query system to generate population, death, and environmental data at small levels of geography. The Youth Risk Behavior Surveillance System also uses an online query system[11] allowing for the creation of state-level panel data by different demographic groups, which was used in one of the early e-cigarette natural experiment studies (24). The Bureau of Labor Statistics' Consumer Expenditure Survey[12] is used for annual report releases containing aggregated geocode-level data (for several metropolitan statistical areas and regions). Though these reports do not provide demographic breakdowns, they still accommodate natural experiments that generally need geocodes but not necessarily aggregated demographic data (though this helps). The first priority in releasing public use aggregated data should always be geocodes, and the second priority should be pairing these geocodes with demographic data (eg, providing state-level data separately for men and women), which will permit stronger, more targeted natural experiment evaluations.

Other data providers make it more difficult or impossible for natural experiment studies. The NHIS query system provides demographic breakdowns but not geographical breakdowns.[13] Improving this query system to include geocoded data would be an important improvement for improving access to the NHIS data for researchers who otherwise do not have the time or resources to access confidential versions of these data sources in research data centers.

Some government-collected survey data are not collected in a way that allows their release through either query systems, reports, or restricted access methods. Currently, the National Youth Tobacco Survey (NYTS)[14] is one survey data source that is collected by the CDC and whose geocodes recently can only be used in evaluations by CDC scientists, which raises scientific integrity questions about replication opportunities and equal access. Such surveys that cannot allow at minimum aggregated data to be released at the area level should be redesigned to prioritize natural experiment research, which is the strongest available study design of retrospective data and can return causal effects (3). For the NYTS in particular, the CDC may need to change its subject assent and consent language to permit future releases of the data to be used for natural experiment research and by a broader array of researchers.

There are sometimes concerns about using population-weighted data at levels smaller than the level at which the data are representative. For example, NYTS data are nationally representative but not state representative, so the CDC may consider this a valid reason to not release aggregated data at the state level because of measurement error in producing state estimates from national data. However, natural experiments are already designed to show the extent of and potentially overcome bias from a variety of sources, including measurement error. For this reason, many natural experiments have been studied at levels smaller than the representativeness of the data (eg, natural experiments using national Youth Risk Behavior Surveillance System and Monitoring the Future data (40-45). Without appropriate weights for the level of aggregation, providing unweighted means will be most useful to researchers in these situations (46).

10 https://wonder.cdc.gov/
11 https://nccd.cdc.gov/Youthonline/App/Default.aspx
12 https://www.bls.gov/cex/
13 https://www.cdc.gov/nchs/nhis/shs.htm
14 https://www.cdc.gov/tobacco/data_statistics/surveys/nyts/index.htm

## Recommendation 2: Approve Projects With Enough Time to Allow for Publication in Journals With Lengthy Peer-Review Times

Some projects require the use of individual-level data with geocodes. For these projects, the National Center for Health Statistics currently approves the use of data with masked geocodes at research data centers for a 3-year period. If researchers wish to continue their project beyond this point, they need to apply for new approval, and there is no guarantee that the project will be approved in its current form (if at all). This review process can take many weeks to complete.

Publication times in leading journals have increased in recent years. In 2020, the average time from submission to publication for a top economics journals was 33.2 months and 23 months for 3 top social science journals (47). This of course does not count the time spent doing the research or pursuing peer review at other journals. With review times this long, 3-year approvals are not enough time in most cases for researchers trying to publish in top social science journals. Some researchers may only benefit professionally from these types of publications rather than historically faster clinical publications. These researchers offer important insights to answer complex questions affecting the health of the nation, and so the National Center for Health Statistics should ensure their policies are not accidently discouraging participation by this important group of researchers.

## Recommendation 3: Improve Communication and Transparency Between Data Providers and the Research Community

To the research community trying to do analyses of natural experiments, decisions about data collection and access can often seem capricious and unclear if the data collectors and providers fully grasp how decisions they make affect whether natural experiments can be performed. It is unclear in many cases if confidentiality decisions rest on interpretation of federal law and the consenting process or if decisions rest more on data providers' understanding of the benefits of the proposed research vs the risks of disclosure. For consent issues, a reasonable question is if the consenting process could be changed to allow more equal access to future releases of the data. For situations that are not based on legal reasoning but rather on interpretation of possible benefits and risks, researchers should have opportunities to make the case to the data provider for the value of their research, especially because new methods are being regularly developed that the data provider may be unfamiliar with.

Government data providers should have a formal and transparent review process for petitions on ways to improve data access without compromising data security. Regardless of the decision hinging on a legal rationale, or more nebulous cost-benefit tradeoff, formal petitions should be responded to in writing, and these responses should be publicly disclosed. This will allow the research community to see the history of the issue and better determine next steps. This also safeguards against data providers making inconsistent decisions regarding data access depending on the researcher. In general, transparency when it comes to decisions regarding data access could reduce frustration (on both sides) and possibly increase creative solutions.

In addition to a formal and public petition process, informal meetings between chief data officers and the research community would also be valuable. Such informal communication may be helpful for brainstorming possible solutions that accomplish both goals of access and security.

## Discussion

The tension between researchers analyzing natural experiments and data providers will not be resolved overnight, but low-lying fruit appears within reach that would help increase data access without compromising security. Such approaches should continue to be brainstormed and pursued.

## Funding

Dr Pesko is supported by the National Institute on Drug Abuse of the National Institutes of Health (NIDA-NIH) (grant no. R01DA045016). The content is solely the responsibility of Dr Pesko and does not necessarily represent the official views of NIDA-NIH.

## Notes

**Role of the funder:** The funder (National Institute on Drug Abuse) provided grant support only.

**Disclosures:** Dr Pesko reports no other conflicts of interest.

**Author contributions**: Dr. Pesko was responsible for all aspects of the manuscript.

## Data Availability

No new data were generated or analyzed in support of this research.

## References

1. Khullar D, Jena AB. "Natural Experiments" in health care research. *JAMA Health Forum*. 2021;112(6):e210290.
2. Tulchinsky TH. John Snow, cholera, the broad street pump; waterborne diseases then and now. In: *Case Studies in Public Health*. Elsevier; 2018:77–99. https://linkinghub.elsevier.com/retrieve/pii/B9780128045718000172. Accessed October 29, 2021.
3. The Prize in Economic Sciences 2021. Press release. The Nobel Prize; October 11, 2021. https://www.nobelprize.org/prizes/economic-sciences/2021/press-release/. Accessed October 29, 2021.
4. Simon K, Soni A, Cawley J. The impact of health insurance on preventive care and health behaviors: evidence from the first two years of the ACA Medicaid expansions. *J Policy Anal Manage*. 2017;36(2):390–417.
5. Courtemanche C, Marton J, Ukert B, Yelowitz A, Zapata D. Early effects of the Affordable Care Act on health care access, risky health behaviors, and self-assessed health. *Southern Econ J*. 2018;84(3):660–691.
6. Tello-Trillo DS. Effects of losing public health insurance on preventative care, health, and emergency department use: evidence from the TennCare disenrollment. *Southern Econ J*. 2021;88(1):322–366.
7. Maclean JC, Halpern MT, Hill SC, Pesko MF. The effect of Medicaid expansion on prescriptions for breast cancer hormonal therapy medications. *Health Serv Res*. 2020;55(3):399–410.
8. Maclean JC, Pesko MF, Hill SC. Public insurance expansions and smoking cessation medications. *Econ Inq*. 2019;57(4):1798–1820.
9. Barnes JM, Johnson KJ, Adjei Boakye E, et al. Early Medicaid expansion and cancer mortality. *J Natl Cancer Inst*. 2021;113(12):1714–1722.
10. Stough R, McBride D. Big data and U.S. public policy. *Rev Pol Res*. 2014;31(4): 339–342.
11. Groves RM, Schoeffel GJ. Use of administrative records in evidence-based policymaking. *Ann Am Acad Political Soc Sci*. 2018;678(1):71–80.
12. National Cancer Institute. Cancer health economics research. National Cancer Institute; 2021. https://healthcaredelivery.cancer.gov/cancer-health/. Accessed January 5, 2022.
13. American Cancer Society. More than 4 in 10 cancers and cancer deaths linked to modifiable risk factors. American Cancer Society; 2017. https://www.cancer.org/latest-news/more-than-4-in-10-cancers-and-cancer-deaths-linked-to-modifiable-risk-factors.html. Accessed December 16, 2021.
14. National Institutes of Health. RFA-OD-13-014: Mentored Research Scientist Career Development Award in Tobacco Control Regulatory Research (K01). National Institutes of Health; 2013. https://grants.nih.gov/grants/guide/rfa-files/RFA-OD-13-014.html. Accessed June 8, 2021.
15. Klein JD. E-cigarettes: a 1-way street to traditional smoking and nicotine addiction for youth. *Pediatrics*. 2018;141(1):e20172850.
16. Beenstock MB, Rahav GR. Testing gateway theory: do cigarette prices affect illicit drug use? *J Health Econ*. 2002;21(4):679–698.
17. Etter J-F. Gateway effects and electronic cigarettes. *Addiction*. 2018;113(10): 1776–1783.
18. Leventhal AM, Strong DR, Kirkpatrick MG, et al. Association of electronic cigarette use with initiation of combustible tobacco product smoking in early adolescence. *JAMA*. 2015;314(7):700–707.
19. Primack BA, Soneji S, Stoolmiller M, Fine MJ, Sargent JD. Progression to traditional cigarette smoking after electronic cigarette use among US adolescents and young adults. *JAMA Pediatr*. 2015;169(11):1018–1023.
20. Barrington-Trimis JL, Urman R, Berhane K, et al. E-cigarettes and future cigarette use. *Pediatrics*. 2016;138(1):e20160379.
21. Unger JB, Soto DW, Leventhal A. E-cigarette use and subsequent cigarette and marijuana use among Hispanic young adults. *Drug Alcohol Depend*. 2016; 163:261–264.
22. Wills TA, Knight R, Sargent JD, Gibbons FX, Pagano I, Williams RJ. Longitudinal study of e-cigarette use and onset of cigarette smoking among high school students in Hawaii. *Tob Control*. 2017;26(1):34–39.
23. United States Surgeon General. E-Cigarette use among youth and young adults: a report of the Surgeon General. Washington, DC: Department of Health and Human Services; 2016.
24. Pesko MF, Hughes JM, Faisal FS. The influence of electronic cigarette age purchasing restrictions on adolescent tobacco and marijuana use. *Prev Med*. 2016;87:207–212.
25. Friedman AS. How does electronic cigarette access affect adolescent smoking? *J Health Econ*. 2015;44:300–308.
26. Dave D, Feng B, Pesko MF. The effects of e-cigarette minimum legal sale age laws on youth substance use. *Health Econ*. 2019;28(3):419–436.
27. Pesko MF, Courtemanche CJ, Catherine Maclean J. The effects of traditional cigarette and e-cigarette taxes on adult tobacco product use. *J Risk Uncertain*. 2020;60(3):229–258.
28. Saffer H, Dench DL, Grossman M, Dave DM. E-cigarettes and adult smoking: evidence from Minnesota. *J Risk Uncertainty*. 2020;60(3):207–228.
29. Pesko MF, Warman C. Re-exploring the early relationship between teenage cigarette and e-cigarette use using price and tax changes. *Health Econ*. 2022; 31(1):137–153.
30. Abouk R, Adams S, Feng B, Maclean JC, Pesko MF. *The Effect of e-cigarette Taxes on Pre-pregnancy and Prenatal Smoking*. National Bureau of Economics Research; 2020. NBER Working Paper Series, No. 26126. http://www.nber.org/papers/w26126. Accessed March 24, 2022.
31. Cotti CD, Courtemanche CJ, Maclean JC, Nesson ET, Pesko MF, Tefft N. *The Effects of e-cigarette Taxes on e-cigarette Prices and Tobacco Product Sales: Evidence from Retail Panel Data*. National Bureau of Economics Research; 2021. NBER Working Paper Series, No. 26724. http://www.nber.org/papers/w26724. Accessed March 24, 2022.
32. Pesko MF, Currie JM. E-cigarette minimum legal sale age laws and traditional cigarette use among rural pregnant teenagers. *J Health Econ*. 2019;66:71–90.
33. Abouk R, Courtemanche C, Dave D, . et al. *Intended and Unintended Effects of E-cigarette Taxes on Youth Tobacco Use*. National Bureau of Economic Research; 2021. NBER Working Paper Series, No. 29216. http://www.nber.org/papers/w29216.pdf. Accessed October 31, 2021.
34. Berry KM, Fetterman JL, Benjamin EJ, et al. Association of electronic cigarette use with subsequent initiation of tobacco cigarettes in US youths. *JAMA Netw Open*. 2019;2(2):e187794.

35. Osibogun O, Bursac Z, Maziak W. E-cigarette use and regular cigarette smoking among youth: population assessment of Tobacco and Health Study (2013-2016). *Am J Prev Med*. 2020;58(5):657–665.
36. Pierce JP, Chen R, Leas EC, et al. Use of e-cigarettes and other tobacco products and progression to daily cigarette smoking. *Pediatrics*. 2021;147(2):e2020025122.
37. Stanton CA, Bansal-Travers M, Johnson AL, et al. Longitudinal e-cigarette and cigarette use among US youth in the PATH study (2013-2015). *J Natl Cancer Inst*. 2019;111(10):1088–1096.
38. HealthyPeople.gov. Adolescent cigarette smoking in past 30 days. Office of Disease Prevention and Health Promotion; 2020. https://www.healthypeople.gov/2020/data/Chart/5342?category=1&by=Total&fips=-1. Accessed June 4, 2021.
39. Glied S. New law enacts recommendations of commission on evidence-based policymaking. AcademyHealth; 2019. https://academyhealth.org/blog/2019-01/new-law-enacts-recommendations-commission-evidence-based-policymaking. Accessed June 11, 2021.
40. Feng B, Pesko MF. Revisiting the effects of tobacco retailer compliance inspections on youth tobacco use. *Am J Health Econ*. 2019;5(4):509–532.
41. Hansen B, Sabia JJ, Rees DI. Have cigarette taxes lost their bite? New estimates of the relationship between cigarette taxes and youth smoking. *Am J Health Econ*. 2017;3(1):60–75.
42. Abouk R, Adams S. Bans on electronic cigarette sales to minors and smoking among high school students. *J Health Econ*. 2017;54:17–24.
43. Mark Anderson D, Matsuzawa K, Sabia JJ. Marriage equality laws and youth mental health. *J Law Econ*. 2021;64(1):29–51.
44. Sabia JJ, Anderson DM. The effect of parental involvement laws on teen birth control use. *J Health Econ*. 2016;45:55–62.
45. Anderson DM, Hansen B, Rees DI, Sabia JJ. Association of marijuana laws with teen marijuana use: new estimates from the youth risk behavior surveys. *JAMA Pediatr*. 2019;173(9):879–881.
46. Solon G, Haider SJ, Wooldridge J. What are we weighting for? *J Hum Resourc*. 2015;50(2):301–316.
47. Hadavand A, Hamermesh D, Wilson W. *Publishing Economics: How Slow? Why Slow? Is Slow Productive? Fixing Slow?* National Bureau of Economic Research; 2021. NBER Working Paper Series, No. 29147. http://www.nber.org/papers/w29147.pdf. Accessed December 19, 2021.