NBER WORKING PAPER SERIES

EXPLAINING THE WIDENING DIVIDES IN US MIDLIFE MORTALITY: IS THERE A SMOKING GUN?

Christopher L. Foote
Ellen Meara
Jonathan S. Skinner
Luke R. Stewart

Working Paper 34553
http://www.nber.org/papers/w34553

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2025

We are grateful to Maximilian Worth for outstanding research assistance. For helpful comments we thank Paul Novosad, Anne Case, Andrew Cherlin, and seminar participants at the University of Chicago's Harris School of Public Policy, the Wennberg International Collaborative, the American-European Health Economics Study Group, the 2024 Empirical Health Economics Workshop, the University of Houston, and the 2025 Russell Sage Conference Beyond Deaths of Despair: New Research on the Health and Well-Being of Low- and Moderately-Educated Americans. We thank Paul Krugman for suggesting the title. The views stated herein are those of the authors and not necessarily those of the Federal Reserve Bank of Boston, the Federal Reserve System, or the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2025 by Christopher L. Foote, Ellen Meara, Jonathan S. Skinner, and Luke R. Stewart. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Explaining the Widening Divides in US Midlife Mortality: Is There a Smoking Gun?
Christopher L. Foote, Ellen Meara, Jonathan S. Skinner, and Luke R. Stewart
NBER Working Paper No. 34553
December 2025
JEL No. I1, I12, I14, I18

ABSTRACT

The education-mortality gradient has increased sharply in the last three decades, with the life-expectancy gap between people with and without a college degree widening from 2.6 years in 1992 to 6.3 years in 2019 (Case and Deaton 2023). During the same period, mortality inequality across counties rose 30 percent, accompanied by an increasing rural health penalty. Using county- and state-level data from the 1992–2019 period, we demonstrate that these three trends arose due to a fundamental shift in the geographic patterns of mortality among college and non-college populations. First, we find a sharp decline in both mortality rates and geographic inequality for college graduates. Second, the reverse was true for people without a college degree; spatial inequality became amplified. Third, we find that rates of smoking play a key role in explaining all three empirical puzzles, with secondary roles attributed to income, other health behaviors, and state policies. Less well-understood is why "place effects" matter so much for smoking (and mortality) for those without a college degree.

Christopher L. Foote
Federal Reserve Bank of Boston
chris.foote@bos.frb.org

Ellen Meara
Harvard University
T.H. Chan School of Public Health
Department of Health Management and Policy
and NBER
emeara@hsph.harvard.edu

Jonathan S. Skinner
Dartmouth College
Department of Economics
and NBER
jonathan.skinner@dartmouth.edu

Luke R. Stewart
University of Pennsylvania
Luke.Stewart1@Pennmedicine.upenn.edu

1 Introduction

The mortality-education gradient has risen dramatically over the past three decades, particularly at midlife (ages 25–64). As Case and Deaton (2023) demonstrated, this divergent trend led to a rise in the educational gradient in life expectancy between 4-year college and non-college graduates from 2.6 years in 1992 to 6.3 years in 2019, a change that translates conservatively to a $39 trillion loss to society.[1] Although the educational gradient in mortality exploded during the COVID-19 pandemic, rising to 8.5 years by 2021 (Case and Deaton 2023), this paper focuses on the longer-term trends prior to COVID-19, as we expect those patterns to persist during the post-COVID-19 period.

Yet the rising educational gradient is not the only troubling mortality pattern to emerge in recent years. Geographic mortality inequality has also increased rapidly, with the county-level coefficient of variation rising from 0.25 in 1992 to 0.33 in 2019. Additionally, we demonstrate below that geographic variation in midlife mortality over time evolves differently by educational group. For college graduates, mortality rates have converged at both the state and county levels.[2] In Montana, for example, the age-adjusted midlife mortality rate for college graduates was 225 per 100,000 population in 1992, well below the equivalent rate in New York (337 per 100,000). But by 2019 the mortality rates for college graduates in the two states had largely converged (135 per 100,000 in Montana compared to 120 in New York). By contrast, for people without a college degree, mortality rates in New York declined from 521 per 100,000 in 1992 to 385 per 100,000 in 2019, while corresponding mortality rates in Montana rose from 393 to 472 per 100,000.

[1] The counterfactual is one in which the gap would have remained constant at 2.6 years, so that the population of people in midlife experienced a loss of 3.7 years (6.3 minus 2.6). Valuing a life year at $100,000 and multiplying that amount times the number of people in 2019 aged 25–64 without a 4-year college degree (105.3 million) yields $39 trillion. This figure is an underestimate if we ignore future generations not yet in midlife who similarly experience lower life expectancy (and life-years past age 85, which is Case and Deaton's upper limit). It is an overestimate if future lost life years are discounted.

[2] As explained below, college mortality has exhibited both beta and sigma convergence. These two concepts are often used in the macroeconomic study of income convergence across countries. In our context, beta convergence occurs when counties with relatively high initial mortality rates experience relatively large declines in mortality. Sigma convergence occurs when the distribution of mortality rates across counties becomes smaller over time.

Rising educational and spatial differences in mortality have coincided with a third fundamental shift. Among Americans at midlife, the modest health and mortality advantage that residents of rural counties historically enjoyed over residents of large urban counties has reversed since the 1990s, with a particularly large rural-urban gap for non-Hispanic Whites and in the South (Monnat 2020; Miller and Vasan 2021). As we show below, midlife mortality rates in the country's most-urban counties were around 300 deaths per 100 thousand population in 2019. In the country's most-rural counties, mortality rates were about 50 percent higher.

Several hypotheses have been offered to explain these trends. First, many explanations for the widening educational divide involve so-called deaths of despair—drug overdoses, alcohol-related diseases, and suicides—which have been described in groundbreaking work by Case and Deaton (2015, 2017, 2021) and others. Second, for the rural-urban divide, selective net migration could arise from healthier people moving from depressed rural areas to cities, or from relatively healthy immigrants from abroad moving to urban areas in the United States. Third, some have argued that mortality trends could be statistical artifacts driven by selection. For the educational gradient, selection might arise from more people attending college, leaving only the presumably less healthy persons in the high-school population (Hoxby 2023).

In this paper, we first demonstrate that none of the three explanations above can explain the patterns we observe in the data. Although deaths of despair contribute to a widening of the education-mortality gradient (as others have shown), they do not explain either the growing geographic divergence in mortality or the worsening rural mortality penalty; low-mortality urban states like New Jersey were also badly hit by the rise in deaths of despair. When we adjust for migration rates, these factors explain only a small fraction of the overall shifts. Finally, we find no evidence that the selection of less-healthy people into the growing population of college graduates (or the shrinking population without college degrees) can explain the rise in the education-mortality gradient.

We consider other potential determinants of the three trends in midlife mortality. One

is the fundamental shift away from manufacturing employment, as the China Shock and related shifts have left many communities unable to provide adequate high-paying and stable jobs (Autor, Dorn, and Hanson 2013; Pierce and Schott 2020). In addition, a growing literature has emphasized the importance of state-level policies, which include tobacco taxes and restrictions, supplemental income programs, Medicaid coverage, and a variety of legislative or regulatory rules (Montez et al. 2020, 2022; Beckfield and Bambra 2016). Because we use county-level data, in the cross-section we are able to include a measure of state-level policies in our regressions. We can also include state fixed effects to capture time-invariant effects of state-level policies, as well as any remaining state-level factors not captured by our right-hand side variables. Yet even with state fixed effects, we find that the three patterns noted above persist quite strongly—the within-state variation by county in mortality rates by education is still very large. For example, the 1992–2019 decline in non-college mortality in urban Cook County, Illinois, from 577 to 462 per 100,000, was in sharp contrast to the increase of 574 to 622 deaths per 100,000 in rural Franklin County, Illinois.

More important than these factors is income (Chetty et al. 2016). At the state level, midlife mortality three decades ago was essentially uncorrelated with per capita income.[3] Since then, however, income has become a highly predictive and economically important factor in explaining variation both within education groups at the aggregate level across regions, and between education groups. Yet even here trends in regional per capita income fail to fully explain the three puzzles we observe in the data. Nor is it easy to interpret an association between income and mortality as causal. What exactly is it about income that causes it to be correlated with mortality? Is it a third factor that is correlated with both mortality and income?

A final leading candidate for rising midlife mortality is health behaviors. There is a

[3]See for example Couillard et al. (2021). It is important to note that the lack of correlation at the aggregate (county or state) level in 1992 was not consistent with the micro-level evidence at the time, which showed a negative gradient between income and mortality. For example, the estimated 10-year mortality risk in the Panel Study of Income Dynamics showed that near the bottom of the income distribution, each additional $10,000 of income was associated with more than a 50 percent reduction in the risk of death in the 1980s and 1990s (Dowd et al. 2011).

lengthy literature on the importance of obesity, exercise, diet, and smoking for the poor health outcomes of Americans (Loprinzi 2016). Of course, these behaviors are in turn likely to be influenced by factors such as income and state policies, working through tobacco taxes or income support programs, or other "fundamental causes" of poor health (Link and Phelan 1995). Yet a focus on health behaviors at least allows us to identify potential key pathways leading to the rising education-mortality gradient. One candidate for these mortality changes is obesity. However, we find that rising obesity rates in the United States cannot explain the increasing educational gradient (nor the rising urban-rural divide), because obesity is rising across the board among urban and college-educated people as well. In our analysis, we also considered other health measures, including inactivity and binge drinking, but these failed to explain variations in mortality.

Instead, smoking emerges as an exceptionally powerful predictor of mortality trends. Our geographic analysis links county and state college and non-college smoking rates to changes in education-specific mortality rates over the past three decades. We find that smoking helps explain the rising education-mortality gradient, the rising geographical dispersion in mortality, and the worsening urban–rural divide. At the county level, we show that smoking is much more closely linked to changes in mortality than local labor market variables or other factors. Higher and persistent smoking rates in rural areas can also explain a substantial portion of the relative change in the rural and urban mortality. At the state level, we find that levels and changes in state-level smoking rates strongly predict changes in state-level education-mortality gaps. Conditional on smoking rates, per capita income is still economically important, but adds little to predictive power and other variables are insignificant. Finally, regional convergence in smoking among college graduates and divergence among non-college graduates is highly predictive of the corresponding geographic convergence in college mortality rates and the divergence in non-college mortality rates.

The emergence of smoking as the most important geographic predictor of mortality raises a question. It is well known that national smoking rates have declined in the non-college

population relative to college graduates in recent years. Although college smoking rates are lower, the gap between current non-college and college rates has been gradually narrowing. How then can smoking explain the widening education-mortality divide? The answer relies on a simple model of health capital in the context of Grossman (1972) and Case and Deaton (2005). People who start smoking at age 18 begin to exhibit higher mortality several decades later, with particularly large effects beginning at ages 45–64 (Lawton et al. 2025). A health-capital model allows the mortality rates of older persons to be determined not only by their current smoking behavior but also by smoking in earlier years. In the United States, smoking rates started falling for college graduates earlier than they did for the non-college population. As a result, college graduates who are now aged 55–64—the age group with a disproportionate share of overall midlife mortality—are far less likely to have ever smoked than comparably aged persons without a degree. This discrepancy in "never-smoking" rates means that college graduates now aged 55–64 have, over the past 30 years, experienced less health depreciation from cigarette consumption. And the mortality rates follow the patterns in smoking; we show below that over the past three decades, mortality rates of college graduates aged 55–64 have declined by more than half.[4]

One may wonder how smoking alone can contribute to these broad patterns of mortality, given that lung cancer is only a small fraction of overall mortality. Indeed the gap in lung-cancer mortality rates between people with and without college degrees is shrinking, not rising. However, with rapidly improving treatments and screening for lung cancer (Howlader et al. 2020), the major impact of smoking over the longer-term—particularly for people aged 55–64—arises from other more-common tobacco-related diseases such as chronic obstructive pulmonary disease (COPD); cardiovascular diseases such as strokes, aneurysms, and heart attacks; diabetes; and other types of cancers (Carter et al. 2015). Perhaps more surprising is that past county-level smoking rates are highly predictive of deaths of despair. This finding,

[4]As we explain below, this rapid decline in mortality among people with college educations cannot be the consequence of selection. If increased college graduation rates over time had reduced the average health of college graduates, we would have expected rising mortality, not falling mortality.

however, is consistent with an emerging literature in biology that points to a causal influence of smoking on drug addiction, which occurs through epigenetic "priming." Mice exposed to nicotine for a week were 78 percent more likely, relative to controls, to prefer living areas where the water was laced with cocaine (Kandel and Kandel 2014; Rajabi et al. 2019).

Although we argue below that smoking is a causal factor in the widening mortality gap, we acknowledge that it is likely not the only such factor. Part of the strong correlation between smoking and mortality across states and counties likely reflects associations rather than causal effects, because factors that cause people to smoke can also lead to higher midlife mortality without a direct causal link to smoking. For example, both smoking and opioid addiction may reflect a common sense of despair.[5] Nonetheless, our results suggest that researchers hoping to explain the otherwise puzzling increase in the education-mortality gradient, the widening spatial disparity in mortality for non–college graduates, and the rising rural mortality penalty should consider whether their candidate explanations are consistent with a strong smoking-mortality correlation. Fortunately, continued declines in smoking rates, particularly among people who did not graduate from college, offer a rare bit of optimism for the future. Using projected state-level smoking rates from Stone et al. (2025), we predict these trends in smoking could lead to an 8-percent decline in the education-mortality gradient by 2035.

The remainder of this paper is structured as follows. Section 2 describes our data, and section 3 outlines the three empirical patterns in mortality addressed in our analysis. Section 4 illustrates how a simple model of health capital is useful in explaining the link between mortality and various health determinants, including smoking. Section 5 presents our main empirical analysis, and section 6 addresses three potential objections to our claim that a significant portion of the smoking-mortality correlation is causal. Section 7 concludes.

[5]As discussed below, regions with high rates of smoking also tend to have higher rates of cancer and Social Security Disability Insurance (SSDI) enrollees, which were both associated with a disproportionate rise in opioid use (Evans and Lieber 2025; Lortet-Tieulent et al. 2016).

2 Mortality Data at the County Level

We use both county- and state-level data in this paper. Creating county-level mortality rates presents some special challenges, given their relatively small sizes. We use the restricted individual-level mortality data from the National Center for Health Statistics (NCHS). Derived from death certificates, these data include the causes of death for each decedent as well as demographic information, including age, sex, race, ethnicity, education, and county of residence. We compute mortality rates in a given county or state by dividing the total number of deaths by the relevant population from the National Cancer Institute's Surveillance, Epidemiology, and End Results Program (SEER). SEER has county-level population data for demographic subgroups, allowing us to compute mortality rates by race, ethnicity, sex, and age. We compute age-specific mortality rates within the age bins of 25–34 years old, 35–44 years old, 45–54 years old, and 55–64 years old, then we age-standardize these rates using the 2000 decennial US Census age distribution for an overall midlife mortality rate. For mortality data from before 1989, we use publicly available data from the Centers for Disease Control and Prevention (CDC) WONDER database. In 1989, death certificates began to include education. Although some states were slow to adopt this policy, by 1992 most states were reporting education level on death certificates. We limit our attention to the 44 states with the most accurate educational reporting.[6]

Education-specific populations by county are not readily available from SEER. To compute these populations, we use data from the US Census and the American Community Survey (ACS). The smallest identifiable geographic area in these data is the Public Use Microdata Area (PUMA), which must have populations of more than 100,000. County boundaries fall cleanly within PUMA boundaries. Frequently, multiple counties are combined to make up a PUMA, but in denser areas, counties map directly to PUMAs or multiple PUMAs make up one county. Generally, we compute the proportion of the population of an age-race-

[6]We omit Alaska, the District of Columbia, and Hawaii in all analyses. For education-specific analyses, we omit Georgia, Oklahoma, Rhode Island, and South Dakota.

sex-PUMA cell by educational attainment, then apply these proportions to PUMA-county crosswalks to compute the proportion of the population of an age-race-sex-county cell by educational attainment. We then multiply these shares by the total population within the cell taken from SEER.

The census data are available only for 1990, 2000, and 2010, while the ACS data are available from 2006 onward (in the form of multiple-year ACS estimates). Due to constraints on the quality of education data in the mortality records for 1989 through 1991, and to avoid complications due to the COVID-19 pandemic, we conduct education-specific analyses for 1992, 2000, 2010, and 2019. To obtain county-level education-specific population estimates for 1992, we linearly interpolate between the 1990 and 2000 census figures. Finally, although changes in the definitions and boundaries of counties are rare, they do occur in our sample. We treat counties that split or combined partway through our years of interest as one combined unit in every year. For further discussion of data sources and a full description of the counties in our sample, see the data appendix.

3 Three Mortality Patterns

3.1 Rising Educational Gradient

Case and Deaton (2023) estimate an increase in the non-college/college life expectancy gap from 2.6 years in 1992 to 6.3 years in 2019. (The gap widened further during the COVID-19 pandemic, but we restrict our attention to the 1992–2019 period.) The increased gradient is unlikely to have resulted solely from causal effects of education.[7] One explanation for the increase is selection effects (Hoxby 2023). Over time, a higher percentage of the midlife population has graduated from college, and college attainment could be more common

[7]Some evidence that college attendance has a causal effect on smoking comes from de Walque (2007), who uses the Vietnam draft as instrument for college attendance. He finds that educated individuals are less likely to start smoking and more likely to quit if they have begun. However, this result cannot explain the different trends by education in midlife mortality.

among healthier people. A selection effect could therefore cause the remaining non-college population in 2019 to be systematically less healthy than the non-college population in 1992.

Yet under reasonable assumptions, increases in educational attainment would also raise mortality risk in the college population. If mortality risk rises as one moves down the educational distribution, then the college population grows by adding people who have higher mortality risk than the people who are already in the college pool. Figure 1 provides evidence against the selection explanation by displaying 1992–2019 mortality rates separately for adults with and without college degrees. The figure shows that the increase in the educational gradient resulted from a *drop* in the college mortality rate that took place alongside a relatively *stable* non-college rate. This pattern is inconsistent with a negative selection effect on the college population. It is possible that all Americans experienced an underlying decline in mortality, but that an adverse selection effect on non-college mortality exactly offset the overall beneficial trend, leaving the non-college mortality rate stable. In the appendix, we evaluate that possibility using two approaches. First, we show that education percentiles changed relatively little for both college and non-college populations from 1992 to 2019, indicating that the rise in college attainment over this period had a small effect on where the typical person in the non-college or college populations placed in the educational distribution. Yet non-college and college mortality trends diverged sharply (Appendix Figure A.1.) Second, we use the bounds-estimation approach in Novosad, Rafkin, and Asher (2022) to show that the decline in college mortality survives the selection correction developed in that paper, and that the relatively stable non-college mortality rate is not driven by mortality among high-school dropouts (Appendix Figure A.2.) The appendix also shows a similar widening educational gradient since 1992 considering non-Hispanic Black and Hispanic populations of any race separately (Appendix Figure A.3).

3.2 Rising Inequality Across States and Counties

Figure 2 documents the increase in spatial inequality across US states and counties since 1992. The blue line depicts the population-weighted coefficient of variation among US counties, which has increased from 0.25 in 1992 to 0.33 in 2019. This increase of more than 30 percent has been especially rapid after 2000.[8] The dashed red line in Figure 2 calculates a modified coefficient of variation, where the numerator is not the raw variance of county-level mortality rates but the variance after absorbing state fixed effects. The growing gap between the blue and red lines indicates that dispersion in state-level mortality rates has been rising over time.[9] The upward trend in the red line indicates that even within states, dispersion in county-level midlife mortality rates has increased.

It is possible that sharp declines in mortality for the non-college population in some states may result from large immigrant inflows, particularly in states such as New York and California, as immigrants are often healthier than natives. More than two-thirds of immigrants during the past several decades have identified as Hispanic or Asian.[10] Consequently, one way to investigate this possibility is to construct state-level mortality rates that are limited to non-Hispanic Black and non-Hispanic White populations. Although the Black-White sample has much a smaller fraction of immigrants than the full sample, the appendix shows that the spatial distribution of mortality in the Black-White sample is nearly identical to the corresponding distribution for the full sample.[11] In our county-level regressions, we also control directly for migration outflow.

3.3 Reversal of the Rural Health Premium at Midlife

Related to the widening county-level dispersion in mortality is the rising rural health penalty. To classify counties along the urban-rural spectrum, we use the six classification

[8] For more on the recent divergence in county-level mortality, see Baltrus et al. (2019).

[9] See Woolf and Schoomaker (2019) for a discussion of rising state-level dispersion in life expectancy, and Couillard et al. (2021) for rising dispersion in state-level midlife mortality.

[10] https://www.pewresearch.org/short-reads/2024/07/22/key-findings-about-us-immigrants.

[11] See Appendix Figure A.4.

codes in the 2013 Urban-Rural Classification Scheme for Counties from the NCHS.[12] These codes are (1) large "central" metro counties in metropolitan statistical areas (MSAs) of 1 million population,[13] (2) large "fringe" counties in MSAs of 1 million or more population that do not qualify as large central metro counties, (3) medium metro counties in MSAs of 250,000–999,999 population, (4) small metro counties in MSAs of less than 250,000 population, (5) counties in micropolitan statistical areas, which are smaller than MSAs, and (6) counties in neither MSAs nor micropolitan statistical areas.

The upper panel of Figure 3 plots midlife mortality rates for each of the six NCHS categories. Red lines correspond to the most-urban designations (codes 1 and 2), blue lines correspond to medium and smaller metro counties (codes 3 and 4), and the most-rural counties are in green (codes 5 and 6). At the start of the sample period, the mortality rate is highest in NCHS code 1 (solid red line), reflecting in part the concentration of AIDS deaths in the largest urban counties. The improvement in AIDS treatments during the 1990s contributes to a striking decline in this mortality rate during this decade, but AIDS deaths alone cannot explain the continued overall decline in mortality in the largest urban counties after 2000.[14] Mortality rates for the smaller urban counties (codes 3 and 4) and the most-rural counties (codes 5 and 6) drift higher after the mid-2010s. At the end of the sample period in 2019, mortality rates line up inversely (and almost monotonically) across the 6 categories with respect to urban status.

The lower panel aggregates the six categories into three, and displays levels and changes in midlife mortality rates by education. The bar charts on the left side of the lower panel display data for non-college populations. In 1992, counties in the most-urban categories on the left (red bars) had higher non-college mortality rates than counties in the most-rural categories on the right (green bars). By 2010, however, this pattern had reversed, as non-college

[12]See Appendix Figure A.5 for a county-level map of NCHS codes, and Centers for Disease Control and Prevention (2014) for the details of the NCHS classification system.

[13]To be classified as a central county, the county must contain the entire population of the largest principal city of the MSA, be completely contained within the largest principal city of the MSA, or contain at least 250,000 residents of any principal city in the MSA.

[14]In Figure 15 below, we show education-specific data on HIV mortality.

mortality in the most-urban counties had declined while it rose in the most-rural counties. The rural-urban gap widened even further between 2010 and 2019. The corresponding bar charts for college populations show a qualitatively similar but smaller reversal, as college mortality was highest in large metro counties in 1992 but lowest in those areas in later years; college mortality fell in each NCHS grouping.

4 Health Capital and Health Depreciation

We motivate our analysis of the three mortality patterns with a model of health capital (Grossman 1972) that emphasizes the depreciation of health over time (Case and Deaton 2005; Galama 2015). Aside from depreciation that occurs from age, its simplest form, health capital depreciates because of poor health behaviors, physically or mentally punishing work, or a lack of investments in health, such as high-quality health care or exercise. Death typically occurs when health capital H drops below some critical value $\bar{H}$. An insight of the health capital model is that health choices often affect mortality with long and variable lags; smoking habits avoided at age 25 reduce mortality rates 30 years later. The historical record illustrates the lag between smoking and lung cancer deaths most clearly. U.S. cigarette sales peaked in 1963 at nearly 11 cigarettes per US adult per day.[15] Yet lung cancer deaths among US men did not peak until 1990; the rate for women peaked eight years later; see Figure 4.

Figure 5 provides some intuition for this model by depicting three hypothetical paths of health capital H over the adult ages. Here we focus on smoking as the relevant factor affecting the depreciation of H. We assume that each person depicted in the graph begins with the same initial health stock H_s, although empirically there is likely to be considerable variation in the initial health stock (Currie 2011). Person A eschews smoking and survives through midlife. Indeed, we might expect this person to live long past age 64 absent any later health shocks. Person B starts smoking at age 25 and quits several years later. Although

[15]Because not all adults smoked, this figure indicates that US smokers were smoking more than 11 cigarettes each day.

Person B's health stock declines while he smokes, his quitting attenuates further declines. Even so, Person B barely makes it to age 64—and given adverse shocks could have died during midlife—so this person is unlikely to live as long as Person A. Finally, Person C continues to smoke throughout her life and dies at an early age.

A health-capital approach is valuable in our context because much of the variation in midlife mortality rates is driven by mortality among persons near the end of midlife, when accumulated differences in health-capital depreciation are most important. Figure 6 shows deaths by 10-year age groups and education. Mortality from deaths of despair (blue bars) are depicted separately from deaths from other causes (red bars). As noted by Case and Deaton (2017) and others, in the non-college population there have been remarkably rapid increases in deaths of despair at both younger and older ages. But especially at older ages, these trends are dwarfed by the much larger prevalence of other deaths, such as those from cardiovascular disease and cancer.[16] The figure makes clear that deaths from these other causes among older age groups are important in driving overall midlife mortality patterns, a fact that in turn points to the importance of past health behaviors in determining current mortality rates.

5 County- and State-Level Mortality Patterns

5.1 County-Level Densities: Levels and Changes

To understand how the distribution of county-level mortality differs by education, Figure 7 plots education-specific kernel densities for county-level mortality rates in 1992 (blue density) and 2019 (red density). Densities for non-college populations appear in the top panel and those for college populations appear in the bottom panel. For people who did not graduate from college, the similar means of the 1992 and 2019 densities are consistent with the relative stability of the mean non-college mortality rate seen earlier in Figure 1. Yet the graph shows

[16] Of course, life expectancy calculations place more weight on early-life mortality.

that the *variance* of non-college mortality increased. In the lower panel, the college densities are consistent with the substantial drop in the mean college mortality rate in Figure 1, as the 2019 college density is shifted substantially to the left relative to the 1992 density. But in contrast to the rising variance of non-college mortality rates, the variance of college mortality rates declined from 1992 to 2019.

5.2 Correlates of County-Level Mortality Changes

What observable factors correlate with county-level changes in mortality? We investigate this question with regressions below, but we set the stage for that analysis with initial graphical evidence on beta convergence in mortality rates. Beta convergence occurs when the cross-sectional units that start with the largest values of some variable experience the largest declines in that variable over time. This concept can be contrasted with sigma convergence, which occurs when cross-sectional variances in the variable get smaller over time. Beta convergence can be overstated when there is transitory measurement error in the initial level. To avoid a spurious correlation between initial mortality levels and the subsequent changes in mortality, we use 1992 mortality as the initial level for each county, but calculate the county's change in mortality between 1993 and 2019.[17] The right panel of Figure 8 shows exceptionally strong evidence of beta convergence in college mortality rates, in contrast to weaker beta convergence for non-college mortality. There is also some modest evidence for beta convergence in non-college rates in the left panel. We therefore include the lagged dependent variable in our county-level regressions to capture beta convergence.

Despite modest beta convergence for non-college mortality rates, the distribution across counties has widened (Figure 7, upper panel). County-level health behaviors are likely to be important factors generating this dispersion in mortality. We therefore compute averages of smoking and obesity rates at the county-education level from the Behavioral Risk Factor Surveillance System (BRFSS). Smokers are defined as those who have smoked at least 100

[17] We calculate 1993 mortality rates using the same 1992 population denominators.

cigarettes in their lives, who consider themselves active smokers, and who smoke regularly. Obesity is defined as having a body mass index of 30 or higher.[18] We use the county identifiers (FIPS codes) that are available in the BRFSS microdata for 1996 through 2010, and we average the individual-level data from these years to estimate smoking and obesity rates for each county-education cell. Consequently, for our county-level analysis, smoking and obesity rates are constant within each county-education cell, because BRFSS sample sizes are not large enough to estimate yearly rates at the county-education level.

The upper-left panel of Figure 9 shows that for the non-college group, average smoking rates and 1992–2019 changes in mortality are highly correlated. In counties with non-college smoking rates below 20 percent, non-college mortality rates fell by about 100 per 100,000 population, close to the mean decline for the college population overall. Conversely, counties with non-college smoking rates of 30 percent or higher saw their non-college mortality rates rise by about 100 deaths per 100,000 population. The upper-right panel shows that the college smoking rates are much less correlated with college mortality changes. The lack of correlation arises primarily from the small amount of dispersion in college smoking rates, which cluster below 10 percent.

The lower panels of Figure 9 show that obesity rates are positively correlated with changes in mortality for both educational groups. Yet obesity is less likely to drive differences in mortality changes across the two groups, because obesity rates are more similar across the groups and because obesity has a smaller correlation with mortality than smoking does (note the difference in vertical scales across the top and bottom panels of Figure 9).

Before turning to the regression analysis of county-level mortality changes, we present time-series data on smoking rates by urban-rural status and education in Figure 10. We define "urban" areas as NCHS codes 1–4 while codes 5–6 designate "rural" areas. Because all smoking rates in this figure are national rates, the sample sizes are large enough to allow the rates to vary by year (unlike the county-level smoking rates, which are constant

[18]Other measures of health behavior such as exercise and binge drinking did a poor job of explaining the patterns we see.

education-specific averages taken over several years). The top panel shows that urban and rural smoking rates were nearly identical in 1992, at about 25 percent. But the decline in the urban smoking rate through 2019 was more than twice the size of the rural decline over the same period (12.4 percentage points vs. 5.0 percentage points). The lower panel presents overall smoking rates for non-college and college populations (solid lines), along with the education-specific rates disaggregated by urban-rural status (dotted lines for rural and dashed lines for urban). As has been noted in previous research, over this period the decline in the non-college smoking rate was somewhat larger than the decline in the college rate, although a large gap in these rates persists.[19] Smoking declines unfold more slowly in rural areas (6.4 percentage point in rural counties vs. 13.6 percentage points in urban counties). And the relatively high rate of smoking among non-college individuals in rural counties persists—it was 23.5 percent in 2019 compared with 15.9 percent for non-college urban residents. Yet by 2019 college smoking rates had fallen to 5.7 percent among rural residents and 3.8 percent for urban residents. Taken together, the two panels of Figure 10 indicate that since 1992, smoking rates across different counties have evolved in ways that mirror the mortality trends motivating our analyses.

5.3 County-Level Regressions

To consider other determinants of county-level mortality rates, we run separate county-level regressions for non-college and college mortality changes from 1993 to 2019. These education-specific regressions include a variety of county-level data, including the 1992 education-specific mortality rate; the log of per capita income in 1992 and the change in this variable from 1993–2019; average smoking and obesity rates from BRFSS over the 1996–2010 period; the county's 1992 employment to population ratio and the change in this variable from 1993 to 2019; the county's 1992 manufacturing share; its China-shock exposure from Autor, Dorn, and Hanson (2013); the county's out-migration rate from 2016 to 2020; and

[19] In our graph, the non-college rate (solid blue line) declines by 12.9 percentage points from 1992 to 2019, while the college rate (solid red line) declines by 9.2 percentage points.

an index of state policy liberalism from Caughey and Warshaw (2016). Each independent variable is standardized so that its coefficient can be interpreted as the effect on the mortality change that occurs from a one-standard-deviation change in that variable. We consider the analysis both with and without dummy variables for NCHS urban-rural codes (indicators for codes 3–4 and for 5–6). Unlike the other regressors, the urban-rural dummies are not standardized; the corresponding regressions without urban-rural controls are reported in the appendix. All county-level regressions are weighted by average education-specific population in 1993 and 2019.

Table 1 depicts regressions for county-level changes in non-college mortality between 1993 and 2019, with the regression in column 1 placing only the 1992 mortality level and the two urban-rural dummies on the right-hand side. Consistent with the negative slope in the non-college panel of Figure 8, the coefficient on the 1992 mortality level is negative and significant (–22.1, std. err. 5.2). Furthermore, the rurality dummy variables exhibt a pronounced increase in midlife mortality from 1993-2019; for example the most-rural areas (codes 5 and 6) experienced mortality changes that averaged more than 116 per 100,000 population greater than the changes in the most-urban areas. All told, the three variables in this column explain just under one-quarter of the variation in population-weighted mortality changes.

Column 2 adds the county's out-migration rate to the regression to investigate potential selection effects arising from differential mobility, but the resulting coefficient is small and insignificant. Adding the county's non-college smoking rate, as in column 3, has a much greater effect. A one-standard deviation increase in the smoking rate increases a county's 1993–2019 non-college mortality change by a highly significant 53 deaths per 100,000 persons, and the R^2 rises to almost 46 percent. Including the smoking rate also attenuates the rural effects; for the most-rural category the coefficient drops from 116 to 80.

Columns 3 through 9 repeat this exercise with different types of variables, including the county's obesity rate, income and employment variables (both 1992 levels and 1993–2019

changes), the county's initial manufacturing share and its China Shock exposure, and a state-policy liberalism index (which is the same for each county in the same state). Many of these variables are significant in these columns, although increases in the R^2 statistics relative to column 1 are typically modest. When all variables are included (Column 10), the partial effect of smoking remains strong, and the R^2 is improved only modestly relative to column 2 (49.9 percent vs. 45.9 percent). While obesity and income remain significant at the 95% level in Column 10, the magnitude of these and other coefficients are much attenuated, with little role for local labor markets. For example, the state policy-liberalism index becomes insignificant (with a coefficient of 6.1); to test the robustness of the index, we also include state-level fixed effects (Column 11), which captures all state-specific factors that predict changes in mortality during this period, but the coefficient estimates are largely unaffected. The implication is that even within states, county-level variations in health behaviors and income (specifically income changes) are most robustly associated with mortality changes.

The corresponding regressions for college graduates in Table 2 suggest much smaller (or insignificant) effects of the set of economic and health behaviors. Indeed, smoking rates exhibit coefficients an order of magnitude less than for non-college graduates and lack statistical significance. As in the non-college regression, the obesity rate and the income change remain significant in the latter columns. Yet the R^2 statistics in the college regressions are generally smaller than in the non-college regressions. As suggested by the college panel of Figure 8, much of the change in college mortality rates is explained by beta convergence, rather than the independent variables considered either separately or jointly.

5.4 State-Level Regressions

Although the county-level results reveal several patterns, estimating a dynamic version of changes in midlife mortality requires both lagged values and changes over time in health behaviors. Because changes in smoking and obesity are not available for county-education cells due to small sample sizes, Table 3 presents state-level regressions using mortality changes

from 1992 and 2019. Unlike the county-level regressions, coefficients are not standardized, so the smoking regressor ranges between zero and 1. Regressions of the non-college mortality rate appear in the first four columns of the table and the college regressions appear in the last four columns. As before, regressions are weighted by education-specific population.[20]

Previous state-level research has shown that income is an important correlate of mortality (Couillard et al. 2021; Chetty et al. 2016), so column 1 places per capita income in 1992 and the change in income between 1992 and 2019 on the right-hand side. These variables explain about half of the variation in non-college mortality changes ($R^2 = .512$). Column 2 replaces the income variables with initial levels and changes in smoking rates, which have highly significant effects; using just these two smoking variables increases the R^2 to 0.728. In the regression, increasing the *level* of non-college smoking in 1992 by 10 percentage points implies a subsequent increase in non-college mortality through 2019 of 98 per 100,000 persons. A decline of 10 percentage points in the *change* in non-college smoking between 1992 and 2019 reduces 2019 mortality by 137 per 100,000 people. Column 3 includes both the income and the smoking variables, which causes the income variables to become insignificant but has a much smaller impact on the smoking variables. Column 4 adds initial levels and changes in obesity rates. These coefficients are individually insignificant and their inclusion has a small effect on the R^2. Interestingly, the smoking coefficients in column 4 increase relative to those in column 3.

Columns 5 through 8 repeat this analysis for college-mortality changes. Similar to the county-level regressions, there are smaller effects of the potential explanatory variables. The R^2 reaches just 0.136 controlling for income, smoking, and obesity in column 8, in contrast to the R^2 of 0.782 for the analogous non-college regression in column 4. And, as in the county-level regressions, the coefficient on smoking rates in 1992 and on changes are much smaller in magnitude and insignificant. Obesity rates also fail to predict mortality trends

[20]Because we will use these regressions to calculate state-specific changes in education-based mortality gaps, we leave initial mortality levels out of the regressions of Table 3 for ease of interpretation. Results with the initial mortality levels included are qualitatively similar and appear in the appendix.

for the college population.

To get a visual sense for the predictive power of initial smoking rates and their changes over time, the top panel of Figure 11 compares actual changes in state-education mortality rates to predicted changes using the smoking-only models of columns 2 and 6 of Table 3. Predicted changes appear on the horizontal axes and actual changes are on the vertical axes. The first thing to note about the non-college chart is the remarkable heterogeneity in non-college mortality changes from 1992 to 2019. Changes range from a decline of 136 per 100,000 in New York to increases of 191 per 100,000 in West Virginia, 164 in Kentucky and 123 in Ohio. Another feature of the non-college chart is the close agreement of predicted and actual changes, consistent with the high R^2 in the non-college smoking-only model in Table 3. By contrast, the college chart illustrates that smoking patterns do less well in predicting the declines in college-graduate smoking, largely because there is so little secular variation in smoking among college graduates. College smoking declined in every state, as did midlife mortality rates. The lower panel of Figure 11 displays similar graphs using both smoking and income variables to predict mortality, that is, the predicted values from columns 3 and 7 of Table 3. The patterns resemble those in the top panel and visually demonstrate that, once one controls for initial levels and changes in smoking rates, income adds little to predictions of state-level mortality changes between 1992 and 2019.[21]

The importance of smoking as a determinant of mortality provides a reason for optimism regarding future mortality differences, because non-college smoking rates are gradually approaching college rates. How will this trend affect future education-mortality gradients? We draw from a recent study by Stone et al. (2025) using state-level smoking rates to predict corresponding smoking rates in 2035. We assign their state-specific proportional change in smoking to both college and non-college graduates, and consider both the average and

[21]In the appendix, we show that the state-level models of mortality changes in Table 3 can also be used to generate predictions for how a state's mortality *gap* changes between 1992 and 2019. Subtracting the predicted mortality change from a college regression from the predicted change from a non-college regression generates a predicted mortality-gap change for each state. As suggested by Figure 11, Appendix Figure A.6 shows that the smoking behavior of a state's residents generates accurate predictions for how a state's mortality gap evolves over time. Income variables add little explanatory power.

distribution of predicted smoking rates. Perhaps not surprisingly, predicted average smoking rates for the 44 states in our sample decline from 0.171 in 2019 to 0.139 in 2035 for non-college graduates, and from 0.039 to 0.031 for college graduates. Predicted mortality similarly declines, attenuating the education-mortality gradient by 8 percent.[22] Despite the average decline, the spatial coefficient of variation among non-college graduates is predicted to widen even further, given that smoking rates in Stone et al. (2025) are estimated to rise in a few states such as Iowa and Kentucky.

6 Is Smoking Causal?

The bottom line of the previous section is that as a statistical matter, county- and state-level differences in non-college smoking rates are most closely correlated with the non-college mortality patterns that drive the increase in the national mortality gap. Yet prediction is not causation. Perhaps the non-college smoking rate appears so strong in our analysis because it proxies for other social or economic forces that actually determine mortality. These factors might include the use of alcohol or illicit drugs, poverty, family instability, job-related stresses, or poor access to health care. At an individual level, the correlation of smoking with many other unhealthy characteristics is close to zero; for example, smokers are no more or less likely to be obese than nonsmokers (Cutler and Glaeser 2005). Yet there are at least three potential challenges to the claim that smoking is causal with respect to midlife mortality divides: (i) the apparent lack of divergence in college and non-college smoking rates in recent decades, (ii) the lack of adverse mortality trends in Europe, given Europeans' enthusiasm for smoking, and (iii) the fact that that US mortality divides are partially due to causes of death that are not normally associated with smoking. In this section, we address each of those challenges in turn.

[22] We use a simplified state-level regression model that predicts education-specific mortality in 2019 as a function of smoking rates and income in 2019.

6.1 The Larger Decline in the Non-College Smoking Rate

Previous attempts to relate smoking and other behavioral factors to widening mortality gaps in the United States have encountered a significant puzzle. Over a longer time period, educational differences in smoking and other behavioral risk factors do not appear to have grown enough to explain the widening mortality gap with respect to education (Cutler et al. 2011). Indeed, as seen earlier in the lower panel of Figure 10, the non-college smoking rate declined by somewhat more than the college rate from 1992 to 2019.

If *current* smoking rates and other behavioral factors are related to current mortality rates, then it is difficult to see how these factors could play any role in driving the widening educational gradient. Yet as the health capital approach emphasizes, health behaviors may affect mortality with long and variable lags; a smoking habit avoided at age 18 does not reduce mortality until 30 years later. Thus smoking (either current or past) should impose the largest mortality penalty on older midlife groups most at risk for diseases associated with smoking, specifically people aged 55–64.

To investigate smoking patterns by age in more detail, we turn to the Tobacco Use Supplement of the Current Population Survey (CPS-TUS). This survey has supplemented the basic CPS in several waves since the early 1990s. We use the waves collected from 1992–1993 (Wave 1) to 2018–2019 (Wave 10).[23] We define a current smoker in the CPS-TUS as someone who smokes "every day" or "some days." In addition to questions about current smoking behavior, the CPS-TUS includes questions on the respondent's smoking throughout his or her lifetime. These questions allow us to separate respondents into current smokers, never-smokers, and former smokers.

Figure 12 disaggregates education-specific smoking rates into three age groups: 25–44, 45–54, and 55–64. The top three panels show that declines in *current* smoking rates for the

[23]Each wave typically adds questions to three months of the CPS. See the appendix for a list of the months that are included in each of the 10 CPS-TUS waves used in this paper. We use a version of the data set that harmonizes questions across waves that was constructed by the National Cancer Institute, which sponsors the CPS-TUS. See National Cancer Institute (2021) for additional information.

younger non-college groups have been the largest. Smoking falls by more than 10 percentage points for the youngest non-college group (blue line in upper left panel) but by less than 5 percentage points for the oldest non-college group (blue line in upper right panel). The lower panel of Figure 12 plots *never-smoking* rates by education for the three age groups. Consistent with previous research (de Walque 2007; Farrell and Fuchs 1982), these panels indicate that college populations gave up smoking decades before non-college populations did. Particularly important for our argument is the college/non-college comparison among 55–64 year-olds, the age group for which the smoking mortality penalty is greatest. In 1992-93, the fraction of college graduates aged 55–64 who were never-smokers was 41 percent, nearly identical to the 43 percent rate for the non-college population in that age group. By 2018-19, however, the fractions of 55–64 year-old never-smokers had diverged sharply: 76 percent for college and 56 percent for non-college. The widening gap in never-smoking rates for older midlife people, combined with the steep smoking mortality penalty for this group, suggests these patterns of *never-smoking* by age could have widened the educational mortality gap, even as *current* midlife smoking rates for college and non-college populations have become more similar over time.

Smoking may also help explain the convergence of college mortality rates we have documented. Although our BRFSS data do not allow us to examine changes in county-level smoking rates, we can test for college convergence in smoking using state-level data. We construct average education-specific smoking rates from 1992–1999 for comparison to rates from 2010–2019. For each educational group, we plot the difference in smoking rates across the two periods against the early period average. Figure 13 shows clear evidence of beta convergence in college smoking rates, but no evidence of convergence among non-college rates.[24]

[24]In Appendix Figure A.9, we repeat this exercise using data from the CPS-TUS. There is again strong evidence of convergence in college smoking rates, and some evidence of *divergence* in non-college rates.

6.2 Smoking and Mortality in an International Context

Another challenge to the claim that smoking is a causal factor for midlife mortality divides is that Europeans often smoke more than Americans, but tend to live longer.[25] The top panel of Figure 14 uses data from Institute for Health Metrics and Evaluation (IHME) (2014) to confirm that recent smoking rates in the United States are indeed lower than in five other peer countries in Europe (France, Germany, Italy, Spain, and the United Kingdom). If smoking has such a deleterious effect on mortality in the United States, why aren't European mortality rates as high as they are here?

Part of the answer is that smoking *rates* do not capture the intensity of smoking, nor how smoking varies across different cohorts or age groups. Even though US smoking rates fell below those of the European countries in the mid-1980s, the number of cigarettes smoked each day by the typical American smoker remained substantially higher than smoking intensity in Europe well into the 21st century. The lower panel of Figure 14 shows that the average US smoker was smoking about 30 cigarettes per day—roughly 1.5 packs—until the mid-1990s. Even in the most recent year in the figure (2012), US smoking intensity remained higher than in the comparison countries.

Additionally, as discussed above, the mortality penalty for smoking is much lower for younger people, and European smoking rates are often highest among the young. In the appendix, we present country-level data from Eurostat and the CPS-TUS on smoking rates and smoking intensities for two 10-year age groups (ages 25–34 and 55–64).[26] To cite one example, the daily smoking rate among French 25–34 year-olds was a stratospheric 30 percent in 2014, while the rate for French 55–64 year-olds was only about half that (15.8 percent). Yet consistent with the relatively low smoking intensities in Europe, only 4.5 percent of French 24–34 year-olds were smoking more than 20 cigarettes per day. Other European countries

[25] See Case and Deaton (2023) for evidence of widening gaps in life expectancy at 25 between Americans and residents of 22 other rich countries.

[26] Appendix Figure A.10 disaggregates the data by age group and country, while Appendix Figure A.11 further disaggregates by education level.

show broadly similar patterns, as daily smoking rates are relatively high for younger Europeans, but these rates reflect mostly moderate levels of smoking. Once again, the European analysis shows that simple comparisons of current smoking rates across two populations can lead to misleading inferences about mortality penalties. It is the accumulated lifetime cigarette consumption (particularly at older ages) that is the primary contributor to elevated midlife mortality risk.

6.3 Smoking and Disaggregated Causes of Death

If smoking is causal with regard to the rising education-mortality gradient, then we might expect the education-based mortality differential to have risen the most for causes of death most closely related to smoking. That is what happened between 1990 and 2000, as two diseases with clear links to smoking—COPD and lung cancer—explained 21 percent of the rising mortality gap at 25–84 between adults who did and did not attend college (Meara, Richards, and Cutler 2008). Yet since the year 2000 there has been a *convergence* in midlife lung cancer deaths. Figure 15 shows the overall increase in the non-college/college gap in midlife mortality relative to 1992 for two age groups (ages 25–54 and 55–64), broken out by broad disease category.[27] For younger ages, most of the increase in the education gradient is due to differential increases in mortality from deaths of despair. For ages 55–64, major contributors to the gap include cancer deaths other than lung cancer, deaths of despair, and especially deaths from all other causes. In both panels, deaths from lung cancer contribute negatively to the change in the mortality gap.

Yet smoking is associated with many other diseases besides lung cancer, so it may still play a causal role in driving mortality differences even though targeted treatments for lung cancer have improved.[28] Figure 16 presents binned scatter plots of 1992–2019 county-level

[27]The 21 percent attributable to lung cancer and COPD in Meara, Richards, and Cutler is larger than what is seen in Figure 15 for a variety of reasons. Most importantly, their paper considers ages groups 25–84, and thus captures changes in mortality dominated by the older age groups (65–84) whose smoking histories differ from our younger population.

[28]Howlader et al. (2020) found that 2-year survival rates from non-small-cell lung cancer (the most

changes in mortality against county average smoking rates, as in figure 9, but stratified by cause of death. Each cause of death in this figure is closely associated with smoking (Carter et al. 2015), and in each panel there is a steep, positive relationship between average smoking rates and the change in non-college mortality. Among the college group, for whom smoking rates are lower and less variable across counties, no such relationship exists (and in some cases the slope is negative).

Figure 17 plots similar graphs for causes of death that are less commonly associated with smoking. Here, the strongest association is between smoking and deaths of despair. This correlation might appear to rule out smoking as a mediator of a major component of the rising education gradient in mortality.[29] Yet emerging evidence from neuroscience, particularly well-controlled mice models, has shown that nicotine reprograms reward circuits, enhancing susceptibility to later drug addiction (Kandel and Kandel 2014; Gould 2023). For example, Levine et al. (2011) demonstrated in mice that nicotine created a "primed" epigenetic environment for enhanced gene expression in response to cocaine.[30] In their study, mice exposed to nicotine were 78 percent more likely to frequent living spaces with cocaine-infused water compared to mice without nicotine exposure.

We note that there are other pathways where smoking is not directly causal, but may have amplified regional effects. For example, publicly released court documents from opioid-related litigation show that Purdue Pharma, the manufacturer of OxyContin, explicitly targeted patients with cancer and their oncologists when marketing OxyContin. And those campaigns succeeded; Arteaga and Barone (2022) show that uptake of prescription opioids was highest in commuting zones with high rates of cancer. Not surprisingly, communities with high rates of smoking also exhibit high rates of cancers (Lortet-Tieulent et al. 2016). Additionally, Evans and Lieber (2025) found that OxyContin was marketed to places where

common type of lung cancer) improved from 26 percent for men diagnosed in 2001 to 35 percent for men diagnosed in 2014. Survival rates for women rose from 35 percent to 44 percent over the same period.

[29]For example, Siddiqi et al. (2019) finds evidence consistent with rising deaths of despair among Whites whose social status is being threatened, rather than the consequence of smoking.

[30]Cocaine is not an opioid, but the broad mechanisms that support addiction show substantial similarities for both (Gould 2023; Zachariou et al. 2006).

high shares of the population received Social Security Disability Insurance (SSDI) before OxyContin became available, and again regions with high rates of smoking are also those with disproportionate rates of SSDI (Coe et al. 2011). These additional factors are likely to have boosted the influence of smoking on mortality for non-college graduates in the United States, but not for college graduates in the United States nor in European countries.

7 Conclusion

The rise in the education gradient in mortality over the last three decades is well-documented in previous studies (Meara, Richards, and Cutler 2008; Galama, Lleras-Muney, and van Kippersluis 2018). The most recent analysis (Case and Deaton 2023) shows an accelerating rise through 2019 and another jump in the gap during the COVID-19 pandemic. The significant increase in geographic mortality inequality and the emergence of a strong rural health penalty since the 1990s are two additional worrisome trends in midlife mortality. In this paper, we use county- and state-level data to show that these three trends are closely related and explained by a fundamental divergence in college and non-college patterns of mortality. For college graduates, mortality continued to decline and exhibited both sigma and beta convergence. By contrast, mortality for non-college graduates barely budged on average from 1992 to 2019, while the spatial variance of mortality grew.

We considered a variety of explanations for the worsening prospects for midlife mortality among non-college graduates. For example, previous attempts to link mortality gaps to regional business cycles or long-term employment prospects have not been successful, and our results confirm that simply residing in a "left-behind" community does not, by itself, appear to directly influence mortality. Instead, smoking was the most predictive factor in explaining the otherwise puzzling changing spatial patterns of mortality. Analogous to findings from the Chetty et al. (2016) study of income, our findings support the view that health effects related to "place" matter far more at the bottom of the distribution than at

the top. And similar to Chetty et al. (2016), our paper argues that health-related behaviors are likely an important reason that college and non-college mortality rates have diverged.

Although we have argued for a causal role for smoking in generating these patterns, the growing mortality gaps still seem too large and the causes of death too varied to blame the patterns on the adverse health effects of tobacco use alone. As noted above, smoking is likely to play a role in *amplifying* the impact of other factors adversely affecting midlife mortality, such as the marketing efforts by opioid manufacturers targeted to areas with high rates of smoking-related illness, coupled with epigenetic changes making smokers more susceptible to opioid use disorders. Still, the strength of our findings that smoking is predictive of spatial trends in midlife mortality points towards different mechanisms needed to explain the troubling trends that have unfolded since 1990.

There has been an increasing interest among economists in "place" effects on mortality (Deryugina and Molitor 2021). For example, research shows that people who move often experience life expectancies that mirror those in their destinations rather than in their previous locations (Deryugina and Molitor 2020; Finkelstein, Gentzkow, and Williams 2021). While identifying and measuring place effects in mortality is an important step forward, we do not yet know why these effects are so large. Do they result from differences in health care quality, peer effects, air pollution, or something else? Health behaviors also exhibit large and significant place effects that cannot be completely explained by the traits of people living there (Duncan, Jones, and Moon 1999; Pearce, Barnett, and Moon 2012). The strong link between place and behaviors likely reflects complex environmental interactions; behavioral responses to state-level policies likely differ across counties and economic circumstances.

How should future research proceed? First, we need new hypotheses for how social and economic forces affect (and interact with) health-related behaviors, and thereby generate mortality differences across places and populations. These forces include not only tobacco norms and policies, but also economic opportunities, the availability and cost of healthy or unhealthy food (Cawley and Frisvold 2023), infrastructure that fosters or inhibits physi-

cal activity, and the supply and safeguards around dangerous substances (e.g, fentanyl) or firearms (e.g., number, safety features, and secure storage). Second, we need a better understanding of how peer effects in health-related behaviors can generate significant variance in mortality across places.[31] Identifying the forces that contribute to health-related behaviors, and how peer effects in these behaviors can magnify mortality differences, are potentially fruitful paths for future research.

[31]See Powell, Tauras, and Ross (2005) and Christakis and Fowler (2007) for evidence of peer effects in smoking and obesity, respectively. Angrist (2014) documents the challenges in estimating peer effects.

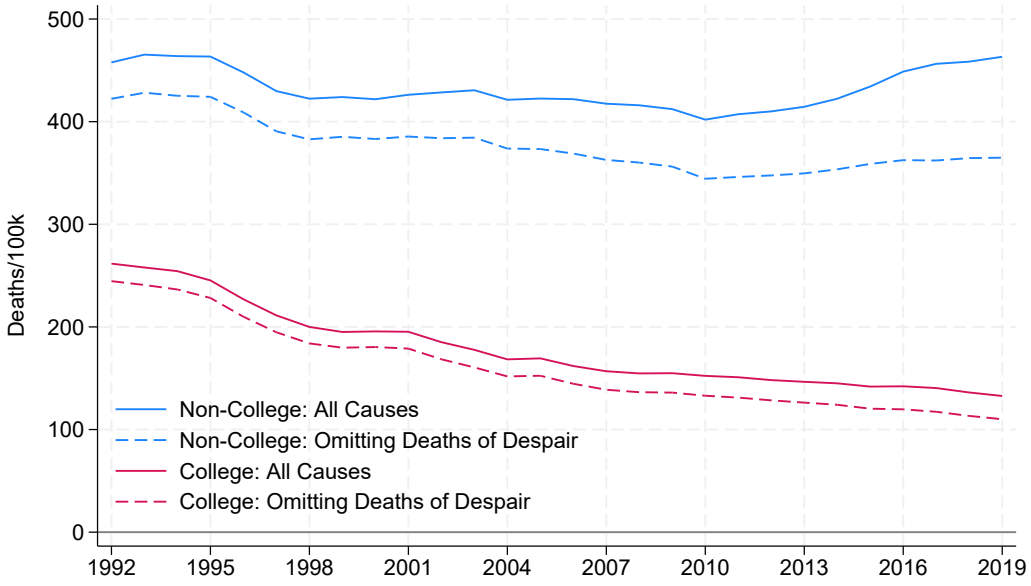

Figure 1. National Midlife Mortality Rates by Education: 1992–2019. Mortality rates are age-adjusted deaths per 100,000 population among adults aged 25–64, standardized using the 2000 US Census age distribution. The college population includes adults with a bachelor's degree or more education; the non-college population includes the remaining adults.

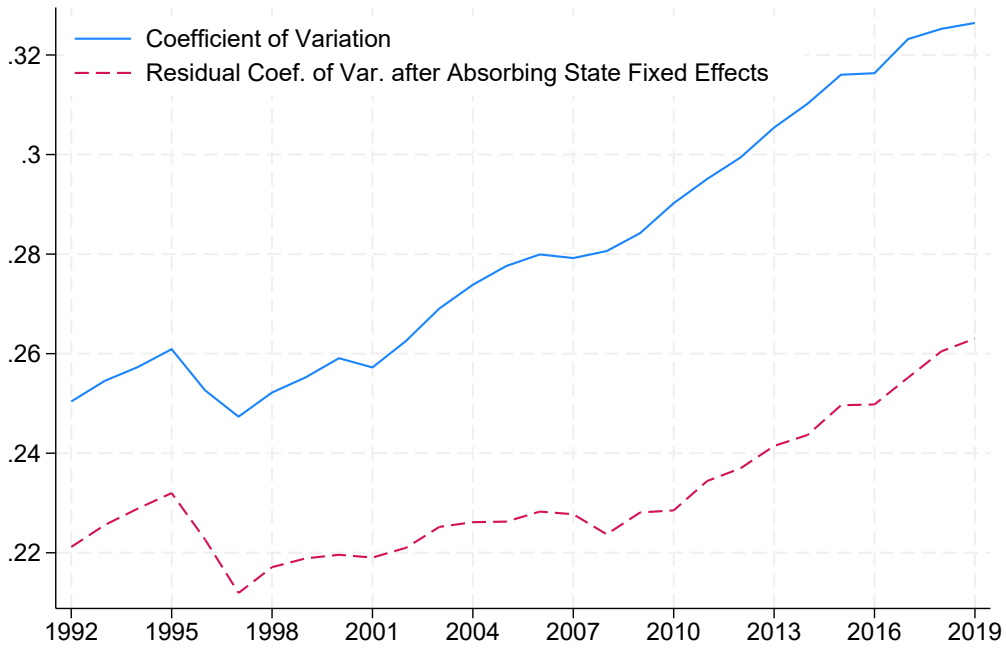

Figure 2. County-Level Coefficients of Variation for Midlife Mortality Rates: 1992–2019. Coefficients of variation are based on county-level age-adjusted deaths per 100,000 population among adults aged 25–64. Data are weighted by county-level population.

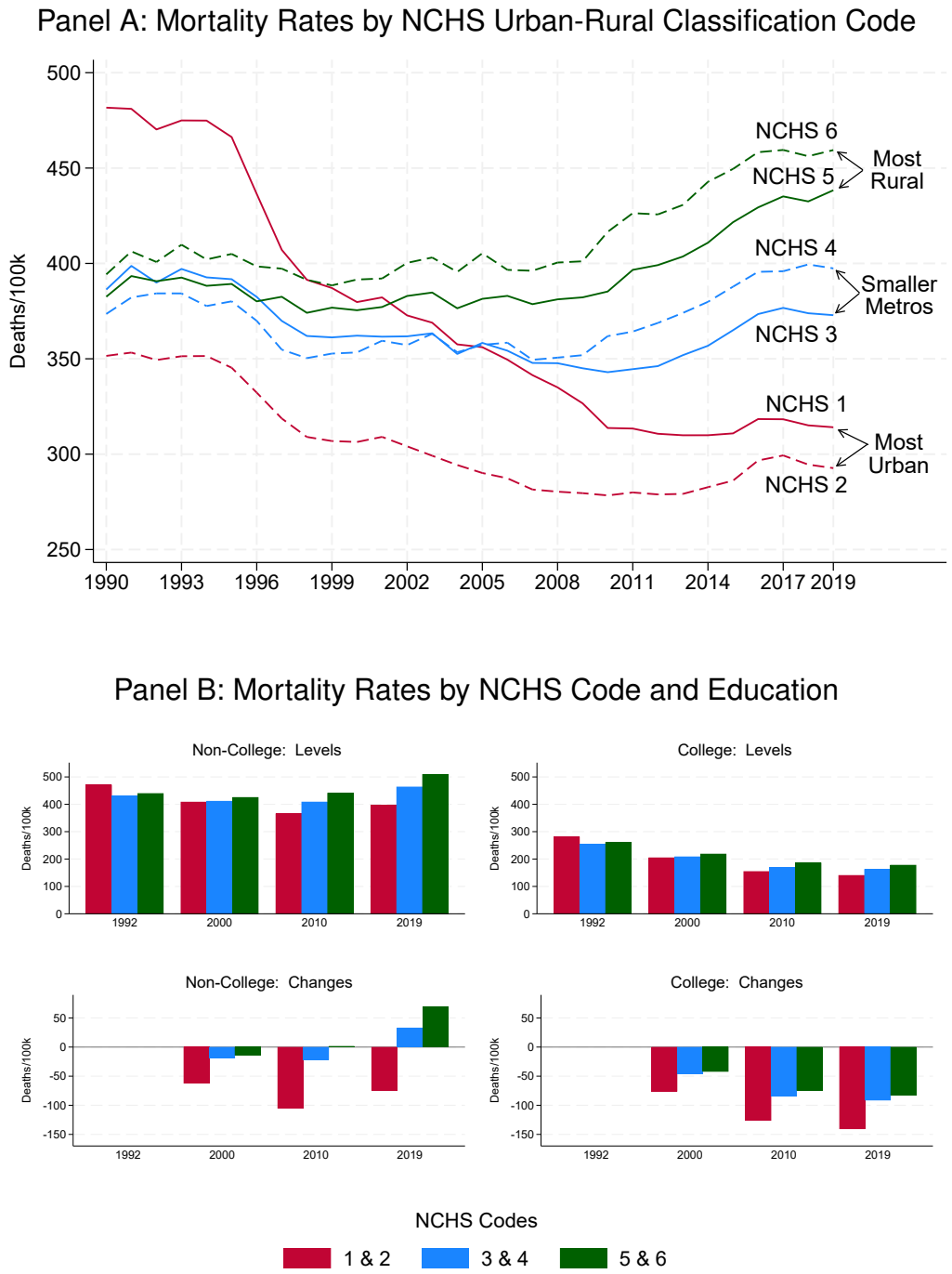

Figure 3. The Rural-Urban Reversal in Midlife Mortality: Mortality Rates Rates by County NCHS Code and Education. Mortality rates are age-adjusted deaths per 100,000 population among adults aged 25–64. NCHS codes correspond to the 2013 Urban-Rural Classification Scheme from the National Center for Health Statistics. See Appendix Figure A.5 for a county-level map of the NCHS codes.

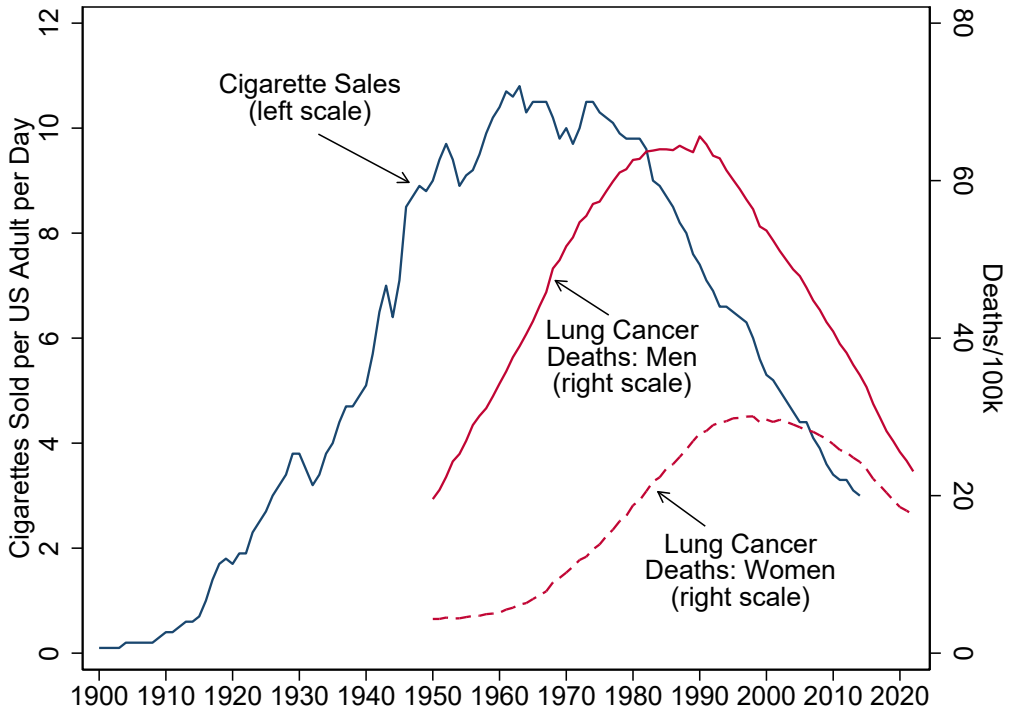

Figure 4. Cigarette Sales and Lung Cancer Deaths: 1900–2021. Left axis shows cigarette sales per person 15 and older; right axis shows deaths due to lung cancer per 100,000 population aged 25–64. Source: Our World in Data (Ritchie and Roser 2023).

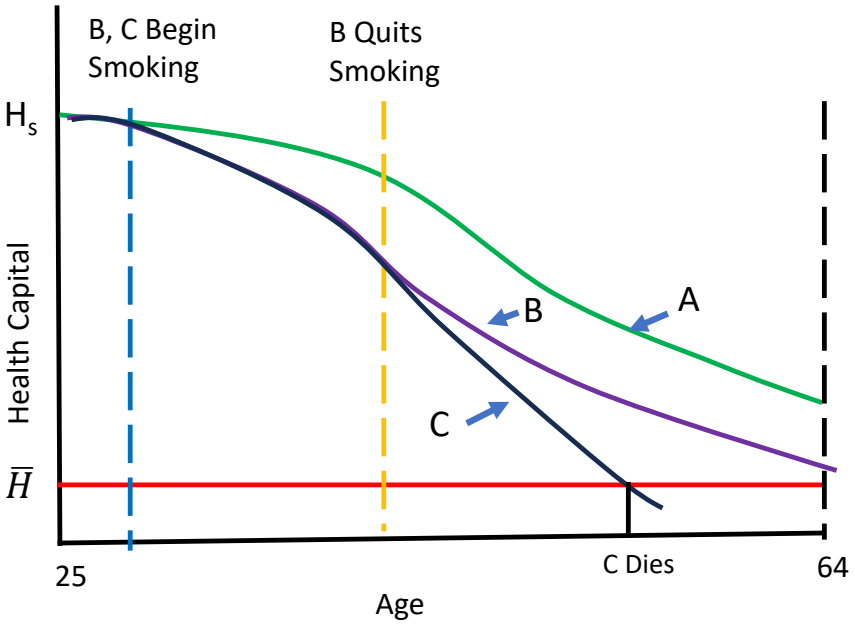

Figure 5. Health Capital for Three Illustrative Individuals: Never Smoked (A), Sometimes Smoked (B), and Always Smoked (C) All three individuals begin at age 25 with H_0. Individuals B and C take up smoking and B later quits; C always smokes. $\bar{H}$ denotes the health level below which the individual dies. Only C dies but B is in poor shape by age 64 and is likely to predecease A.

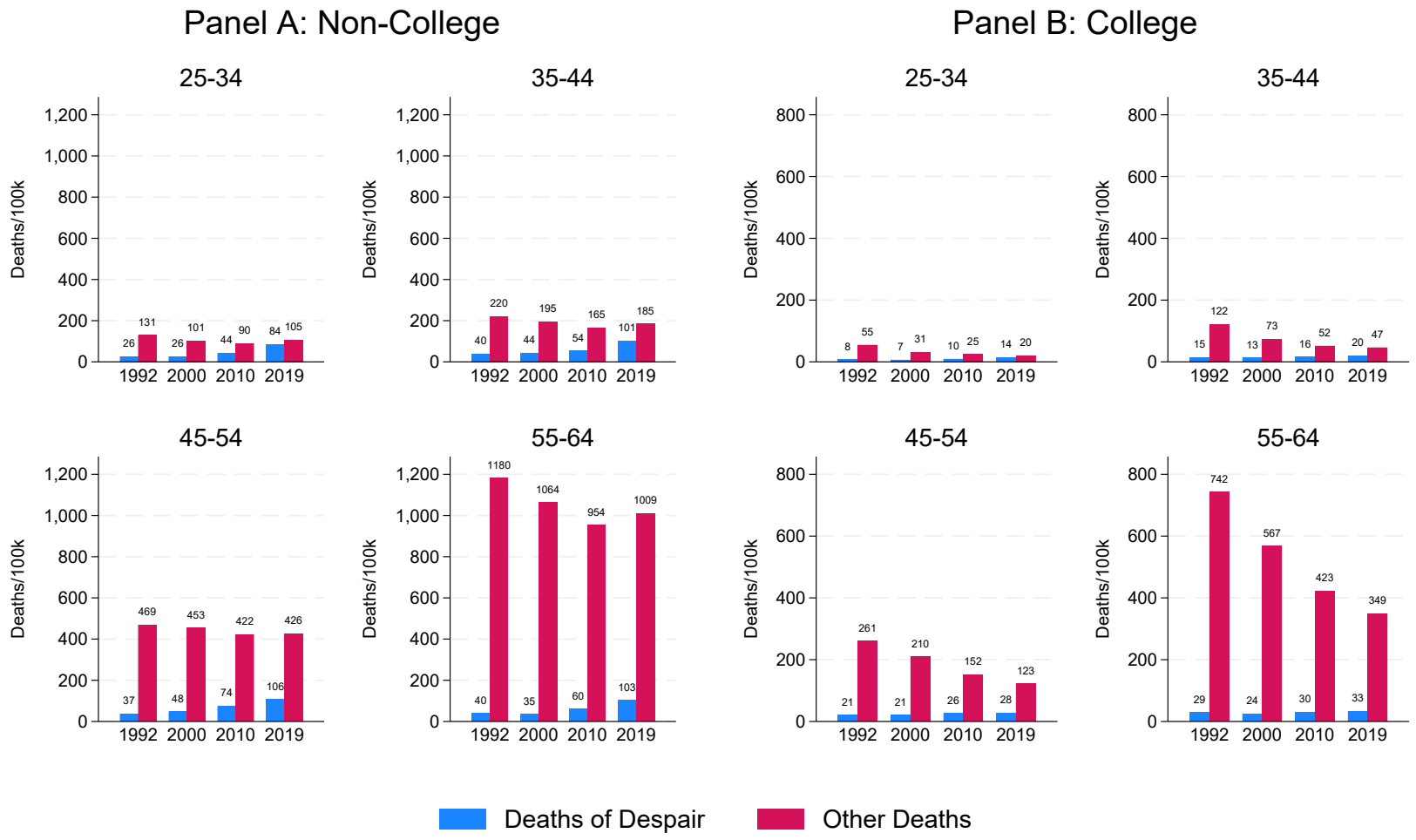

Figure 6. National Mortality Rates by 10-year Age Group and Education: 1992–2019. Vertical scales differ for the non-college population (Panel A) and the college population (Panel B).

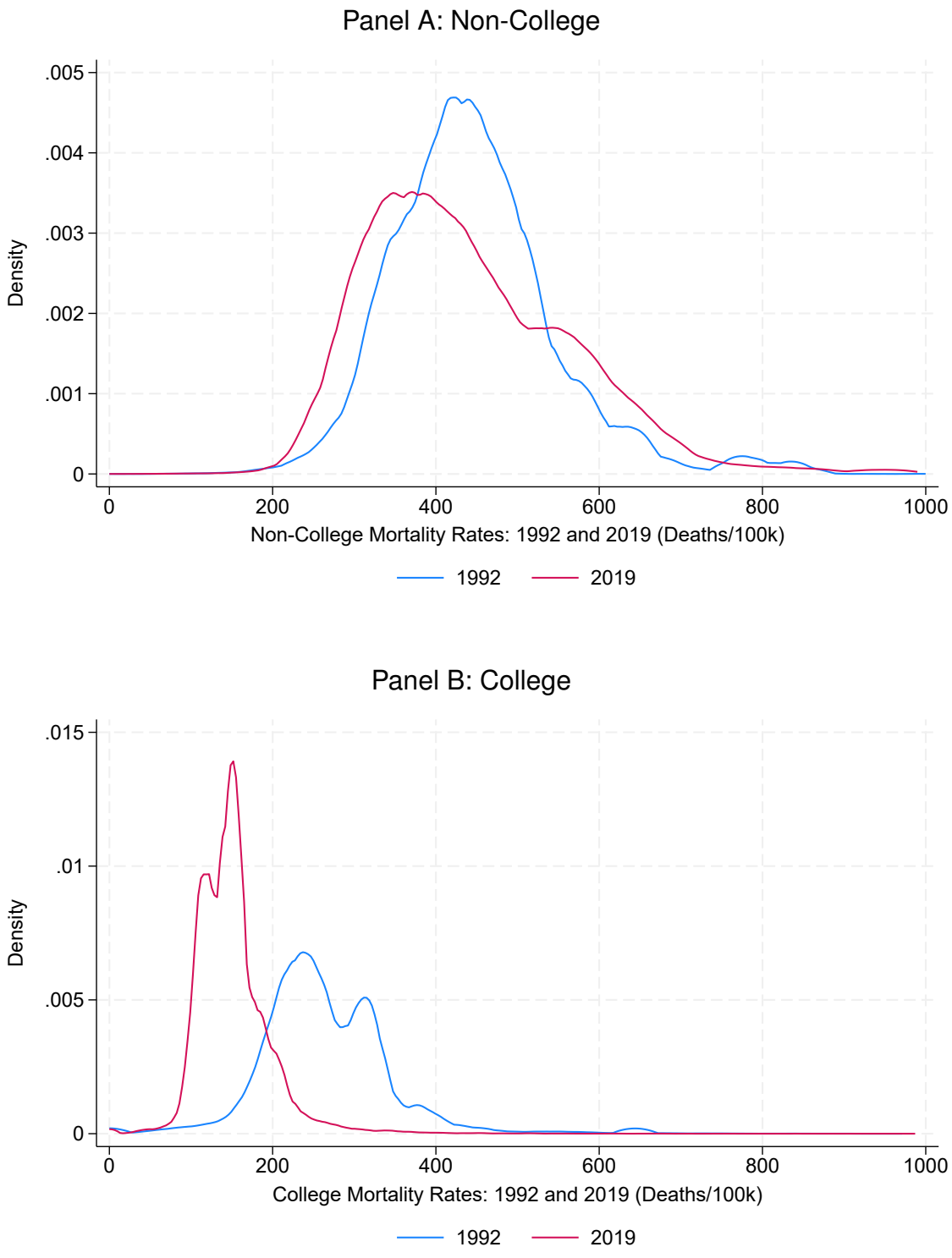

Figure 7. County-Level Kernel Densities of Midlife Mortality Rates per 100,000 Population in 1992 and 2019 by Education. Mortality rates are age-adjusted deaths per 100,000 population among adults aged 25–64. Densities are weighted by education-specific population in the given years.

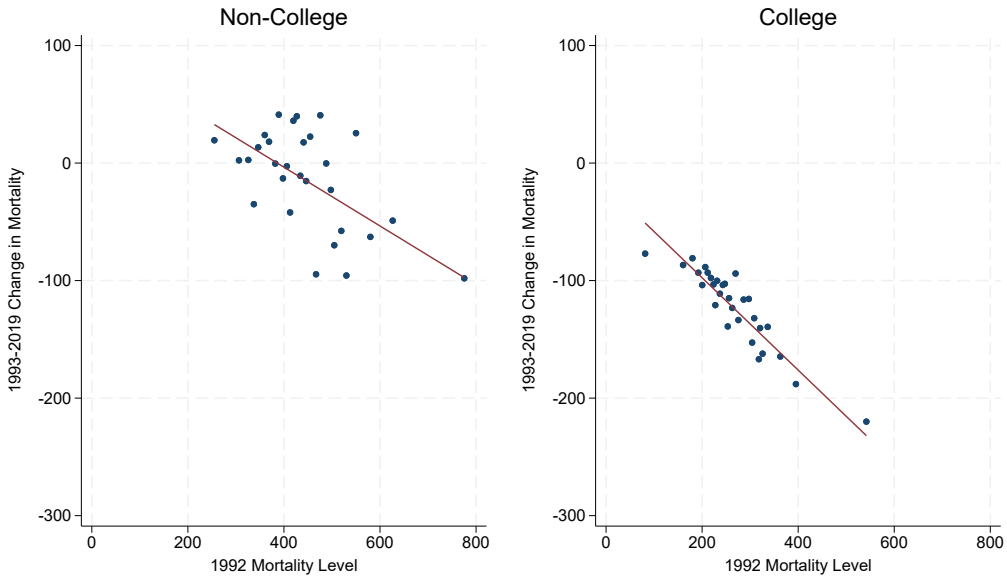

Figure 8. County-Level Correlations of Education-Specific Mortality-Rate Changes with Initial (1992) Mortality Rates. Data are weighted by average education-specific population in 1993 and 2019. Because measurement error in the 1992 mortality rate can engender a spurious negative correlation between the 1992 rate and the 1992–2019 change, we plot 1993–2019 changes on the vertical axis.

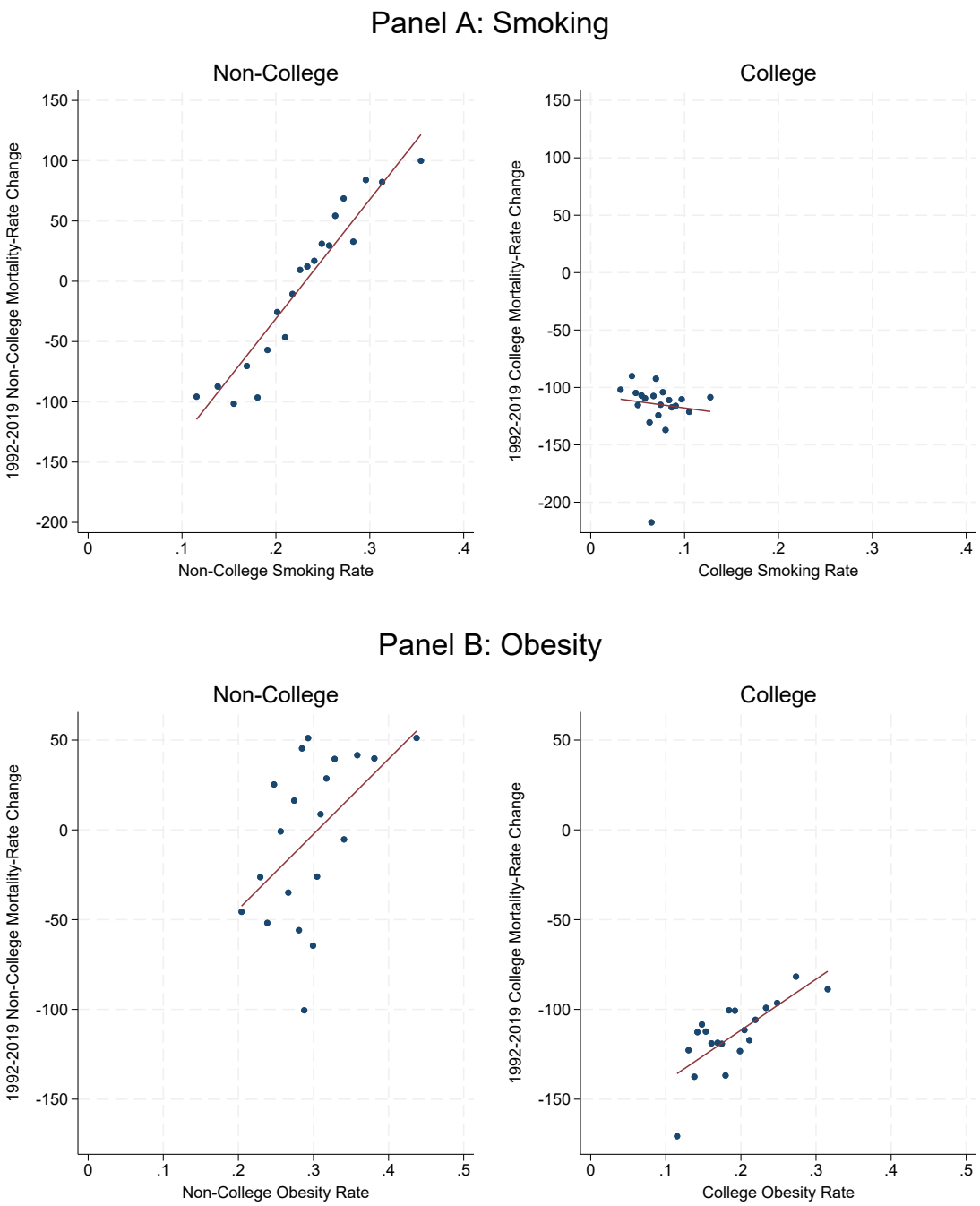

Figure 9. County-Level Correlations of Education-Specific Mortality-Rate Changes with Smoking and Obesity Rates. Data are weighted by average education-specific population in 1992 and 2019. Smoking and obesity rates are derived from individual-level responses collected by the Behavioral Risk Factor Surveillance System (BRFSS) and correspond to averages within the county and educational group from 1996 to 2010. Vertical and horizontal scales for Panel A are different from those for Panel B.

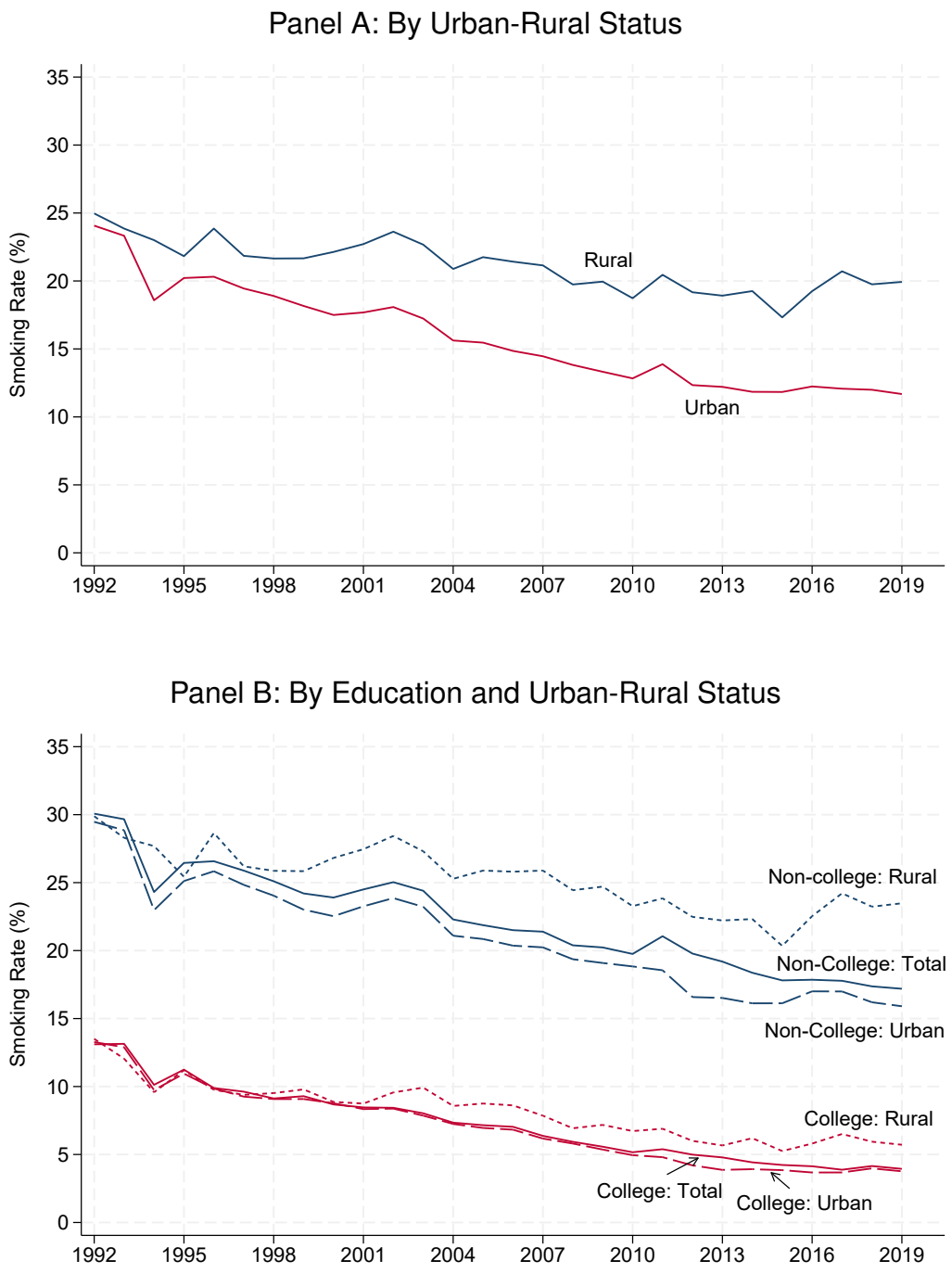

Figure 10. Smoking Rates by Urban-Rural Status and Education. Urban counties correspond to NCHS codes 1-4; rural counties correspond to codes 5 and 6. Smoking rates are from the Behavioral Risk Factor Surveillance System (BRFSS).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
1992 Mortality Rate	–22.14***	–21.99***	–31.94***	–28.23***	–28.50***	–22.47***	–21.60***	–22.06***	–27.02***	–38.65***	–36.20***
	(5.161)	(5.158)	(4.848)	(5.117)	(4.993)	(5.176)	(5.131)	(5.247)	(5.271)	(4.222)	(5.667)
Avg. 1-yr Outmig Rate (2016-20)		4.019								–0.429	–3.717
		(5.937)								(3.424)	(3.474)
Smoking Rate			53.84***							48.86***	41.35***
			(4.952)							(4.542)	(6.647)
Obesity Rate				18.07***						7.792*	9.289**
				(4.289)						(3.712)	(2.826)
Ln 1992 Per Capita Income					–20.24***					–15.17**	–9.334
					(4.855)					(5.251)	(5.573)
1993-2019 Chg Ln Per Cap Income					–25.34***					–14.30***	–15.48***
					(4.249)					(2.270)	(2.705)
1992 Emp-Pop Ratio						–9.012**				3.502	–0.462
						(3.169)				(3.312)	(3.397)
1993-2019 Chg Emp-Pop Ratio						–13.53*				11.02*	8.524+
						(6.351)				(4.471)	(4.386)
1992 Manufacturing Empl Share							13.60***			2.823	0.129
							(3.382)			(3.410)	(3.734)
China Shock								1.409		–0.767	–2.989
								(3.889)		(2.758)	(2.706)
State Policy-Liberalism Index									–24.53*	–6.097	
									(11.18)	(4.437)	
NCHS Codes 3-4	79.75***	78.92***	51.30***	71.40***	50.34***	67.01***	76.48***	79.50***	68.62***	36.95***	35.47***
	(9.647)	(9.046)	(8.485)	(10.07)	(11.76)	(9.112)	(10.51)	(9.813)	(11.80)	(9.114)	(9.803)
NCHS Codes 5-6	116.3***	115.1***	80.07***	95.90***	77.15***	96.41***	105.3***	115.6***	100.9***	54.99***	50.74***
	(16.75)	(15.19)	(10.70)	(18.22)	(18.59)	(14.08)	(17.42)	(17.45)	(14.71)	(10.20)	(9.759)
Constant	–58.61***	–58.18***	–44.23***	–52.91**	–43.49**	–51.62***	–55.92**	–58.42**	–52.80***	–35.95***	4.612
	(16.56)	(15.73)	(5.938)	(16.69)	(14.65)	(14.59)	(16.17)	(16.78)	(13.96)	(6.764)	(8.792)
N	2081	2081	2081	2081	2081	2081	2081	2081	2081	2081	2081
R-sq	0.243	0.244	0.459	0.263	0.322	0.261	0.257	0.243	0.288	0.499	0.569
State Fixed Effects?	N	N	N	N	N	N	N	N	N	N	Y

Table 1. County-Level Mortality-Change Regressions, with Standardized Coefficients and NCHS Urban-Rural Dummies: Non-College. + 0.10 * 0.05 ** 0.01 *** 0.001. The dependent variable is 1993–2019 mortality-rate change for the county's non-college population. Standard errors are clustered by state and the data are weighted by education-specific average population in 1993 and 2019. The dependent variable is 1993–2019 mortality-rate change for the county's non-college population. Standard errors are clustered by state and the data are weighted by education-specific average population in 1993 and 2019. All regressors except the NCHS indicator variables are standardized. The average 1-year out-migration rate is a 5-year average from the 2016–2020 American Community Survey. The China shock variable is from Autor, Dorn, and Hanson (2013), and the state policy-liberalism index is from Caughey and Warshaw (2016). Because the regressors are standardized, coefficients denote effects of 1-standard-deviation increases in the regressors.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
1992 Mortality Rate	–33.09***	–33.39***	–34.34***	–34.30***	–32.31***	–31.51***	–32.17***	–33.02***	–33.44***	–30.20***	–26.17***
	(7.518)	(7.534)	(8.146)	(7.064)	(6.827)	(6.156)	(7.424)	(7.532)	(7.671)	(6.275)	(6.618)
Avg. 1-yr Outmig Rate (2016-20)		–3.891								1.227	–1.715
		(2.818)								(3.320)	(3.989)
Smoking Rate			4.499							–0.687	–1.425
			(3.621)							(1.970)	(2.336)
Obesity Rate				14.66***						10.27***	6.865*
				(3.997)						(2.546)	(2.707)
Ln 1992 Per Capita Income					–0.215					9.722***	8.820**
					(2.797)					(2.166)	(2.709)
1993-2019 Chg Ln Per Cap Income					–16.75***					–14.10***	–12.36***
					(2.072)					(1.677)	(1.822)
1992 Emp-Pop Ratio						–8.112**				–6.031+	–8.432**
						(2.838)				(3.020)	(2.832)
1993-2019 Chg Emp-Pop Ratio						–8.588*				0.943	–0.915
						(3.704)				(2.955)	(3.357)
1992 Manufacturing Empl Share							8.465***			6.329**	5.303**
							(1.765)			(1.842)	(1.853)
China Shock								1.660		1.554	3.343
								(1.782)		(1.975)	(2.138)
State Policy-Liberalism Index									–3.806	–0.443	
									(5.530)	(4.092)	
NCHS Codes 3-4	22.61***	23.28***	21.29***	13.83**	14.27*	14.06**	20.78**	22.24***	20.91***	16.41**	16.84**
	(6.023)	(6.271)	(5.376)	(4.820)	(5.371)	(4.575)	(6.775)	(6.261)	(5.144)	(5.535)	(5.653)
NCHS Codes 5-6	32.18***	33.14**	30.64***	9.735	26.98***	17.42**	25.17*	31.25**	30.03***	19.43**	17.72*
	(8.968)	(9.530)	(8.048)	(6.665)	(7.448)	(5.855)	(9.577)	(9.438)	(6.872)	(7.027)	(7.991)
Constant	–132.2***	–132.5***	–131.8***	–128.1***	–129.6***	–128.8***	–131.2***	–132.1***	–131.6***	–129.6***	–115.3***
	(6.528)	(6.890)	(5.835)	(4.670)	(3.901)	(5.325)	(6.650)	(6.676)	(5.487)	(3.449)	(4.166)
N	2081	2081	2081	2081	2081	2081	2081	2081	2081	2081	2081
R-sq	0.207	0.209	0.210	0.234	0.250	0.225	0.218	0.207	0.209	0.271	0.308
State Fixed Effects?	N	N	N	N	N	N	N	N	N	N	Y

Table 2. County-Level Mortality-Change Regressions, with Standardized Coefficients and NCHS Urban-Rural Dummies: College. + 0.10 * 0.05 ** 0.01 *** 0.001. The dependent variable is 1993–2019 mortality-rate change for the county's college population. Standard errors are clustered by state and the data are weighted by education-specific average population in 1993 and 2019. See the notes to Table 1 for a description of the regressors. All regressors except the NCHS indicator variables are standardized.

	(1) Non-Coll	(2) Non-Coll	(3) Non-Coll	(4) Non-Coll	(5) College	(6) College	(7) College	(8) College
Ln 1992 State Per-Capita Income	–337.2***		–130.6	–148.0	–70.80		–85.67	–72.80
	(93.29)		(82.91)	(86.32)	(73.41)		(96.98)	(118.2)
1992-2019 Chg. in Ln State Income	–587.7**		–19.61	58.27	–94.48		–138.5	–96.70
	(175.9)		(139.3)	(125.9)	(97.64)		(120.6)	(146.4)
1992 Non-College Smoking Rate		977.1***	843.7**	978.9**				
		(177.4)	(251.4)	(278.9)				
1992–2019 Chg in Non-College Smoking Rate		1369.3***	1192.3***	1214.4***				
		(177.1)	(143.3)	(134.0)				
1992 Non-College Obesity Rate				–536.3				
				(325.2)				
1992–2019 Chg in Non-College Obesity Rate				160.7				
				(175.6)				
1992 College Smoking Rate						412.5	–448.5	–716.2
						(457.4)	(697.1)	(701.6)
1992–2019 Chg in College Smoking Rate						488.3	–357.0	–714.7
						(461.2)	(887.7)	(945.6)
1992 College Obesity Rate								–142.9
								(378.1)
1992–2019 Chg in College Obesity Rate								236.0
								(165.8)
Constant	3803.1***	–106.1	1280.5	1443.4	652.0	–141.3**	854.4	678.9
	(969.6)	(59.65)	(919.3)	(953.0)	(753.0)	(43.58)	(1030.4)	(1286.3)
N	44	44	44	44	44	44	44	44
R-sq	0.512	0.728	0.759	0.782	0.0890	0.0207	0.0984	0.136

Table 3. State-Level Mortality-Change Regressions. The dependent variable for each regression is the state-level mortality-rate change from 1992 to 2019 for non-college (columns 1–4) or college (columns 5–8) populations. Data are weighted by average education-specific population in 1992 and 2019 and robust standard errors are in parentheses. State-level smoking rates for 1992 and for 2019 (used for the 1992–2019 change) are calculated from BRFSS.

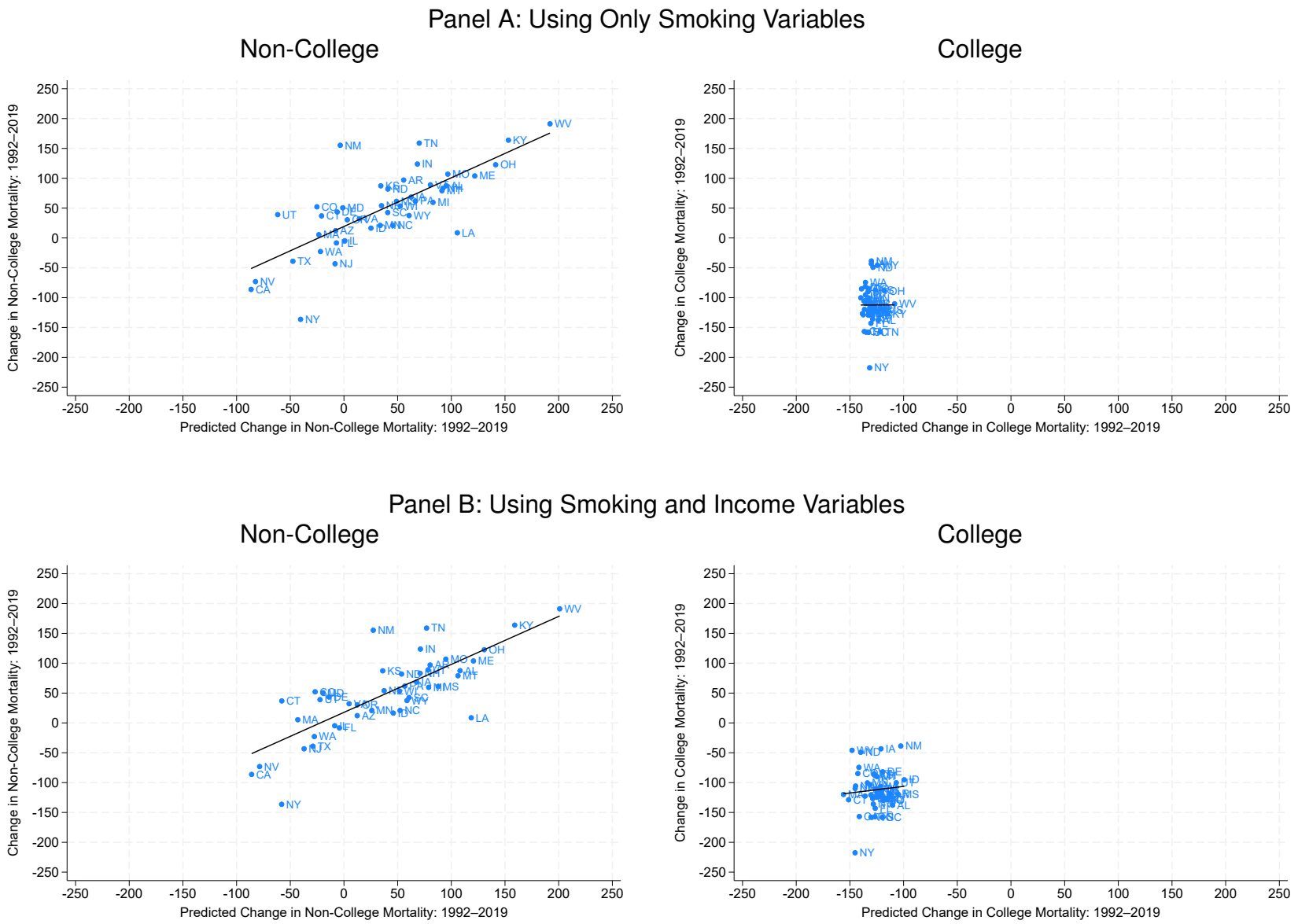

Figure 11. Predicted and Actual Values for 1992–2019 State-Level Education-Specific Mortality-Rate Changes. Predicted values for Panel A use the regression models in columns 2 (for non-college) and 6 (for college) in Table 3. These models use only the 1992 level and the 1992–2019 change in education-specific smoking rates. Panel B uses predictions from the models in columns 3 and 7, which add the 1992 level and 1992–2019 change in state per-capita income to the regression models.

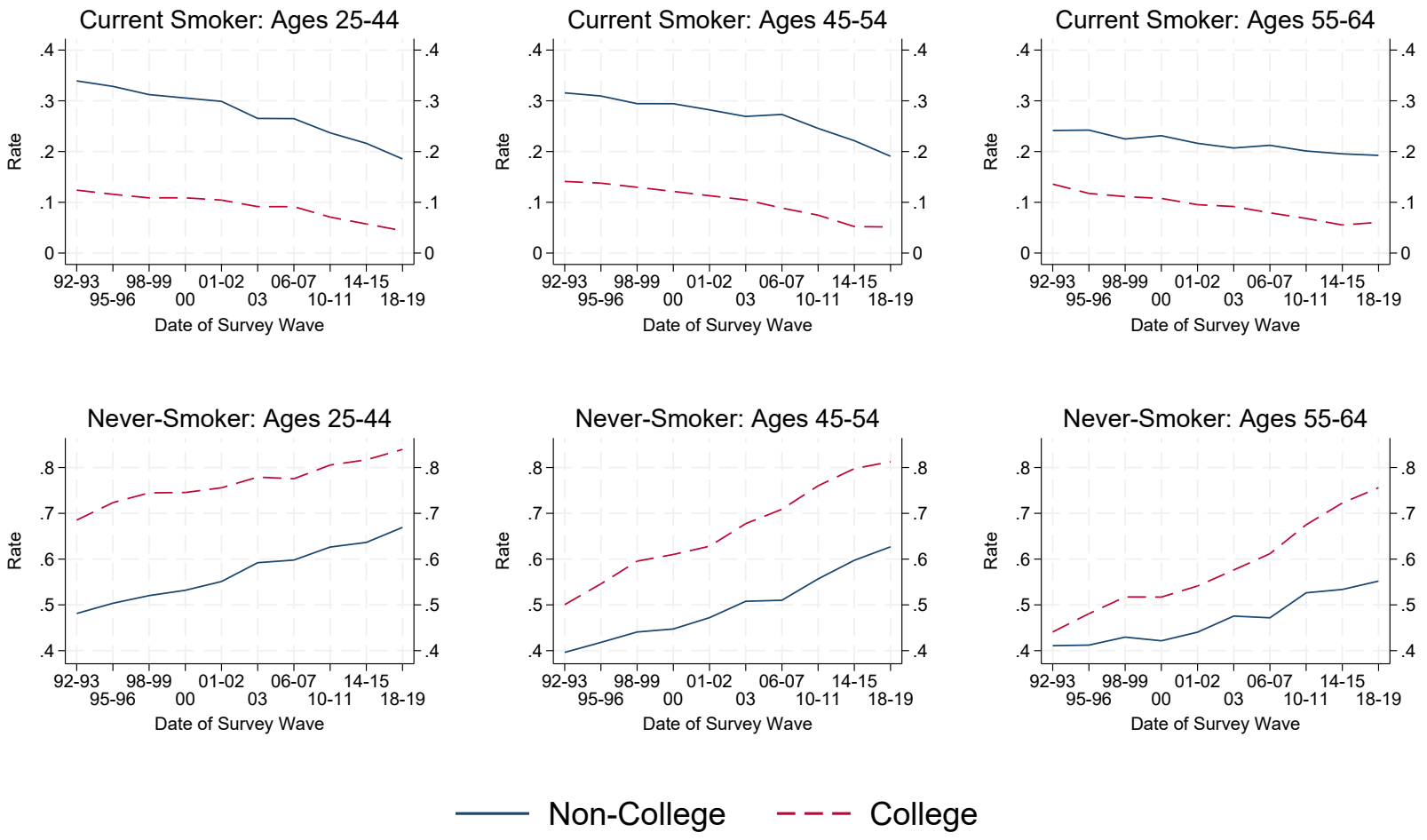

Figure 12. Current and Never-Smoking Rates at Midlife by Education and Age Group. Source: Current Population Survey Tobacco-Use Supplements.

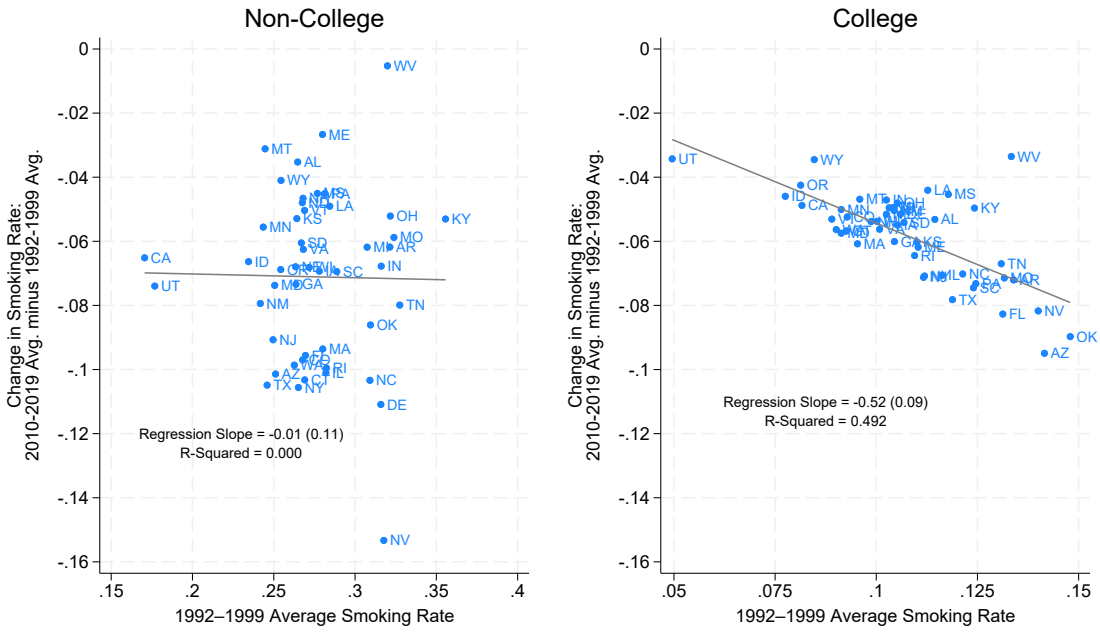

Figure 13. State-Level Changes in Current-Smoking Rate at Midlife and Initial Current-Smoking Rate by Education. *Notes:* Data are age-adjusted and regressions are unweighted. *Source:* Behavioral Risk Factor Surveillance System (BRFSS) dataset, years 1992–1999 and 2010–2019.

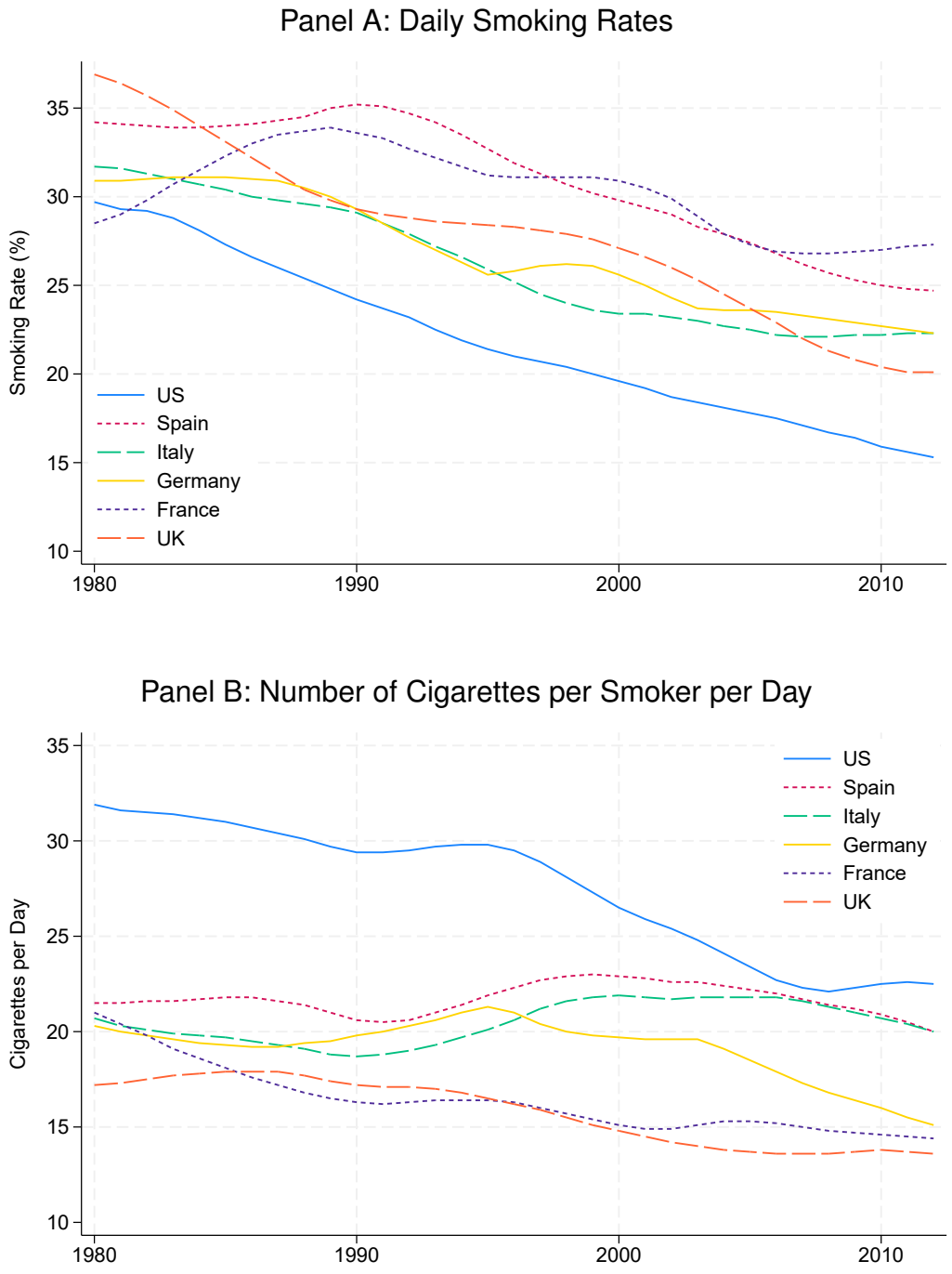

Figure 14. Daily Smoking Rates and Number of Cigarettes Smoked per Day in Six Countries: 1980–2012. *Source:* Institute for Health Metrics and Evaluation (IHME) (2014).

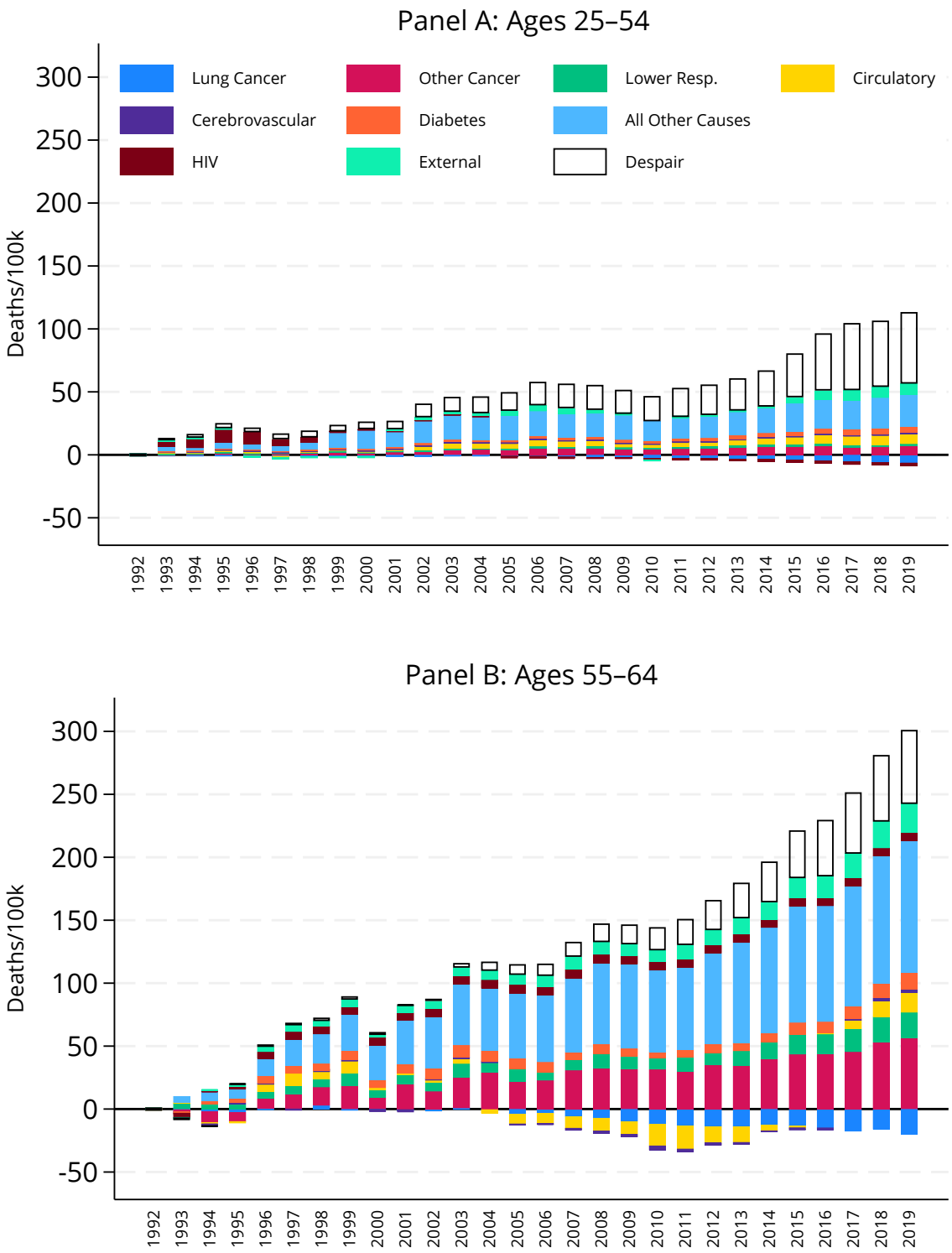

Figure 15. Contributions of Mortality Causes to Growth in National Mortality Gaps Since 1992: 25–54 Year-Olds and 55–64 Year-Olds. Bars above the horizontal axis denote causes of death that have contributed positively to the rising educational mortality gap since 1992; that is, the change in the cause-specific mortality rate for the non-college population has been higher than the corresponding change for the college population. Bars below the axis denote mortality causes that have risen for college populations more than non-college populations. The total height of bars above the horizontal axis, less the height of bars below the axis, equals the change in the mortality gap from 1992 to the given year.

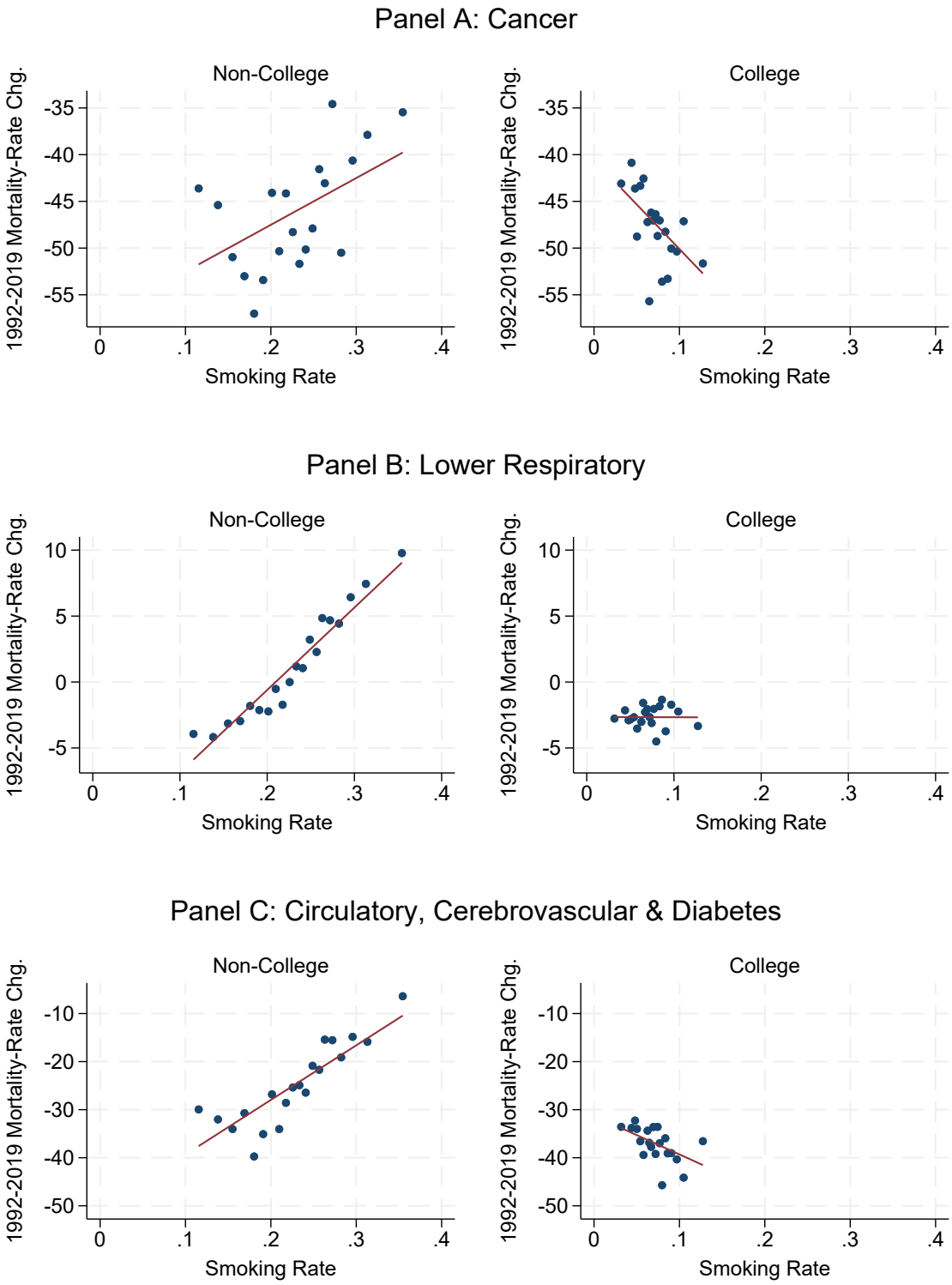

Figure 16. County-Level Correlations of Cause-Specific Mortality-Rate Changes with Smoking Rates: Causes Closely Related to Smoking. Data are weighted by average education-specific population in 1992 and 2019. Smoking and obesity rates are derived from individual-level responses collected by the Behavioral Risk Factor Surveillance System (BRFSS) and correspond to averages within the county and educational group from 1996 to 2010. Vertical scales are different across panels.

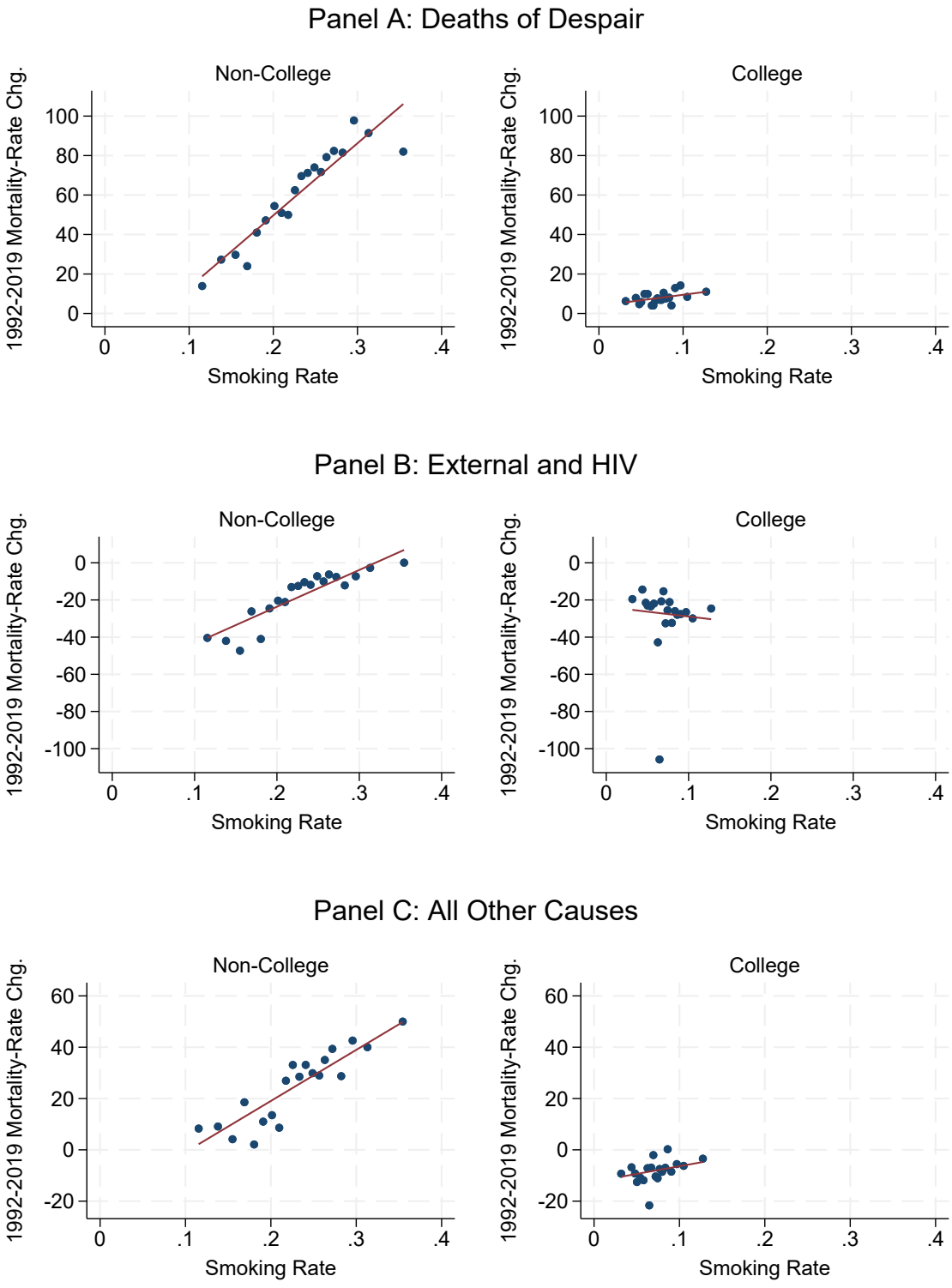

Figure 17. County-Level Correlations of Cause-Specific Mortality-Rate Changes with Smoking Rates: Causes Less Closely Related to Smoking. Data are weighted by average education-specific population in 1992 and 2019. Smoking and obesity rates are derived from individual-level responses collected by the Behavioral Risk Factor Surveillance System (BRFSS) and correspond to averages within the county and educational group from 1996 to 2010. Vertical scales are different across panels. "All Other Causes" in Panel C denotes causes that are included neither in Figure 16 nor in Panels A or B of this figure.

References

Angrist, Joshua D. 2014. "The Perils of Peer Effects." *Labour Economics* 30: 98–108.

Arteaga, Carolina, and Victoria Barone. 2022. "A Manufactured Tragedy: The Origins and Deep Ripples of the Opioid Epidemic." University of Toronto Working Paper. Available at http://www.carolinaarteaga.com/s/Opioids_ArteagaBarone_July2022-l76k.pdf,.

Autor, David H., David Dorn, and Gordon H. Hanson. 2013. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *American Economic Review* 103(6): 2121–2168.

Baltrus, Peter, Khusdeep Malhotra, George Rust, Robert Levine, Chaohua Li, and Anne H. Gaglioti. 2019. "Identifying County-Level All-Cause Mortality Rate Trajectories and Their Spatial Distribution Across the United States." *Preventing Chronic Disease* 16. Available at https://www.cdc.gov/pcd/issues/2019/18_0486.htm.

Beckfield, Jason, and Clare Bambra. 2016. "Shorter Lives in Stingier States: Social Policy Shortcomings Help Explain the US Mortality Disadvantage." *Social Science & Medicine* 171: 30–38.

Carter, Brian D., Christian C. Abnet, Diane Feskanich, Neal D. Freedman, Patricia Hartge, Cora E. Lewis, Judith K. Ockene, Ross L. Prentice, Frank E. Speizer, Michael J. Thun, and Eric J. Jacobs. 2015. "Smoking and Mortality—Beyond Established Causes." *New England Journal of Medicine* 372(7): 631–640.

Case, Anne, and Angus Deaton. 2005. "Broken Down by Work and Sex: How Our Health Declines." In *Analyses in the Economics of Aging*, David A. Wise, ed., 185–212. University of Chicago Press.

Case, Anne, and Angus Deaton. 2017. "Mortality and Morbidity in the 21st Century." *Brookings Papers on Economic Activity* 2017(Spring): 397–443.

Case, Anne, and Angus Deaton. 2021. "Life Expectancy in Adulthood Is Falling for Those Without a BA Degree, but as Educational Gaps Have Widened, Racial Gaps Have Narrowed." *Proceedings of the National Academy of Sciences* 118(11). Available at https://www.pnas.org/doi/epdf/10.1073/pnas.2024777118.

Case, Anne, and Angus S. Deaton. 2015. "Rising Morbidity and Mortality in Midlife Among White Non-Hispanic Americans in the 21st Century." *Proceedings of the National Academy of Sciences* 112(49): 15078–15083.

Case, Anne, and Angus S. Deaton. 2023. "Accounting for the Widening Mortality Gap Between American Adults With and Without a BA." *Brookings Papers on Economic Activity* 2023(Fall): 1–44.

Caughey, Devin, and Christopher Warshaw. 2016. "The Dynamics of State Policy Liberalism, 1936–2014." *American Journal of Political Science* 60(4): 899–913.

Cawley, John, and David Frisvold. 2023. "Review: Taxes on Sugar-Sweetened Beverages: Political Economy, and Effects on Prices, Purchases, and Consumption." *Food Policy* 117(May): 102441.

Centers for Disease Control and Prevention. 2014. "2013 NCHS Urban-Rural Classification Scheme for Counties." National Center for Health Statistics, *Vital and Health Statistics*. Series 2, No. 166. Available at https://www.cdc.gov/nchs/data/series/sr_02/sr02_166.pdf.

Chetty, Raj, Michael Stepner, Sarah Abraham, Shelby Lin, Benjamin Scuderi, Nicholas Turner, Augustin Bergeron, and David Cutler. 2016. "The Association Between Income and Life Expectancy in the United States, 2001–2014." *Journal of the American Medical Association* 315(16): 1750–1766.

Christakis, Nicholas A., and James H. Fowler. 2007. "The Spread of Obesity in a Large Social Network Over 32 Years." *New England Journal of Medicine* 357(4): 370–379.

Coe, Norma, Kelly Haverstick, Alicia H. Munnell, and Anthony Webb. 2011. "What Explains State Variation in SSDI Application Rates?" Boston College Center for Retirement Research at Working Paper No. 2011-23.

Couillard, Benjamin K., Christopher L. Foote, Kavish Gandhi, Ellen Meara, and Jonathan Skinner. 2021. "Rising Geographic Disparities in US Mortality." *Journal of Economic Perspectives* 35(4): 123–146.

Currie, Janet. 2011. "Inequality at Birth: Some Causes and Consequences." *American Economic Review* 101(3): 1–22.

Cutler, David M., and Edward Glaeser. 2005. "What Explains Differences in Smoking, Drinking, and Other Health-Related Behaviors?" *American Economic Review* 95(2): 238–242.

Cutler, David M., Fabian Lange, Ellen Meara, Seth Richards-Shubik, and Christopher J. Ruhm. 2011. "Rising Educational Gradients in Mortality: The Role of Behavioral Risk Factors." *Journal of Health Economics* 30(6): 1174–1187.

de Walque, Damien. 2007. "Does Education Affect Smoking Behaviors? Evidence Using the Vietnam Draft as an Instrument for College Education." *Journal of Health Economics* 26(5): 877–895.

Deryugina, Tatyana, and David Molitor. 2020. "Does When You Die Depend on Where You Live? Evidence from Hurricane Katrina." *American Economic Review* 110(11): 3602–33. doi:10.1257/aer.20181026.
Available at https://www.aeaweb.org/articles?id=10.1257/aer.20181026.

Deryugina, Tatyana, and David Molitor. 2021. "The Causal Effects of Place on Health and Longevity." *Journal of Economic Perspectives* 35(4): 147–170.

Dowd, Jennifer B., Jeremy Albright, Trivellore E. Raghunathan, Robert F. Schoeni, Felicia Leclere, and George A. Kaplan. 2011. "Deeper and Wider: Income and Mortality in the USA over Three Decades." *International Journal of Epidemiology* 40(1): 183–188.

Duncan, Craig, Kelvyn Jones, and Graham Moon. 1999. "Smoking and Deprivation: Are There Neighbourhood Effects?" *Social Science & Medicine* 48(4): 497–505.

Evans, William N, and Ethan M.J. Lieber. 2025. "Prescription for Disaster: The SSDI Rate, Pain, and Prescribing Practices." National Bureau of Economic Research Working Paper No. 34265.

Farrell, Phillip, and Victor R. Fuchs. 1982. "Schooling and Health: The Cigarette Connection." *Journal of Health Economics* 1(3): 217–230.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams. 2021. "Place-Based Drivers of Mortality: Evidence from Migration." *American Economic Review* 111(8): 2697–2735.

Galama, Titus J. 2015. "A Contribution to Health-Capital Theory." CESR-Schaeffer Working Paper Working Paper No. 2015-004.

Galama, Titus J., Adriana Lleras-Muney, and Hans van Kippersluis. 2018. "The Effect of Education on Health and Mortality: A Review of Experimental and Quasi-Experimental Evidence." National Bureau of Economic Research Working Paper No. 24225.

Gould, Thomas J. 2023. "Epigenetic and Long-Term Effects of Nicotine on Biology, Behavior, and Health." *Pharmacological Research* 192: 106741.

Grossman, Michael. 1972. "On the Concept of Health Capital and the Demand for Health." *Journal of Political Economy* 80(2): 223–255.

Howlader, Nadia, Gonçalo Forjaz, Meghan J. Mooradian, Rafael Meza, Chung Yin Kong, Kathleen A. Cronin, Angela B. Mariotto, Douglas R. Lowy, and Eric J. Feuer. 2020. "The effect of Advances in Lung-Cancer Treatment on Population Mortality." *New England Journal of Medicine* 383(7): 640–649.

Hoxby, Caroline M. 2023. "Comment on Anne Case and Angus Deaton: 'Accounting for the Widening Mortality Gap Between American Adults With and Without a BA'." Brookings Papers on Economic Activity. 2023:(Fall).

Institute for Health Metrics and Evaluation (IHME). 2014. "Global Smoking Prevalence and Cigarette Consumption: 1980–2012." Processed by *Our World in Data*. Available at https://ghdx.healthdata.org/record/ihme-data/global-smoking-prevalence-and-cigarette-consumption-1980-2012.

Kandel, Eric R., and Denise B. Kandel. 2014. "A Molecular Basis for Nicotine as a Gateway Drug." *New England Journal of Medicine* 371(10): 932–943.

Lawton, Ralph, Jennifer Dowd, Jonathan Skinner, and Ellen Meara. 2025. "Trends in Smoking and Age-Specific Mortality in the United States." Manuscript, Cambridge, Mass.

Levine, Amir, YanYou Huang, Bettina Drisaldi, Edmund A. Griffin Jr, Daniela D. Pollak, Shiqin Xu, Deqi Yin, Christine Schaffran, Denise B. Kandel, and Eric R. Kandel. 2011. "Molecular Mechanism for a Gateway Drug: Epigenetic Changes Initiated by Nicotine Prime Gene Expression by Cocaine." *Science Translational Medicine* 3(107): 107ra109–107ra109.

Link, Bruce G, and Jo Phelan. 1995. "Social Conditions as Fundamental Causes of Disease." *Journal of Health and Social Behavior* 80–94. Special Issue.

Loprinzi, Paul D. 2016. "Health Behavior Characteristics and All-Cause Mortality." *Preventive Medicine Reports* 3: 276–278.

Lortet-Tieulent, Joannie, Ann Goding Sauer, Rebecca L. Siegel, Kimberly D. Miller, Farhad Islami, Stacey A. Fedewa, Eric J. Jacobs, and Ahmedin Jemal. 2016. "State-Level Cancer Mortality Attributable to Cigarette Smoking in the United States." *JAMA Internal Medicine* 176(12): 1792–1798.

Meara, Ellen R., Seth Richards, and David M. Cutler. 2008. "The Gap Gets Bigger: Changes in Mortality and Life Expectancy, By Education, 1981–2000." *Health Affairs* 27(2): 350–360.

Miller, Charlotte E, and Ramachandran S Vasan. 2021. "The Southern Rural Health and Mortality Penalty: A Review of Regional Health Inequities in the United States." *Social Science & Medicine* 268: 113443.

Monnat, Shannon M. 2020. "Trends in US Working-Age Non-Hispanic White Mortality: Rural-Urban and Within-Rural Differences." *Population Research and Policy Review* 39(5): 805–834.

Montez, Jennifer Karas, Jason Beckfield, Julene Kemp Cooney, Jacob M. Grumbach, Mark D. Hayward, Huseyin Zeyd Koyta, Steven H. Woolf, and Anna Zajacova. 2020. "US State Policies, Politics, and Life Expectancy." *The Milbank Quarterly* 98(3): 668–699.

Montez, Jennifer Karas, Nader Mehri, Shannon M. Monnat, Jason Beckfield, Derek Chapman, Jacob M. Grumbach, Mark D. Hayward, Steven H. Woolf, and Anna Zajacova. 2022. "US State Policy Contexts and Mortality of Working-Age Adults." *PLOS ONE* 17(10): e0275466.

National Cancer Institute. 2021. "Tobacco Use Supplement to the Current Population Survey Harmonized Data, 1992–2019." Available at https://cancercontrol.cancer.gov/brp/tcrb/tus-cps/.

Novosad, Paul, Charlie Rafkin, and Sam Asher. 2022. "Mortality Change Among Less Educated Americans." *American Economic Journal: Applied Economics* 14(4): 1–34.

Pearce, Jamie, Ross Barnett, and Graham Moon. 2012. "Sociospatial Inequalities in Health-Related Behaviours: Pathways Linking Place and Smoking." *Progress in Human Geography* 36(1): 3–24.

Pierce, Justin R, and Peter K Schott. 2020. "Trade Liberalization and Mortality: Evidence from US Counties." *American Economic Review: Insights* 2(1): 47–63.

Powell, Lisa M., John A. Tauras, and Hana Ross. 2005. "The Importance of Peer Effects, Cigarette Prices and Tobacco Control Policies for Youth Smoking Behavior." *Journal of Health Economics* 24(5): 950–968.

Rajabi, Abdolhalim, Mohsen Dehghani, Azadeh Shojaei, Mojtaba Farjam, and Seyed Abbas Motevalian. 2019. "Association Between Tobacco Smoking and Opioid Use: A Meta-Analysis." *Addictive Behaviors* 92: 225–235.

Ritchie, Hannah, and Max Roser. 2023. "Smoking." *Our World in Data.* Available at https://ourworldindata.org/smoking.

Siddiqi, Arjumand, Odmaa Sod-Erdene, Darrick Hamilton, Tressie McMillan Cottom, and William Darity Jr. 2019. "Growing Sense of Social Status Threat and Concomitant Deaths of Despair Among Whites." *SSM-Population Health* 9: 100449.

Stone, Matthew D, John P Pierce, Brian Dang, Sara B McMenamin, Candice D Donaldson, Xueying Zhang, David R Strong, Yuyan Shi, Karen Messer, and Dennis R Trinidad. 2025. "State and Sociodemographic Trends in US Cigarette Smoking With Future Projections." *JAMA Network Open* 8(4): e256834–e256834.

Woolf, Steven H., and Heidi Schoomaker. 2019. "Life Expectancy and Mortality Rates in the United States, 1959–2017." *Journal of the American Medical Association* 322(20): 1996–2016.

Zachariou, Venetia, Carlos A. Bolanos, Dana E. Selley, David Theobald, Michael P. Cassidy, Max B. Kelz, Tamara Shaw-Lutchman, Olivier Berton, Laura J. Sim-Selley, Ralph J. Dileone, et al. 2006. "An Essential Role for ΔFosB in the Nucleus Accumbens in Morphine Action." *Nature Neuroscience* 9(2): 205–211.

A Appendix: Additional Results

Appendix Figures A.1 and A.2 evaluate the selection argument for the rising educational gradient in midlife mortality. Appendix Figure A.1 plots the education-specific mortality rate (y-axis) on education percentiles (x-axis) by year for college and non-college graduates. The figure demonstrates that the education percentiles of college and non-college populations, defined here as the midpoint of each education group's percentile range, changed little, falling from the 88th to 83rd percentile for college graduates and from the 38th to the 33rd percentiles for non-college graduates. Yet mortality rates decreased monotonically for college graduates from 1992 to 2019 (from 262 to 133 deaths per 100,000). For non-college graduates, the decline slowed after 2000, and deaths per 100,000 rose after 2010, such that mortality *increases* slightly comparing 1992 with 2019 (458 versus 463 deaths per 100,000). Appendix Figure A.2 uses the bounds-based selection correction of Novosad, Rafkin, and Asher (2022) to provide additional evidence that mortality patterns were not driven by selection. In the raw data, the widening educational mortality gradient is driven by a relatively stable non-college rate and a declining college rate. The green areas in each panel of the figure show that if we define the high-education group to be persons in the top 30 percent of educational attainment, rather than college graduates, we also see declines in mortality over time. The blue and red areas provide bounds for persons in the bottom 10 percent of the educational distribution and the 10th–70th percentiles. These areas are essentially stable, as is the national non-college mortality rate in the unadjusted data.

Appendix Figure A.3 shows that the mortality gap increase for Black and Hispanic populations is similar to the increase for the entire population. In addition, increases in the Black and Hispanic gaps arise from declining college mortality rates combined with more-stable non-college rates, which is also the pattern in the overall population.

Appendix Figure A.4 shows that mortality patterns for a largely non-immigrant population (Blacks and Whites) are nearly identical to the pattern for the entire population. The implication is that full-sample results are not likely to be biased by states with large numbers of immigrants, who tend to have lower mortality rates.

Appendix Figure A.5 displays a map of county codes from the 2013 Urban-Rural Classification Scheme from the National Center for Health Statistics.

Appendix Tables A.1 and A.2 present county-level mortality-change regressions for non-college and college populations excluding the rural/urban controls. (Regressions with these controls appear as Tables 1 and 2.) This sensitivity analysis is motivated by the potential for measurement error. If a right-hand side variable is measured with error, then some of its true effect may be absorbed within the urban-rural controls.

Appendix Table A.6 provides the state-level analysis of mortality gaps that is referenced in footnote 21. The horizontal axis of each chart depicts a prediction for the change in a state's mortality-education gap between 1992 and 2019. This prediction is constructed by subtracting the mortality-change prediction from a college regression model in Table 3 from that from a non-college model. The vertical axes depict actual mortality-gap changes. The

top chart uses models that contain only smoking variables (columns 2 and 6 of Table 3). The bottom chart uses predictions from models with both the smoking and income variables (columns 3 and 7). A comparison of the two charts shows that most of the information needed to predict the change in a state's mortality gap is contained in the smoking variables.

Appendix Table A.3 and Appendix Figures A.7 and A.8 replicate the state-level analysis of Table 3, Figure 11, and Appendix Figure A.6 using regression models that also include 1992 education-specific mortality rates as right-hand-side variables. The inclusion of the lagged mortality rate in the college regressions significantly improves the mortality-change predictions for those populations, due to the strong beta convergence in college mortality. But the main lesson of the baseline state-level analysis survives, as smoking behavior alone (particularly non-college smoking) does a good job of predicting how state-level mortality evolves between 1992 and 2019.

Appendix Figure A.9 replicates the analysis of potential convergence in education-specific state-level smoking rates in Figure 13 using data from the CPS-TUS. (Figure 13 uses data from BRFSS.)

Appendix Figures A.10 and A.11 present daily 2014 smoking rates for the United States and five European countries. The first of these figures corresponds to all education levels and the second is disaggregated by college/non-college. The number at the top of each bar is the total daily smoking rate for the given category of persons. This number is equal to the combined height of the red and blue sections of the bar. The lower number corresponds to the share of persons who are daily smokers that smoke more than 20 cigarettes per day. The number corresponds to the blue section of the bar.

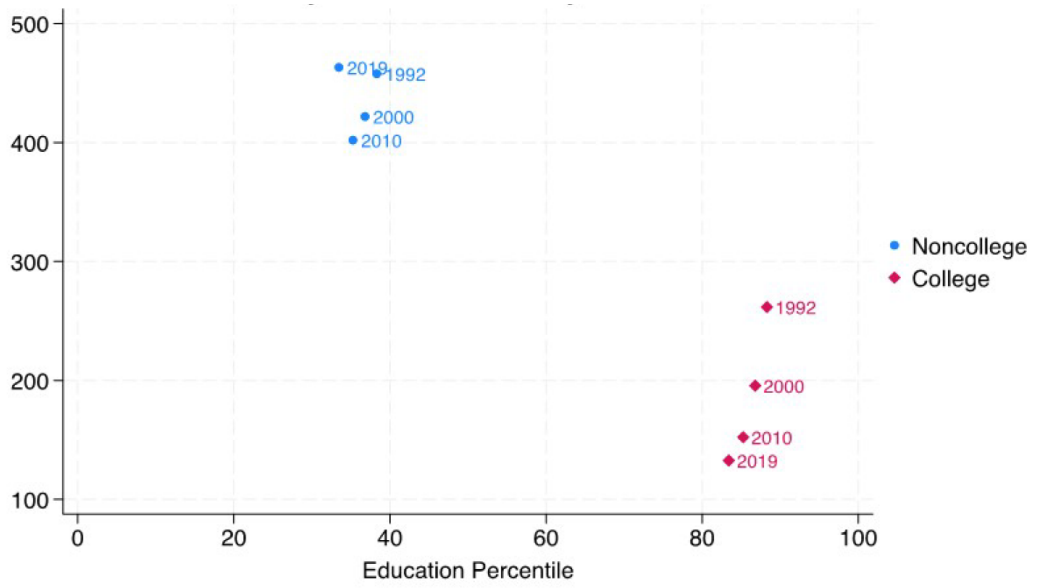

Figure A.1. Midlife Mortality Rates by Education and Education Percentile. Education percentile is defined as the midpoint of each education group's percentile range. Mortality rates are age-adjusted deaths per 100,000 population among adults aged 25–64.

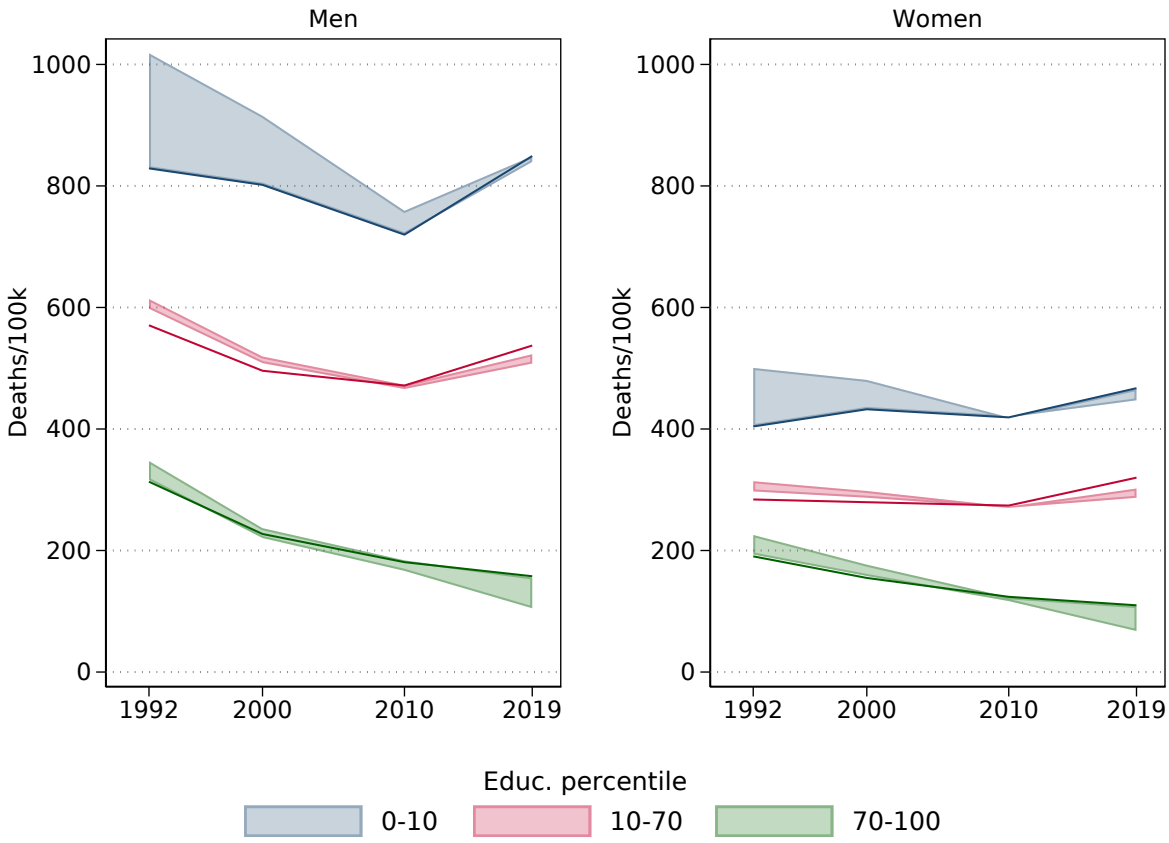

Figure A.2. Midlife Mortality by Education and Year with Novosad et al. (2022) Selection Correction. Notes: The selection model in Novosad, Rafkin, and Asher (2022) develops bounds on the education-mortality gradient that are anchored on the 10th and 70th percentiles. Assuming that health is monotonically related to education percentiles, the authors are able to estimate upper and lower bounds on the percentile estimates. These bounds are widely spaced when the actual rate of (say) high school dropouts is different from the 10th education percentile, as it is for men in 1992.

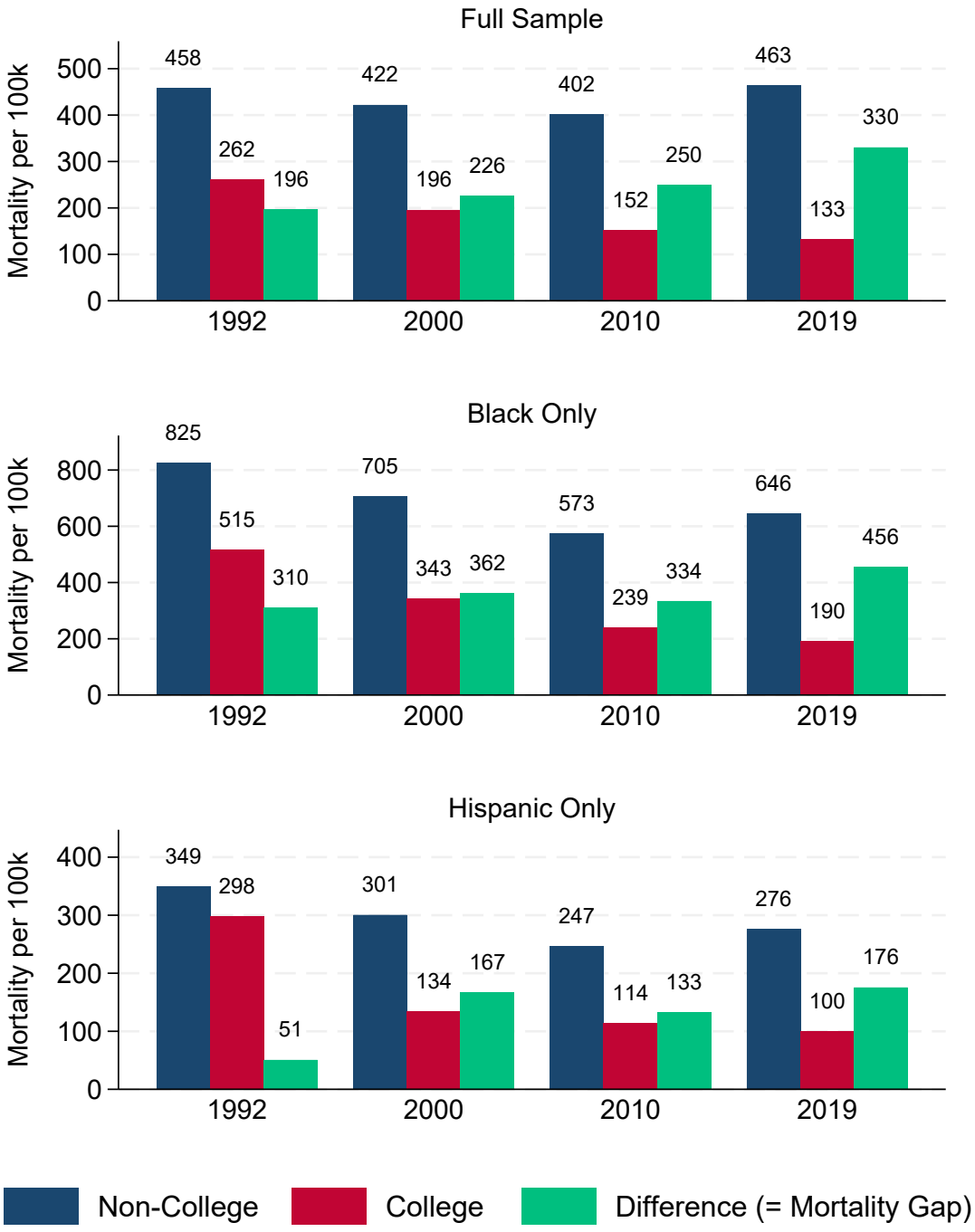

Figure A.3. Midlife Mortality Rates by Education and Year for the Full Sample and non-Hispanic Black and Hispanic Populations of Any Race Separately.

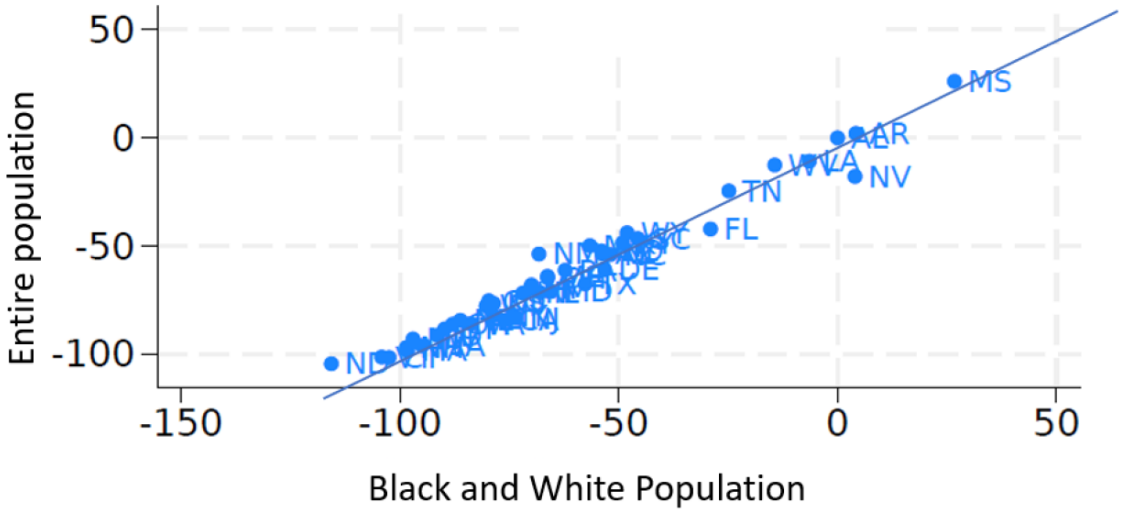

Figure A.4. State-Level Mortality for Non-Hispanic Black and Non-Hispanic White Populations Compared to the Full Population. Notes: The mortality estimates are centered on Alaska.

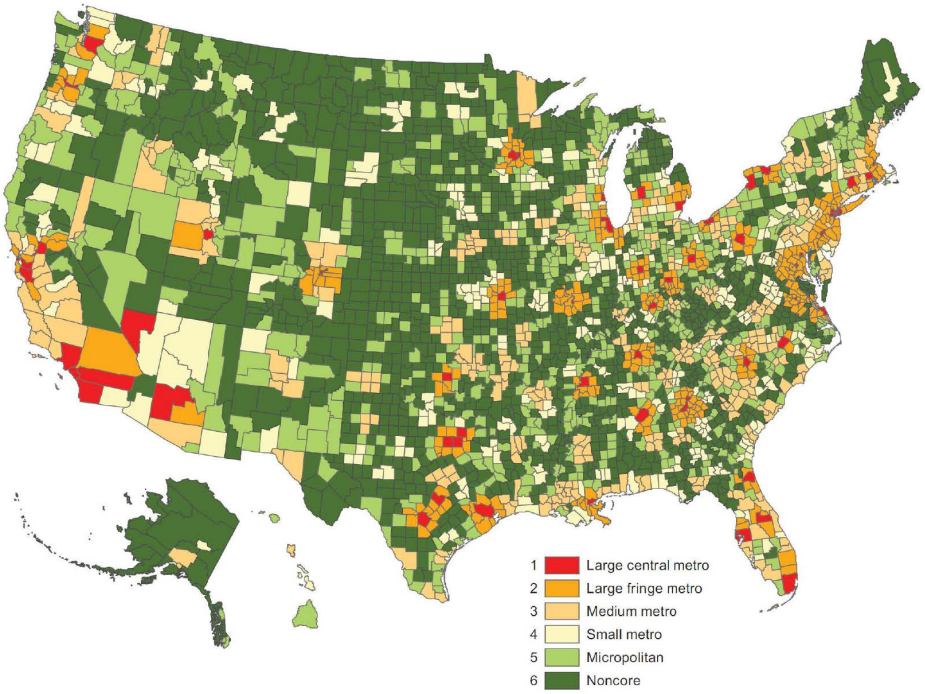

Figure A.5. County Codes from the 2013 Urban-Rural Classification Scheme from the National Center for Health Statistics.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
1992 Mortality Rate	–26.61***	–26.15***	–36.28***	–37.30***	–34.38***	–26.24***	–25.25***	–26.01***	–32.77***	–43.57***	–41.60***
	(6.174)	(6.090)	(5.304)	(5.758)	(5.157)	(5.990)	(5.828)	(6.205)	(5.572)	(4.108)	(5.435)
Avg. 1-yr Outmig Rate (2016-20)		9.634								0.462	–2.432
		(8.559)								(3.885)	(3.939)
Smoking Rate			63.10***							51.41***	42.57***
			(4.979)							(4.901)	(7.132)
Obesity Rate				34.65***						8.954*	10.89***
				(4.930)						(3.685)	(2.924)
Ln 1992 Per Capita Income					–38.53***					–27.85***	–21.67**
					(3.661)					(6.242)	(6.347)
1993-2019 Chg Ln Per Cap Income					–27.31***					–12.88***	–14.29***
					(3.750)					(2.581)	(2.801)
1992 Emp-Pop Ratio						–14.81***				6.950*	3.206
						(3.647)				(3.321)	(3.646)
1993-2019 Chg Emp-Pop Ratio						–31.35***				5.879	3.826
						(7.370)				(4.886)	(4.777)
1992 Manufacturing Empl Share							24.35***			2.856	–0.420
							(3.678)			(3.616)	(3.562)
China Shock								9.541+		0.258	–2.049
								(4.784)		(2.726)	(2.432)
State Policy-Liberalism Index									–35.40**	–5.071	
									(11.86)	(5.047)	
Constant	–16.08	–16.08	–16.08**	–16.08	–16.08	–16.08	–16.08	–16.08	–16.08	–16.08***	28.91***
	(15.85)	(15.21)	(5.455)	(14.80)	(11.66)	(13.77)	(14.15)	(15.69)	(11.61)	(4.491)	(4.879)
N	2081	2081	2081	2081	2081	2081	2081	2081	2081	2081	2081
R-sq	0.0590	0.0668	0.383	0.150	0.273	0.163	0.108	0.0666	0.160	0.477	0.553
State FEs?	N	N	N	N	N	N	N	N	N	N	Y

Table A.1. County-Level Mortality-Change Regressions, with Standardized Coefficients: Non-College. + 0.10 * 0.05 ** 0.01 *** 0.001. The dependent variable is 1993–2019 mortality-rate change for the county's non-college population. Standard errors are clustered by state and the data are weighted by education-specific average population in 1993 and 2019. The average 1-year out-migration rate is a 5-year average from the 2016–2020 American Community Survey. The China shock variable is from Autor, Dorn, and Hanson (2013), and the state policy-liberalism index is from Caughey and Warshaw (2016). Because the regressors are standardized, coefficients denote effects of 1-standard-deviation increases in the regressors.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
1992 Mortality Rate	–33.78***	–34.00***	–35.46***	–34.88***	–33.46***	–31.41***	–32.59***	–33.60***	–34.27***	–31.01***	–26.80***
	(8.037)	(8.060)	(8.641)	(7.149)	(6.798)	(6.094)	(7.833)	(8.027)	(7.916)	(6.123)	(6.486)
Avg. 1-yr Outmig Rate (2016-20)		–2.743								1.986	–0.681
		(2.749)								(3.320)	(4.091)
Smoking Rate			6.268							–0.526	–1.597
			(3.897)							(1.981)	(2.465)
Obesity Rate				16.80***						10.68***	7.266**
				(3.901)						(2.360)	(2.538)
Ln 1992 Per Capita Income					–5.086*					5.930***	5.382*
					(2.367)					(1.659)	(2.050)
1993-2019 Chg Ln Per Cap Income					–17.02***					–13.94***	–12.31***
					(1.833)					(1.556)	(1.662)
1992 Emp-Pop Ratio						–9.081**				–5.416+	–8.014**
						(2.833)				(2.935)	(2.767)
1993-2019 Chg Emp-Pop Ratio						–11.73**				–1.097	–2.802
						(3.868)				(2.972)	(3.363)
1992 Manufacturing Empl Share							10.42***			6.235**	4.899*
							(1.600)			(1.893)	(1.868)
China Shock								3.508*		2.096	3.955+
								(1.734)		(1.922)	(1.972)
State Policy-Liberalism Index									–6.349	–0.166	
									(5.937)	(4.267)	
Constant	–123.7***	–123.7***	–123.7***	–123.7***	–123.7***	–123.7***	–123.7***	–123.7***	–123.7***	–123.7***	–106.6***
	(5.356)	(5.509)	(4.759)	(3.836)	(3.372)	(4.821)	(5.079)	(5.427)	(4.526)	(2.922)	(3.180)
N	2081	2081	2081	2081	2081	2081	2081	2081	2081	2081	2081
R-sq	0.184	0.185	0.189	0.229	0.241	0.219	0.201	0.186	0.190	0.265	0.303
State FEs?	N	N	N	N	N	N	N	N	N	N	Y

Table A.2. County-Level Mortality-Change Regressions, with Standardized Coefficients: College. + 0.10 * 0.05 ** 0.01 *** 0.001. The dependent variable is 1993–2019 mortality-rate change for the county's college population. Standard errors are clustered by state and the data are weighted by education-specific average population in 1993 and 2019. See the notes to Table A.1 for a description of the average 1-year out-migration rate and the sources of the child economic-connectedness and state policy-liberalism regressors. Because the regressors are standardized, coefficients denote effects of 1-standard-deviation increases in the regressors.

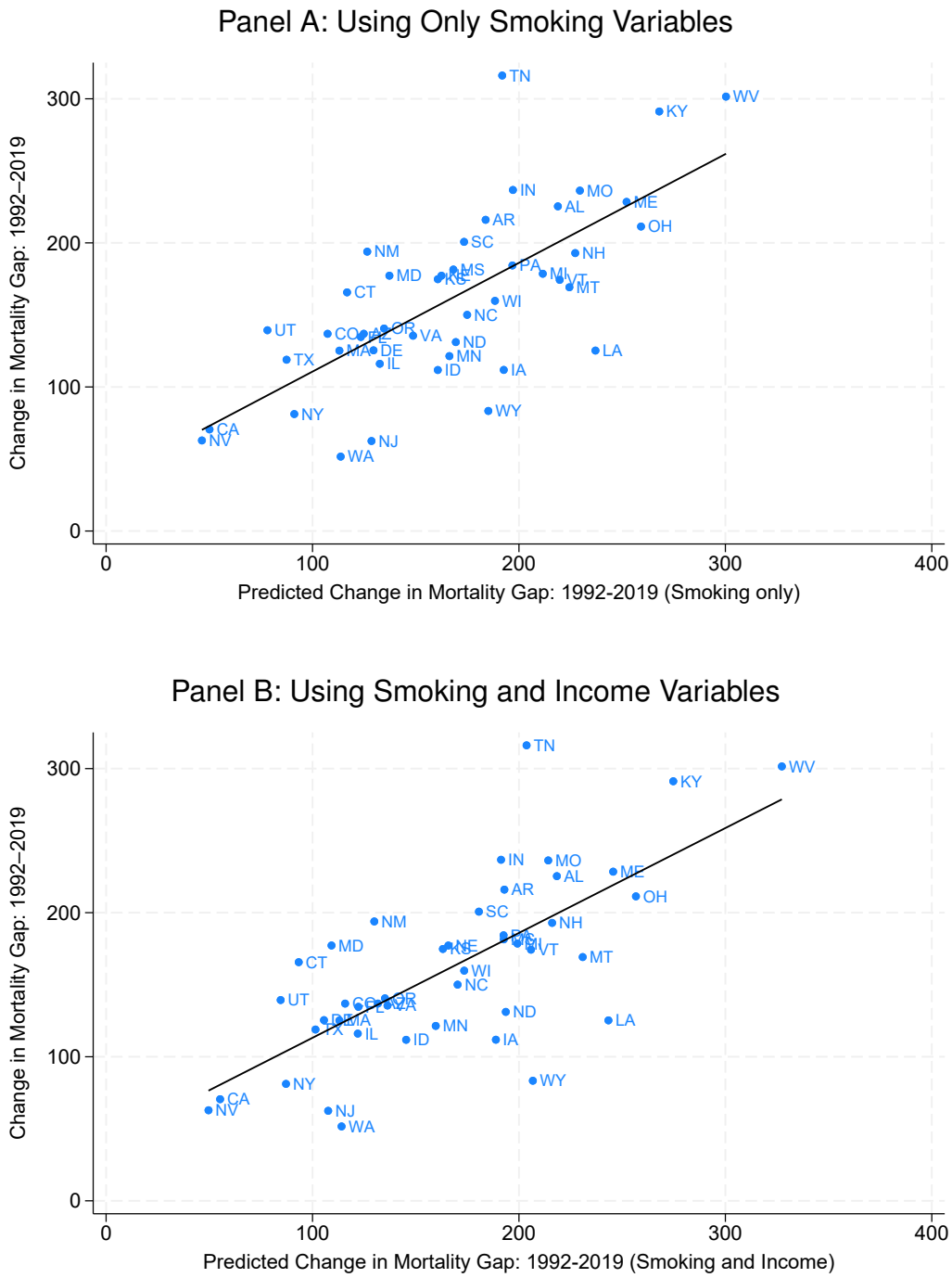

Figure A.6. Predicted and Actual Values for 1992–2019 State-Level Changes in Mortality-Rate Gaps. Predicted values for mortality-gap changes in Panel A are the differences between non-college mortality-change predictions generated by column 2 of Table 3 and college mortality-change predictions generated by column 6. The predicted values for Panel B are the differences between non-college predicted values generated from column 3 and college predicted values generated from column 7.

	(1) Non-Coll	(2) Non-Coll	(3) Non-Coll	(4) Non-Coll	(5) College	(6) College	(7) College	(8) College
Ln 1992 State Per-Capita Income	–340.2***		–151.2*	–158.9*	–88.24***		–71.78***	–85.63***
	(70.81)		(68.30)	(74.09)	(14.84)		(16.71)	(22.02)
1992-2019 Chg. in Ln State Income	–674.3***		–109.3	–40.09	–73.74*		–30.84	–42.60
	(165.0)		(127.6)	(131.0)	(31.24)		(37.29)	(42.67)
1992 Non-College Smoking Rate		1189.7***	981.1***	1070.5***				
		(168.0)	(220.4)	(257.3)				
1992–2019 Chg in Non-College Smoking Rate		1273.8***	1012.6***	1034.5***				
		(123.9)	(119.7)	(124.2)				
1992 Non-College Obesity Rate				–402.0				
				(278.7)				
1992–2019 Chg in Non-College Obesity Rate				171.5				
				(209.5)				
1992 College Smoking Rate						883.9***	414.0*	395.2
						(196.9)	(196.4)	(231.4)
1992–2019 Chg in College Smoking Rate						926.9***	440.9*	374.2
						(186.9)	(177.3)	(227.1)
1992 College Obesity Rate								–124.2
								(83.02)
1992–2019 Chg in College Obesity Rate								–3.346
								(69.36)
1992 Non-College Mortality Rate	–0.531***	–0.402*	–0.435**	–0.414**				
	(0.134)	(0.159)	(0.147)	(0.141)				
1992 College Mortality Rate					–0.863***	–0.887***	–0.880***	–0.878***
					(0.0479)	(0.0933)	(0.0507)	(0.0501)
Constant	4118.2***	0.632	1671.9*	1716.4*	1050.8***	70.28**	849.0***	1007.7***
	(767.8)	(78.59)	(782.3)	(844.9)	(152.8)	(23.42)	(183.4)	(254.4)
N	44	44	44	44	44	44	44	44
R-sq	0.628	0.787	0.826	0.841	0.916	0.880	0.925	0.928

Table A.3. State-Level Mortality-Change Regressions Including 1992 Mortality Rate. The dependent variable for each regression is the state-level mortality-rate change from 1992 to 2019 for non-college (columns 1–4) or college (columns 5–8) populations. Data are weighted by average education-specific population in 1992 and 2019 and robust standard errors are in parentheses. State-level smoking rates for 1992 and for 2019 (used for the 1992–2019 change) are calculated from BRFSS.

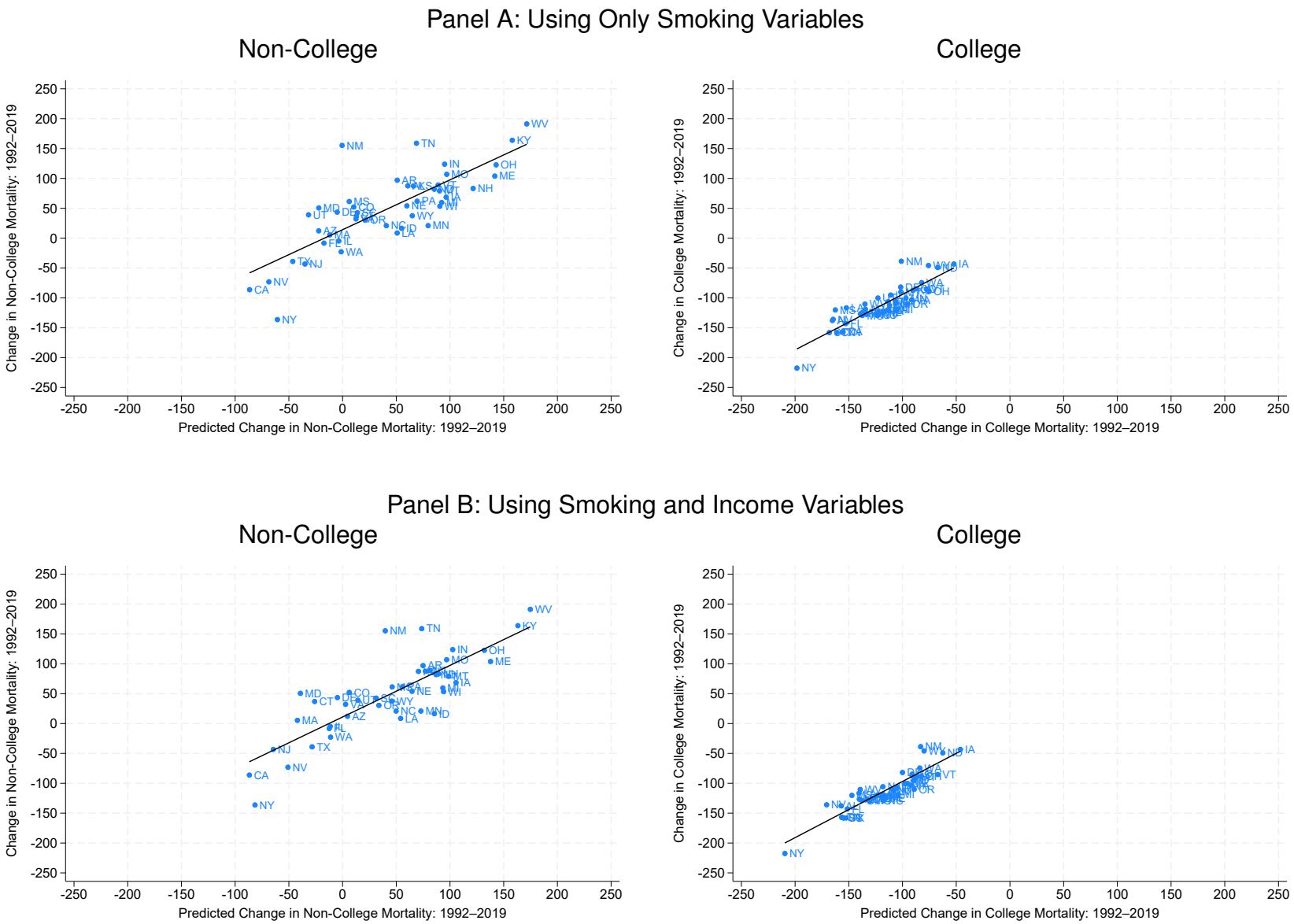

Figure A.7. Predicted and Actual Values for 1992–2019 State-Level Education-Specific Mortality-Rate Changes, From Models with 1992 Mortality Rates Included. Predicted values for Panel A use the regression models in columns 2 (for non-college) and 6 (for college) in Table A.3. These models use only the 1992 level and the 1992–2019 change in education-specific smoking rates. Panel B uses predictions from the models in columns 3 and 7, which add the 1992 level and 1992–2019 change in state per-capita income to the regression models.

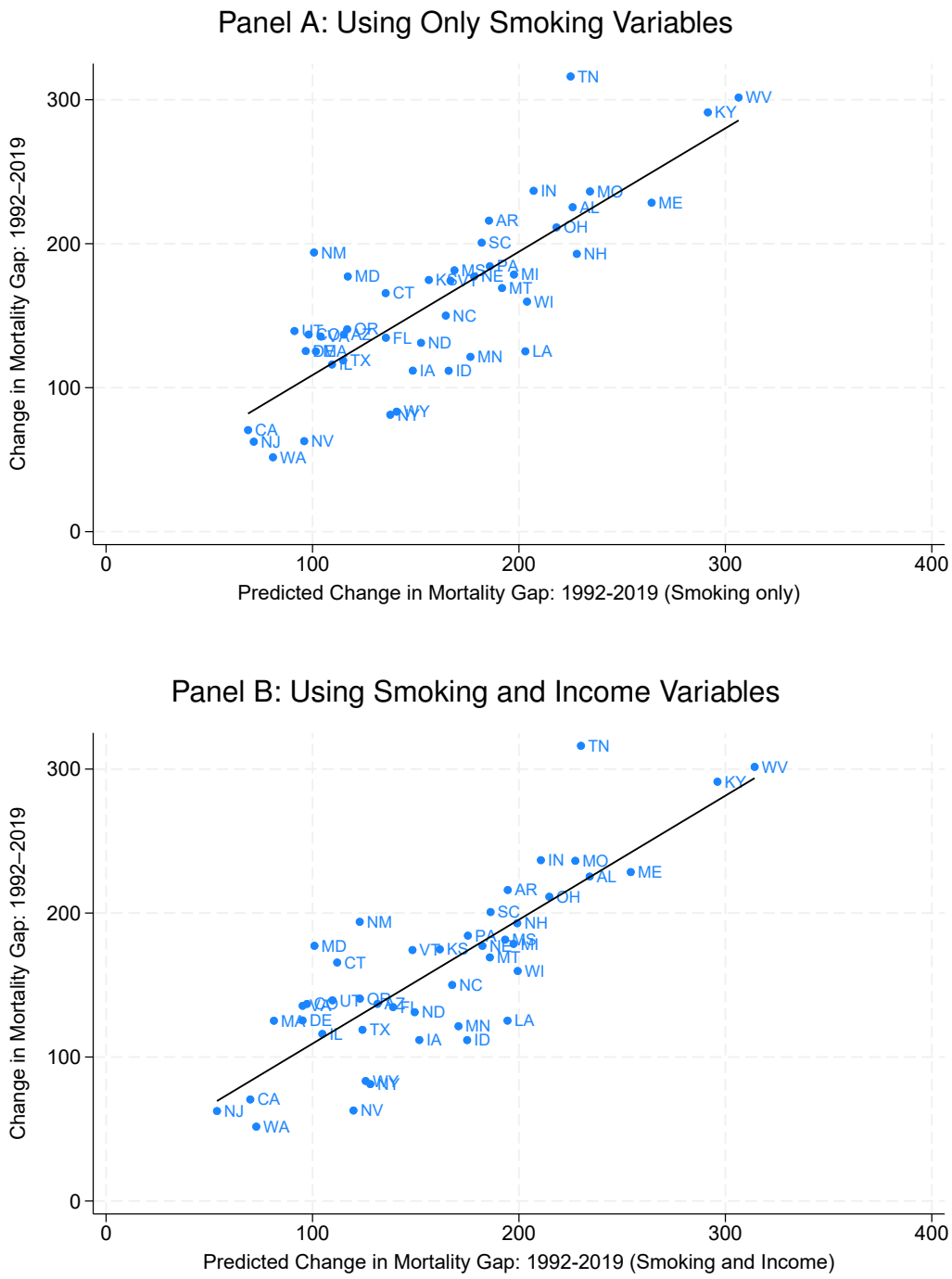

Figure A.8. Predicted and Actual Values for 1992–2019 State-Level Changes in Mortality-Rate Gaps, from Models with 1992 Mortality Rates Included. Predicted values for mortality-gap changes in Panel A are the differences between non-college mortality-change predictions generated by column 2 of Table A.3 and college mortality-change predictions generated by column 6. The predicted values for Panel B are the differences between non-college predicted values generated from column 3 and college predicted values generated from column 7.

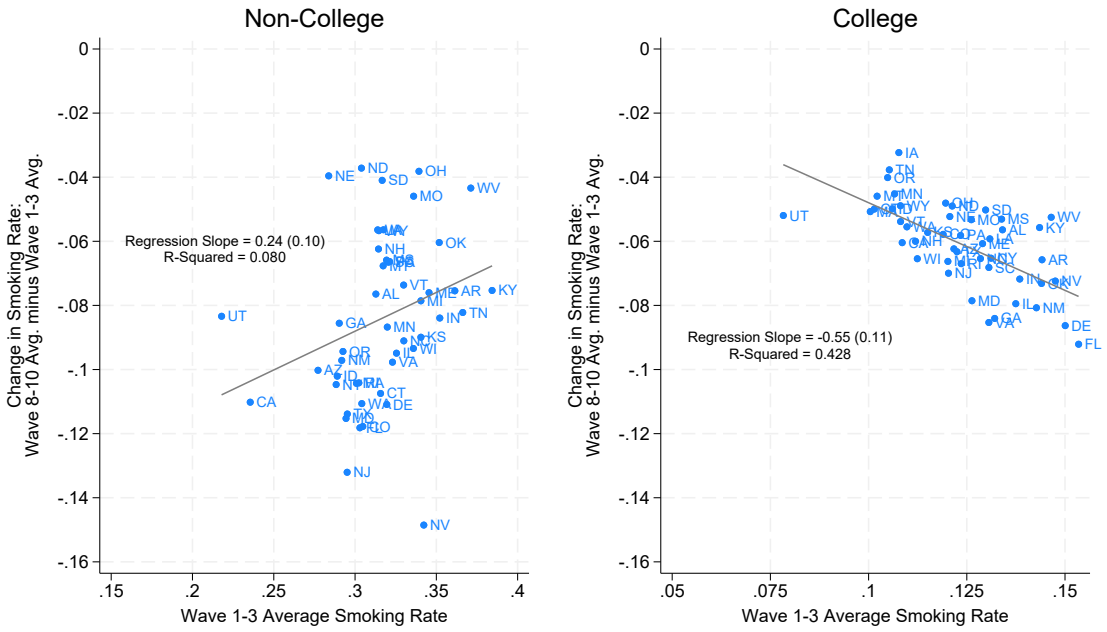

Figure A.9. State-Level Changes in Current-Smoking Rate at Midlife and Initial Current-Smoking Rate by Education, using CPS-TUS data. *Notes:* Data are age-adjusted and regressions are unweighted. *Source:* Current Population Survey Tobacco-Use Supplements Waves 1 (1992–1993), 2 (1995–1996), 3 (1998–1999), 8 (2010–2011), 9 (2014–2015), and 10 (2018–2019).

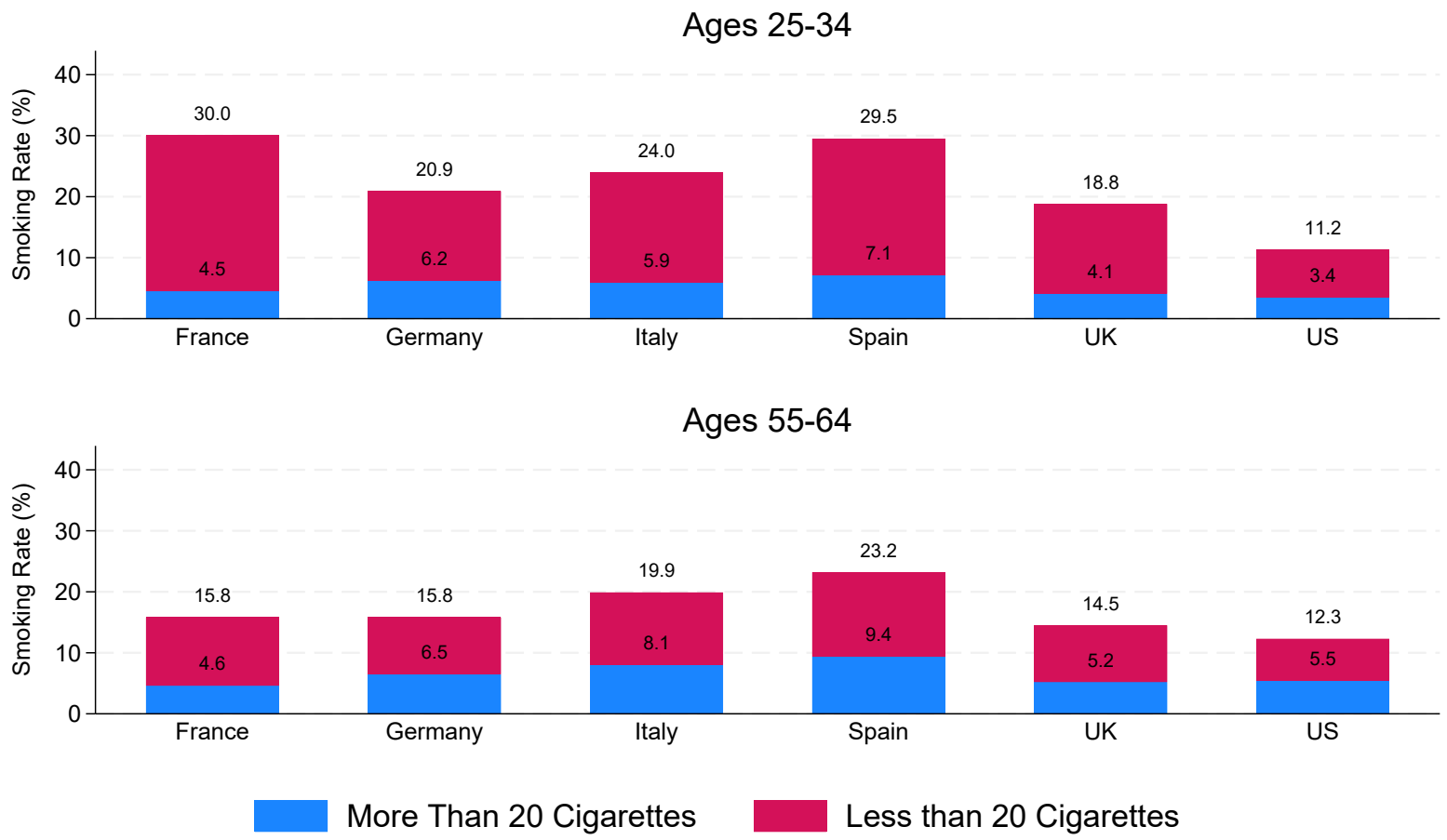

Figure A.10. Smoking Rates by Age Group for Six Selected Countries in 2014. *Notes:* The number at the top of each bar is the total daily smoking rate for the given age group and country. This number is equal to the combined height of the red and blue sections of the bar. The lower number corresponds to the population share of persons who smoke more than 20 cigarettes per day. This number corresponds to the blue section of the bar. *Sources:* European data are from Eurostat. US data are from the Tobacco-Use Supplement of the Current Population Survey.

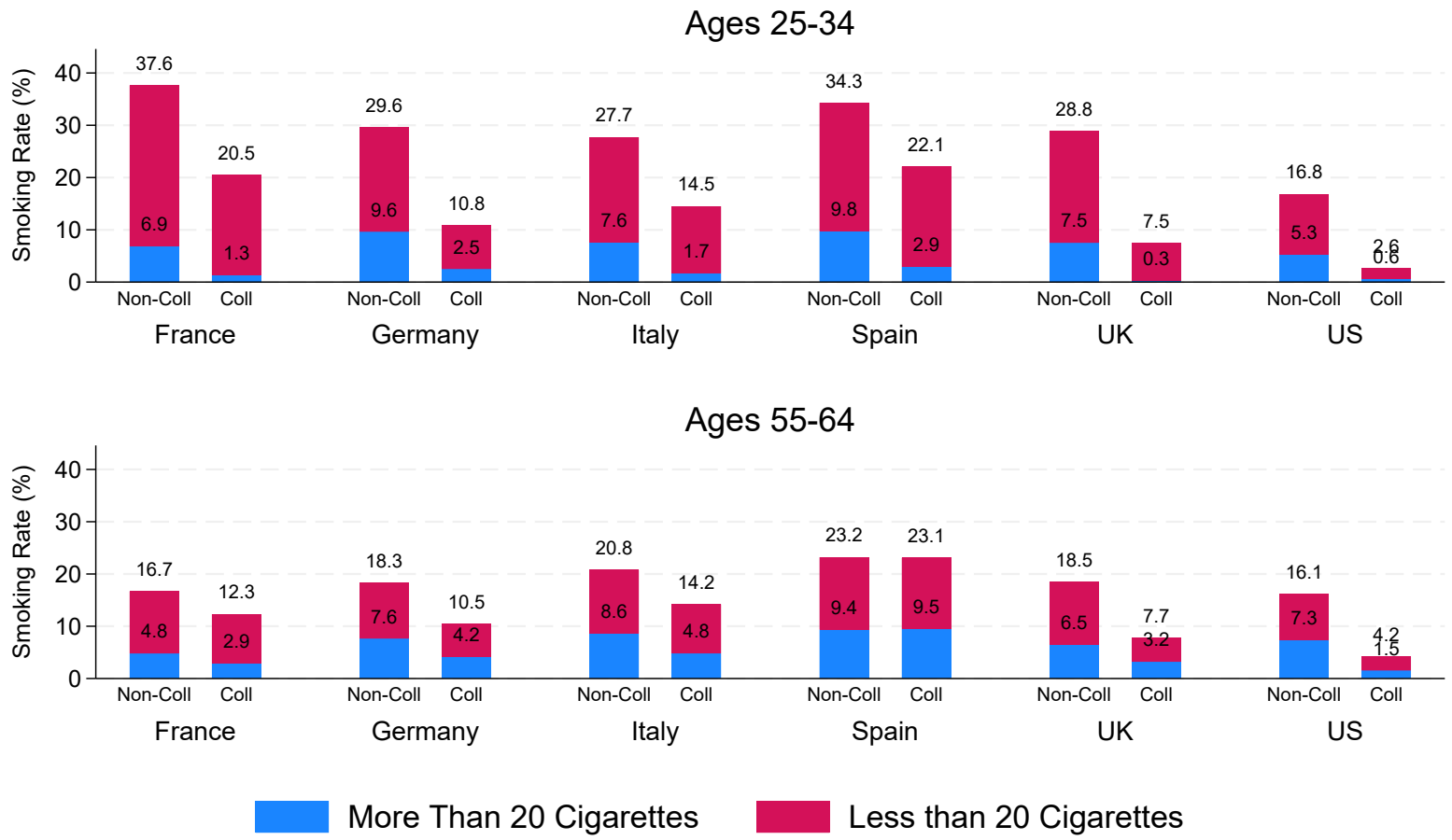

Figure A.11. Smoking Rates by Age Group and Education Level for Six Selected Countries in 2014. *Notes:* The number at the top of each bar is the total daily smoking rate for the given education level, age group and country. This number is equal to the combined height of the red and blue sections of the bar. The lower number corresponds to the population share of persons who smoke more than 20 cigarettes per day. This number corresponds to the blue section of the bar. For the European countries, the non-college rate is the population-weighted average of smoking rates for persons with education corresponding to ISCED11 categories 0–2 (lower secondary education and less) and categories 3–4 (upper secondary and post-secondary non-tertiary education). The college category corresponds to ISCED11 categories 5–8 (tertiary education). *Sources:* European data are from Eurostat. US data are from the Tobacco-Use Supplement of the Current Population Survey.

B Appendix: Data

B.1 County-level Data

For any county-level variable, we require a set of consistent county definitions over time so that our observations correspond to the same geographic area throughout the panel. Fortunately, county definitions are more consistent than any other sub-state geographic unit over time. However, some changes do occur. Furthermore, different data systems use different conventions for independent cities or other local areas that are similar to counties. Therefore, we must adopt the conventions of a given system and convert between the conventions of other systems to account for this.

A small number of counties combined or split during the sample period. To account for this, we use the largest definition of a given county or set of counties throughout the whole period. For example, in 1983, Yuma County (FIPS 4027) splits into a smaller Yuma County and La Paz County (FIPS 4012). For years in which the two are split in the data, we combine them and assign them the FIPS code of the original (4027). The following counties are edited across all data sources:[32]

- Miami-Dade county is recoded from 12025 to 12086 in all periods due to a renaming in 1997.
- Counties with FIPS codes 36005, 36047, 36061, 36081, and 36085 (which are the 5 FIPS codes associated wit New York City) are combined in all periods.
- Yuma County (FIPS 4027) and La Paz County (FIPS 4012) are combined in all periods.
- Counties 8001, 8013, 8014, 8059, and 8123 (alternately known as 8911, 8912, 8913, 8914, and 8014) are combined for all periods.
- Counties 30113, 30067, and 30031 are combined for all periods.
- Counties 35006 and 35061 are combined in all periods.
- Records with FIPS 13999 are dropped from mortality data, because this special county designation was given to all HIV deaths in Georgia in some years.
- County 46113 is recoded to 46102 in all periods.
- Counties 55078 and 55115 are combined into 55901 in all periods.

The state of Virginia has counties and independent cities that are considered county-equivalents by the Census. Some data sources, like SEER and the NCHS micro-level mortality data, do not combine these regions. Others, like the BEA, combine these into slightly larger regions. We aggregate to the higher level, using BEA definitions. However, in rare cases we must further aggregate these regions for intertemporal consistency. Furthermore, in early SEER data (before the early 80s), some county/city combinations in Virginia had different codes

[32]Thanks to David Dorn for compiling a list of county changes that was helpful in developing these definitions. See the list here: https://www.ddorn.net/data/FIPS_County_Code_Changes.pdf.

than the BEA data but were nonetheless combined. These codes need to be reassigned to the BEA definitions. We use a crosswalk to convert between the two definitions where necessary.[33]

After these combinations are made, the final panel contains 3063 counties with data from 1969 to 2021.

B.2 Mortality Rates

In general, to calculate mortality rates in a given county and year, we need to obtain mortality counts and population in that county and year. Before 1989, we obtain data on both population and mortality from the CDC's WONDER database. From 1989 onward, we use Restricted-Use Vital Statistics data from the National Vital Statistics System of the National Center for Health Statistics for mortality data. We gather population data from the Surveillance, Epidemiology, and End Results program of the National Cancer Institute for this period. In both systems, we drop data from Alaska, DC, and Hawaii for data reliability concerns.

B.2.1 Pre-1989 Mortality Rates

The WONDER database contains both deaths and population by county, age group, and cause from 1968 to 1988. We pull data by county for 25–34, 35–44, 45–54, and 55–64 age groups (which we use as age bins throughout), separately for all-cause mortality and for a set of specific causes. Some records are regarded as unreliable due to a small number of deaths occurring. We choose to not treat these differently than other records. After standardizing county definitions, we calculate the deaths and populations by year and age group for our counties, separately by cause. We then calculate the mortality rate per 100,000 people by county, year, and age group for each cause.

B.2.2 1989-2021 Mortality Rates

The NCHS Restricted-Use Data provides us with individual death records that contain, among other things, county of residence, county of death, demographics such as age and race, cause of death, and educational attainment, from 1989 to 2021 for all recorded deaths during that period. We convert both county of residence and county of death into our county basis before we aggregate deaths to the county and 10-year age group level for all-cause mortality and for a select group of causes for each year. For population, we take data from SEER from 1989 to 2020. We convert SEER to our county basis, then aggregate to the county and age group level by year to merge with the deaths data.

Data on the 2021 population is not available from SEER. To address this, we use Census population data from 2020 and 2021. We aggregate these data to our county definitions in 5 year age bins. We convert to county definitions for Connecticut, which changed its statistical areas in the 2020 Census from counties to planning regions. We then merge both years of data with SEER data from 2020. We estimate a linear regression of SEER population in

[33] A blog post on the topic and the crosswalk itself can be downloaded here: https://www.economy.com/support/blog/buffet.aspx?did=869A03D1-5D74-4376-A606-00A8C64DDB0B.

2020 on Census population in 2020, then predict SEER population in 2021 using that model and the Census population in 2021. This gives us an estimate of what SEER population would be in 2021 data. We then aggregate these population estimates to our 10-year age bins.

We combine our SEER population data from 1989 to 2020 with our population estimates from 2021 then merge on our death counts from the microlevel mortality data. Some counties have 0 deaths with nonzero population for certain age groups, which is plausible for small populations. After aggregating to 10-year age bins, no counties have a population of 0 with non-zero deaths, suggesting our population estimates are reliable. We then calculate the death rate per 100,000 by cause and age-group for each county and year.

Finally, we append the mortality rates that are calculated by county, year, and 10-year agegroup from the WONDER database with those from the microlevel mortality data.

B.2.3 Age-standardizing Mortality Rates

To adjust for the fact that older age groups tend to have higher mortality rates, which would distort relative mortality rates for counties with different age distributions, we age-standardize mortality rates using the 25–34, 35–44, 45–54, and 55–64 age bins and the 2000 Census standard age distribution. Specifically, we calculate

$$M_{cst} = \sum_{a=1}^{4} \frac{P_{a,2000}}{P_{2000}} M_{acst}$$

where M_{acst} is the age-group-specific mortality rate in year t, county c, state s, and age group a, $P_{a,2000}$ is the population in age group a in 2000, and P_{2000} is the total population in 2000.[34] Therefore, we now have age-standardized midlife mortality rates by cause for select causes by county and year.

B.2.4 Causes of Death

We combine multiple ICD-8, ICD-9 and ICD-10 codes into our own cause-of-death categories that are mostly consistent across years. For the most part, we are interested in all-cause mortality. Table B.4 details our categories and the corresponding ICD codes for the three systems.

We choose these categories and the corresponding ICD codes to represent major causes of death which are mutually exclusive.[35].

To properly convert between ICD-9 and ICD-10 mortality counts, we multiply death counts in each category by comparability ratios provided by the CDC, which inflate/deflate death counts by major cause groups so that they are consistent through time.[36]

[34]See this CDC statistical note for details: https://www.cdc.gov/nchs/data/statnt/statnt20.pdf

[35]For deaths related to alcohol and drugs, as well as suicide deaths, we use the list of ICD codes used in this Senate Joint Economic Committee report: https://www.jec.senate.gov/public/index.cfm/republicans/methodological-appendix-to-long-term-trends-in-deaths-of-despair/

[36]See this CDC page for more details on ICD comparability: https://www.cdc.gov/nchs/nvss/mortality/comparability_icd.htm

We calculate the mortality rate that is not otherwise specified by our select causes (by subtracting deaths of each cause by the total), and, where this is less than zero due to rounding and the use of comparability ratios, we set it equal to 0. This minor adjustment affects very few counties.

B.3 Mortality Rates by Education

Educational attainment is provided for some records in all years, but is frequently missing in 1989, 1990, and 1991. Two different definitions for educational attainment are used depending on the data year, one from 1989 and one from 2003. We map both onto our own education variable that records education into six categories: less than high school, some high school, high school graduate, some college, bachelor's degree, and advanced degree. We typically compare those with a college degree to those without.

Most of the processing associated with these mortality rates is conceptually the same as the combined mortality rates described above. We calculate the number of deaths by county, year, age group, and education then we calculate the relevant populations for this group. We then take the ratio of these to get the mortality rate, which we age-adjust to be at the county-year level.

However, there are some key differences. First, due to the lack of consistently available data on population by education, county, and year, we use data from the Census and ACS as described in the text. At the state level, we can calculate the relevant populations using the Current Population Survey (CPS) yearly from 1992 through 2021.

Second, some records are missing information on education. In the 1990s, this happens somewhat frequently, but becomes less common in later years. If we simply assigned deaths to the category of education which is recorded in the death record, we would undercount total deaths (and mortality rates) in a county-year by a potentially large amount.

B.3.1 State and County Populations by Education

To calculate both state and county populations by education (and age group), we calculate shares of the population by educational attainment, and multiply those shares by SEER population counts to get estimates of the respective populations.

At the county level, we use data from the decennial Census (for 1990 and 2000) and from the ACS (for 2010 onwards) that records population counts in a PUMA by age group, sex, race, and educational attainment for the 25–34, 35-44, 45-54, and 55-64 age groups. We compute the population of each educational group within each PUMA cell. Then, we convert to a county-basis using a geographic crosswalk from the Missouri Census Data Center Geocorr Applications. Each PUMA is mapped to one or more counties, and its population is divided into those counties with an allocation factor. The allocation factor is the share of the PUMAs total population that lies within the county. We multiply the PUMA-cell populations by this allocation factor then compute the total population within a cell at the county level.

When a county is only within one PUMA, this county-cell population is simply the PUMA-cell population multiplied by the allocation factor. When a county contains multiple PUMAs, this county-cell population is the sum of all the PUMA-cell populations that

it contains. Finally, we compute the share of the population by educational attainment within each county-cell to multiply by the total population computed from SEER. By doing this, our education-specific populations match the totals in the overall populations. In performing these calculations, we make two key assumptions: 1) The population distribution across counties within a PUMA is the same for all demographic cells and 2) The education distribution is the same within the same demographic cells in all counties within a PUMA.

For population estimates in 1992, we compute population estimates in 1990 and 2000, then estimate the 1992 population within each county-education-cell as 0.8 times the 1990 estimate plus 0.2 times the 2000 estimate (equivalent to a linear interpolation across that 10-year period).

At the state level, we use data from the Current Population Survey to estimate the share of the population in each state with a college degree for each of our original 10-year age groups, by year from 1992 to 2021. We use the monthly CPS and aggregate to a year level. Because the CPS faces considerable measurement error and year-to-year noise, we smooth our estimates using a moving average with 1 lag and 1 lead.[37] We then merge these estimated shares with SEER data to calculate population counts.

B.3.2 Adjusting for Missing Education

We must also adjust the way we calculate death counts by education. Many records are missing education data, so we take a similar approach as in the population data. We calculate the share of deaths in each county and age-race-sex cell by educational attainment (using records that are not missing education), then multiply it by the total count of deaths within the county-year-age group cell. This methodology implicitly assumes that the records with missing education data have the same education distribution those that have education data.

We then merge these estimated death counts with the population data (after aggregating to the state level, for state-level data), calculate age-group specific mortality rates by geographic unit and year, and age-adjust with the 2000 Census age distribution.

Some counties are missing education for many records, especially in 1992. This causes concern over the reliability of estimates associated with these counties. Therefore, if a county is missing education data in more than half of all records within an age group for a given year, we record mortality rates as missing for all education categories. When aggregating death counts to the state or national level, we do not exclude the counts from such counties. However, a small number of county-demographic cells are missing education data for all entries. Therefore, after aggregating death counts to the state and nation level, the total number of deaths across education categories is below that computed without regards to education by a small amount. Given the small number of records this affects, it will have a negligible effect on our analysis.

[37] On the boundaries of our time series, we use just the current value and the lag, or the current value and the lead, depending on which boundary.

B.4 Per Capita Income

Besides mortality rates, we also extensively use per capita income at both the state and county level. We take both of these from the Bureau of Economic Analysis. Because we use the BEA's broader definition of counties, less processing must be done to harmonize the county definitions with our conventions. When we combine two or more counties together, we take the average of their per capita incomes, weighted by population.

B.5 Behavioral and Health Outcomes from the Behavioral Risk Factor Surveillance System (BRFSS)

We compute rates of smoking, obesity, and inactivity using survey data from BRFSS. We create indicator variables using the following definitions.

- Smoking: The respondent must have smoked at least 100 cigarettes in their lifetime, consider themselves an active smoker, and smoke regularly. This is the basis of a calculated variable in the BRFSS data.
- Obesity: The respondent has a body mass index over 30.
- Inactivity: The respondent has not exercised or performed recreational activities outside of work in the last 30 days.

We then compute means of each indicator variable within counties (both overall and conditional on education) using provided survey weights, across all years in which county is identified in the data, 1996 to 2010. We set a cell's given rate to missing if fewer than 50 records are used to compute that rate across the 15 year period.

At the state level, we are able to compute rates that change over time. We compute weighted means of each indicator variable by state and year, overall and across educational attainment, from 1992 to 2021.

Cause category	**ICD-8 Codes**	**ICD-9 Codes**	**ICD-10** Codes
Circulatory	390-398, 402, 404, 410-413, 420-429, 440, 441, 450	390-398, 402, 404, 410/425.4, 425.6/429.9	I00/I09.9, I11, I13, I20/I42.5, I42.7/I51.9
Cancer	140/162.0	140/162.0, 163/208	C00/C33.9, C35/C97
Lung cancer	162.1	162.2/162.9	C34
Lower Resp.	490-493, 518, 519.3	490/494	J40/J47
External	E800/E849, E861/E949, E960/E979, E980.4/E999	E800/E849, E861/E949, E960/E979, E980.6/E999	V01/X39, X46/X59, X85/Y09, Y16/Y86, Y88/Y89
Cerebrovascular	430/434, 436/438	430/434, 436/438	I60/I69
Diabetes	250	250	E10/E14
Alcohol	291, 303, 571.0, E860	291, 303, 305.0, 357.5, 425.5, 535.3 571.0/571.3, 790.3, E860	E24.4 F10, G31.2, G62.1, G72.1, I42.6, K29.2, K70, K85.2, K86.0, O35.4, P04.3, Q86, R78.0, X45, Y15
Suicide	E950/E959	E950/E959	X60/X84, Y87
Drugs	304, E850/E859, E980.0/E980.3	292, 304, 305.2/305.9, E850/E858, E980.0/E980.5	F11/F16, X40/X44, Y10/Y14

Table B.4. ICD Codes by Causes of Death.