

Online Appendix for:
“Rising Geographic Disparities in US Mortality”

Benjamin K. Couillard Christopher L. Foote Kavish Gandhi
Ellen Meara Jonathan Skinner

This version: October 26, 2021

1 Data Sources

Replication files for this paper will be available at a web page maintained by the *Journal of Economic Perspectives*. All data are publicly available online, with the exception of the restricted-use mortality data from the National Vital Statistics System (NVSS). In this section, we outline our main data sources.

Mortality data: Mortality data are from NVSS’s confidential microdata files for individual decedents (Centers for Disease Control and Prevention, National Center for Health Statistics 2017b). These data can be accessed via application at <https://www.cdc.gov/nchs/nvss/nvss-restricted-data.htm>. Where possible, results in the paper were validated using public-use data from the CDC Wonder Underlying Cause of Death Compressed Mortality Data, found at <https://wonder.cdc.gov/mortSQL.html>. Comparability ratios, to allow comparison of mortality rates for different causes of death across the ICD-9 and ICD-10 revisions, were obtained via a FTP request to https://ftp.cdc.gov/pub/health_statistics/NCHS/Datasets/Comparability/icd9_icd10/Comparability_Ratios_tables.xls. We also implement an alternate method of mapping causes of death to one another using the General Equivalence Mappings between ICD-9 and ICD-10, obtained from <https://www.cms.gov/Medicare/Coding/ICD10/Downloads/2018-ICD-10-CM-General-Equivalence-Mappings.zip>.

Life Expectancy: State-level life expectancy at birth is from the United States Mortality Database (2021). Data can be downloaded as .txt files after registering at <https://usa.mortality.org>. National life expectancy rates through 2020 are from the National Center for Health Statistics and were retrieved using the Haver data service. The Haver mnemonics are USLE, USLEWX, USLEBX, USLEH, USLEB, USLEW in the USECON database.

Population: Population estimates used in the denominators of mortality-rate estimates are from National Cancer Institute’s Surveillance, Epidemiology, and End Results Program (SEER). Breakdowns into state-education cells were estimated using Current Population Survey microdata from IPUMS (Flood et al. 2020). The SEER data can be downloaded from <https://seer.cancer.gov/popdata/download.html>, and the IPUMS-CPS data are available at <https://cps.ipums.org/cps/>. State-level population estimates through 2020, used in the population-weighting of Figure 1, are from the US Census Bureau’s Resident Population Annual Estimates program and can be downloaded from the Federal Reserve Bank of St. Louis at <https://fred.stlouisfed.org/release/tables?rid=118&eid=259194>.

State-level Covariates: Our exploratory regressions are discussed near the end of the main text and outlined in more detail in Section 4 of this appendix. The state-level covariates used in these regressions include:

- *Income:* The per capita income variable used in the exploratory regressions is the same

income variable used throughout the main text. Real per capita personal income is based on estimates from the Bureau of Economic Analysis (BEA) of total personal income received by the residents of individual states in each year. These data can be downloaded from BEA’s Regional Data Tables (Table SAINC1) at <https://apps.bea.gov/itable>. In addition to total personal income by state, this table also includes *per capita* personal income, defined by BEA as total personal income divided by state population as of July 1 of the given year. We express that per capita figure in 2012 dollars using BEA’s price deflator for personal consumption expenditures, which can be downloaded from the St. Louis Fed at <https://fred.stlouisfed.org/series/PCEPI>.

- *Smoking and obesity*: State-level smoking and obesity data come from the Behavioral Risk Factor Surveillance System (BRFSS), an annual set of telephone surveys conducted by the Centers for Disease Control and Prevention (2021). We use BRFSS’s post-stratification weights to construct state-level shares of daily smokers and to calculate obesity rates. Daily smokers are defined as respondents who report smoking every day and having smoked at least 100 cigarettes throughout their lifetimes. Obesity is defined as having a body mass index greater than 30.0.¹ BRFSS data are downloadable from https://www.cdc.gov/brfss/annual_data/annual_data.htm.
- *College shares*: State-level college shares are estimated separately from CPS data and from the Census and ACS Educational Attainment tables. The data for 1940 through 2000 are downloadable as Table 6 from <https://www.census.gov/data/tables/2000/dec/phc-t-41.html>. For 2010 through 2018, the data are in Table S1501 at the Census website (<https://data.census.gov/cedsci/table?q=S1501>).
- *Poverty*: State-level poverty rates from 1980 through 2019 are from the Census Bureau’s Historical Poverty Tables and are based on the CPS’s Annual Social and Economic Supplements (ASEC, also known as the March CPS). The data are at <https://www.census.gov/data/tables/time-series/demo/income-poverty/historical-poverty-people.html>.
- *Manufacturing shares*: Yearly manufacturing shares are drawn from the BEA’s estimates of employment by state and industry from 1969 through 2018 (Table SAEMP25), which can be downloaded from BEA’s Regional Data Tables web page (<https://apps.bea.gov/itable/iTable.cfm?ReqID=70&step=1>). The manufacturing share is defined as the percentage share of nonfarm employment, where employment is classified by the Standard Industrial Classification System (SIC) through 1997 and the North American Industrial Classification System (NAICS) after 1997.

¹Body mass index is defined as (weight in kilograms) \div (height in meters).²

- *Prescription quality:* Data from the Dartmouth Atlas of Medicare Prescription Drug Use (Munson et al. 2013) are used to measure prescribing quality based on 2008–2010 data for Medicare enrollees aged 65 and older. We include an index of both high-quality prescribing (for example, prescriptions for beta blockers and statins in the first six months following a heart attack) and poor-quality prescribing (for example, prescriptions for high-risk medications for which clinical risks exceed benefits, such as muscle relaxants for frail elderly patients). The Dartmouth data reference only the years 2008 through 2010, and be accessed at dartmouthatlas.org/downloads/tables/DAP_prescription_drug_tables_2010.zip.

2 Life Expectancy and Mortality by Race/Ethnicity

Figure A.1 parallels Figure 1 of the main text, which shows that national life expectancy was essentially flat after 2010 and fell slightly after 2014. Figure A.1 shows that this general pattern is also found among individual racial and ethnic groups.

Figure A.2 parallels Figure 6 of the main text, which documents rising state-level dispersion in both all-cause mortality and in deaths of despair. The appendix figure shows that white non-Hispanic and Black populations also display increasing state-level dispersion; by contrast, such dispersion *decreases* in the Hispanic population. This pattern demonstrates that racial patterns alone do not explain the results in Figure 6, while also illustrating some relevant intra-group dynamics in mortality that are beyond the scope of our paper.

3 Decomposition of State-Level Mortality Dispersion

Figure 4 in the main text is a decomposition of dispersion in log state-level mortality rates since 1992. The underlying model for this decomposition starts by noting that the mortality rate for persons aged 25 to 64 in state i and year t , m_{it} , can be written as the weighted average of education-specific mortality rates:

$$m_{it} = s_{it}^C m_{it}^C + s_{it}^N m_{it}^N, \quad (1)$$

where s_{it}^C is the population share of college-educated adults among the state’s 25- to 64-year-olds, $s_{it}^N = 1 - s_{it}^C$ is the non-college population share, and m_{it}^C and m_{it}^N are the midlife mortality rates for the state’s college and non-college populations, respectively. The individ-

ual mortality rates for the two education groups can be modelled in log form as

$$\begin{aligned}\ln(m_{it}^C) &= \mu_t + \epsilon_{it}^C \\ \ln(m_{it}^N) &= \mu_t + \lambda_t + \epsilon_{it}^N.\end{aligned}\tag{2}$$

In this system, the average mortality rate for college-educated Americans in year t is denoted by μ_t , while λ_t reflects a national non-college mortality “penalty,” which is well known to have been rising since the early 1990s. The notation implies that the national non-college rate in year t is $\mu_t + \lambda_t$. Consequently, the residuals ϵ_{it}^C and ϵ_{it}^N denote deviations from education-specific national rates in the given state and year.

As noted in the text, this structure allows us to decompose the sources of rising geographic disparities in mortality into four channels: (a) state-level changes over time in college shares s_{it}^C , whether because of higher educational attainment within the state or because of migration; (b) a widening of the national non-college mortality penalty λ_t , which will benefit states with high (or increasing) shares of college graduates; (c) an increase in the standard deviation of the state-year college residuals ϵ_{it}^C ; and (d) an increase in the standard deviation of residuals for non-college residents ϵ_{it}^N . Because our ultimate objective is to explain rising dispersion in *log* state-level mortality rates, changes in the national college effect μ_t do not matter, because this term affects (the log of) college and non-college mortality rates equally.

In the text, we use a series of counterfactual experiments to explore the relative contributions of these four channels to rising dispersion in state-level mortality. These experiments start by holding all model elements constant at their 1992 values. We then replace those baseline values with actual values of model elements in a series of steps, starting with actual college and non-college shares s_{it}^C and s_{it}^N . Using actual shares changes the implied values of m_{it} due to the presence of these shares in equation (1). Additional counterfactuals use actual values of λ_t , ϵ_{it}^C , and ϵ_{it}^N , which change the implied state-level mortality rates via their presence in the two-equation system (2). When actual values of the latter three elements are used, we exponentiate the new log mortality rates and then feed the resulting rates into equation (1), which generates a new set of state-specific mortality rates m_{it} . In all of the counterfactuals, we take logs of the implied m_{it} s before figuring the standard deviation, because Figure 4 depicts the standard deviation of log state-level mortality rates.

As noted in the main text, adding actual college and non-college shares s_{it}^C and s_{it}^N has only a modest impact on the standard deviation of state-level mortality, while the widening national college differential λ_t adds a bit more. Rising variation in the college residuals ϵ_{it}^C adds an additional amount of dispersion, but most of the increase in dispersion in state-level mortality over time is explained by a fanning out of the non-college residuals ϵ_{it}^N .

Our series of counterfactuals is not a formal variance decomposition, which would require us to account for changing covariances among the various model elements. One disadvantage

of our simpler approach is that the results of the exercise depend on the order in which actual model elements are introduced. In **Figure A.3**, however, we show that different orderings of the decomposition do not change our bottom line: Rising dispersion is caused largely by rising variance in the residual elements ϵ_{it}^C and ϵ_{it}^N , not by changing college shares or the rising educational penalty λ_t .

The model above can be modified to analyze the importance of state-wide “place effects,” if we are willing to assume a specific way in which place effects manifest themselves. For example, if place effects have equal effects on mortality among the state’s college and non-college populations, then we can replace the two-equation system (2) with

$$\begin{aligned}\ln(m_{it}^C) &= \phi_{it} + u_{it}^C \\ \ln(m_{it}^N) &= \phi_{it} + \lambda_t + u_{it}^N.\end{aligned}\tag{3}$$

In the new system, the place effect ϕ_{it} captures “average” mortality in the state, once the national non-college penalty λ_t has been accounted for. Conceptually, this place effect reflects the equally weighted average of the residuals ϵ_{it}^C and ϵ_{it}^N from the earlier two-equation system (2).² Because our place effect ϕ_{it} reflects average mortality, the residuals in the new system, u_{it}^C and u_{it}^N , must sum to zero and are therefore equal in absolute value and of opposite sign.

We can use the new system to construct a series of counterfactuals similar to those that appear in the main text. The main objective now is to see how much the new place effect ϕ_{it} matters for the widening dispersion of state-level log mortality rates. Results of this exercise appear in **Figure A.4**. As before, using actual college shares has little effect on state-level dispersion, and changes in the national education penalty λ_t also have a limited effect. Including actual values of place effects ϕ_{it} , however, matters a great deal. Our estimated place effect relies on the equal-weight assumption, and the results in the text (as well as previous research) indicate that place effects could well be stronger at the lower end of the income or educational distribution. Even so, we view these additional results as evidence that place effects with wide-ranging effects throughout a state could be an important driver of increased mortality dispersion over time.

Finally, note that our decomposition in (1) is only approximate due to the age-adjustment of the mortality rates. In practice, this means that the college shares used in the decomposition exercise in the main text are not exactly equal to actual college shares. However, we can verify the thrust of our main results with counterfactuals that use crude mortality rates, which allow for a decomposition with exact college shares. **Figure A.5** shows a robustness check using the model with separate college and non-college residuals, while **Figure A.6** depicts the robustness check using crude rates for the combined place-effect model.

²Rather than using equal weights, an alternative method would define ϕ_{it} as an average that uses the national share of college and non-college graduates over the time period considered.

4 Exploratory Regressions

The main text discusses some exploratory regressions that project the log of state-level overall mortality rates on various state-level explanatory variables, including real per capita personal income, the poverty rate, the manufacturing employment share, the smoking and obesity rates, and pharmaceutical-prescription quality. Results of these regressions are shown in **Table A.1**. Each column of the table depicts a regression of log mortality from a single year: 1992, 2000, or 2016. The right-hand-side variables are dated as of the same year, except the prescription-quality variables, which reference only the years 2008 through 2010.

The table has two major takeaways. First, comparing columns (1), (3), and (5) demonstrates the increasing importance of income over time. A significantly negative income coefficient appears by 2000, and this coefficient becomes even more negative in 2016. Second, comparing columns (2), (4), and (6) shows that when other variables besides income are included, the negative effect of state-level income is reduced. This reduction indicates that high-income states differ from low-income states along a variety of dimensions relevant for health, which are being captured in some ways by these additional variables.

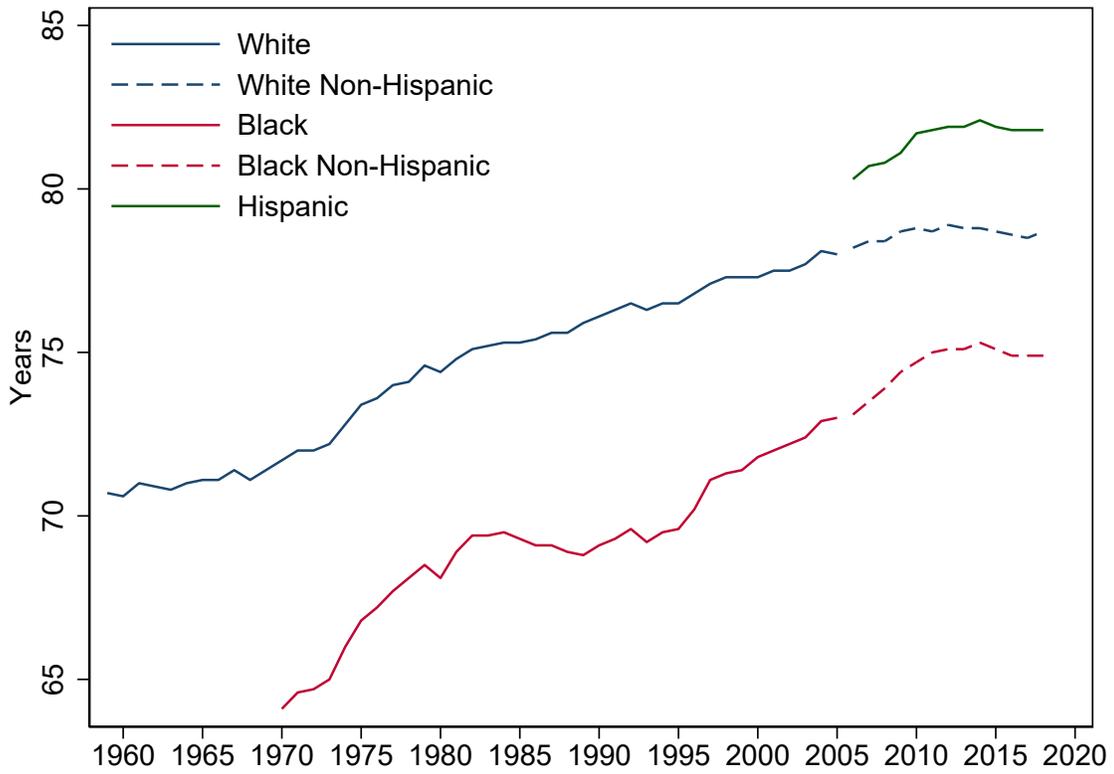


Figure A.1. US LIFE EXPECTANCY AT BIRTH BY RACE/ETHNICITY.

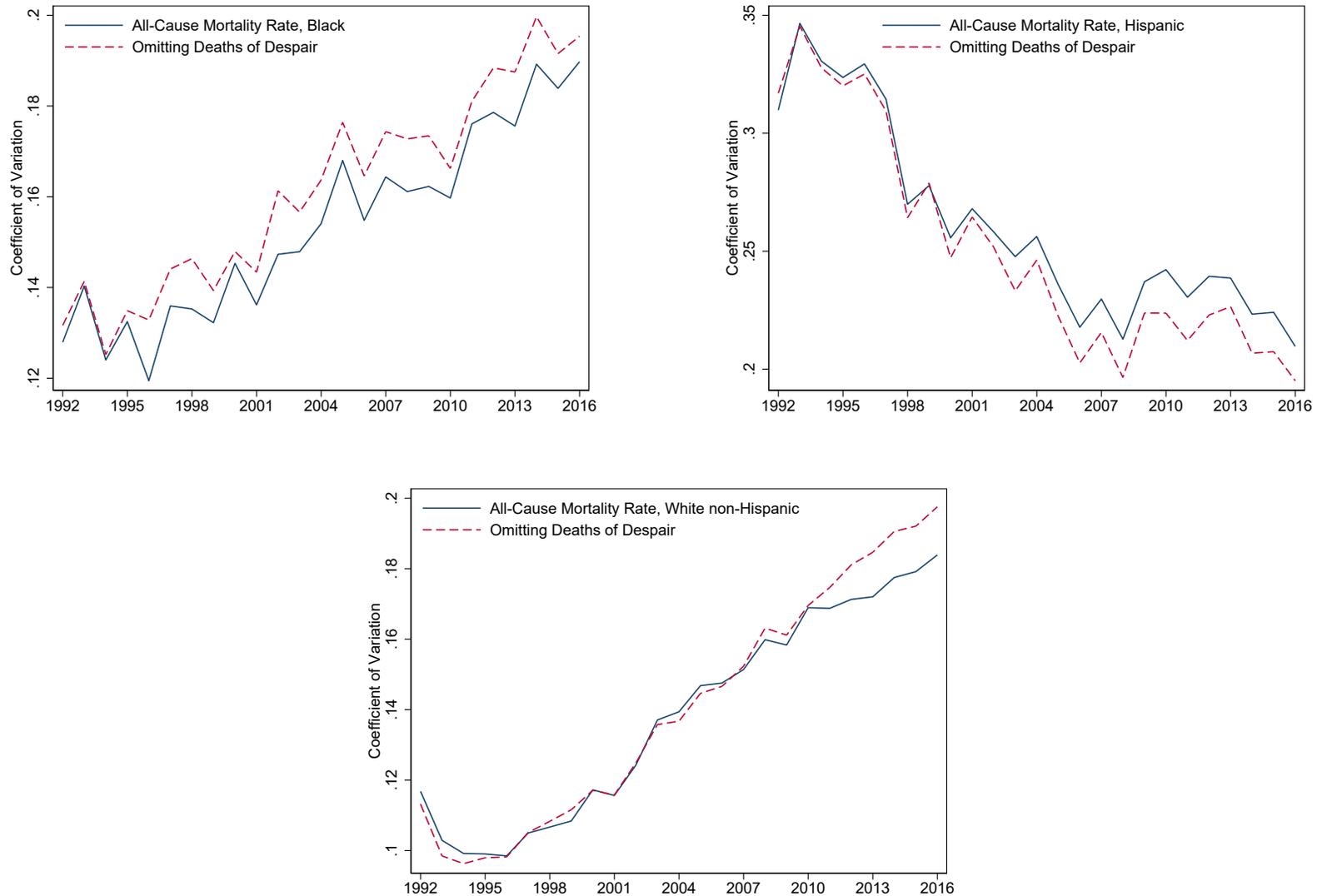


Figure A.2. COEFFICIENTS OF VARIATION FOR ALL-CAUSE MORTALITY AND MORTALITY OMITTING DEATHS OF DESPAIR, BY RACE/ETHNICITY. *Note:* This figure shows that state-level dispersion in mortality rises over time among Black and white non-Hispanic populations, while mortality dispersion declines in the Hispanic population. Mortality rates are age adjusted and correspond to persons aged 25 to 64. Deaths of despair are deaths attributed to poisonings, suicide, or cirrhosis (ICD9: 571, E950-959, E850-860 and E980-982; ICD10: K70 and K73-74, X60-84 and Y87.0, X40-X45 and Y10-15). Coefficients of variation are population weighted.

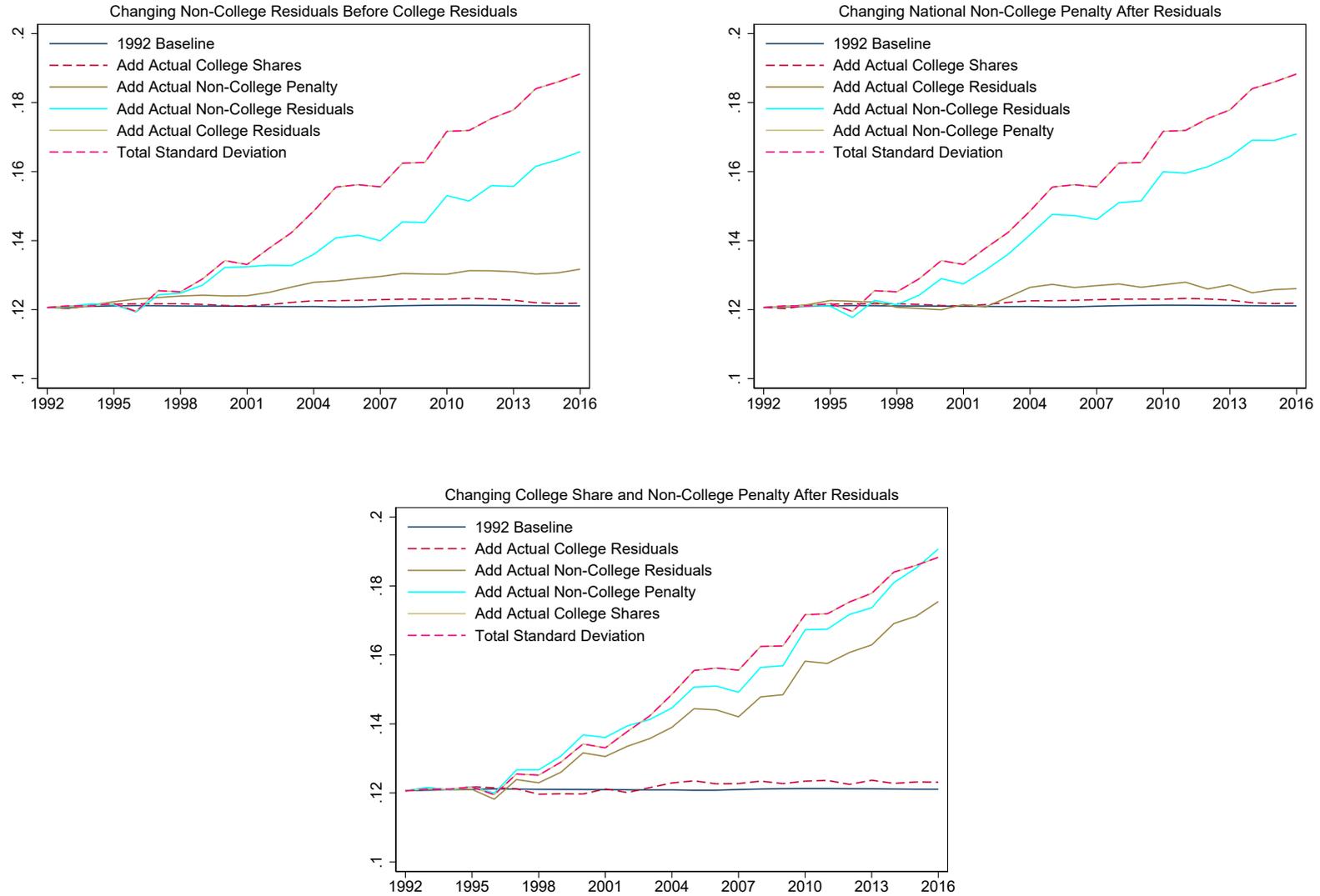


Figure A.3. CHANGING THE ORDER OF MODEL ELEMENTS IN THE DECOMPOSITION OF MORTALITY DISPERSION. *Note:* Each panel displays standard deviations for a series of counterfactual sets of log state-level mortality rates, which are built up using the statistical framework outlined in section 3 of this appendix and discussed more briefly in the main text. The panels are distinguished by the order in which actual values of model elements are incorporated into the statistical framework. Across all panels, rising state-level dispersion is driven largely by rising dispersion in college and non-college residuals, not by changing college shares or the rising non-college mortality penalty.

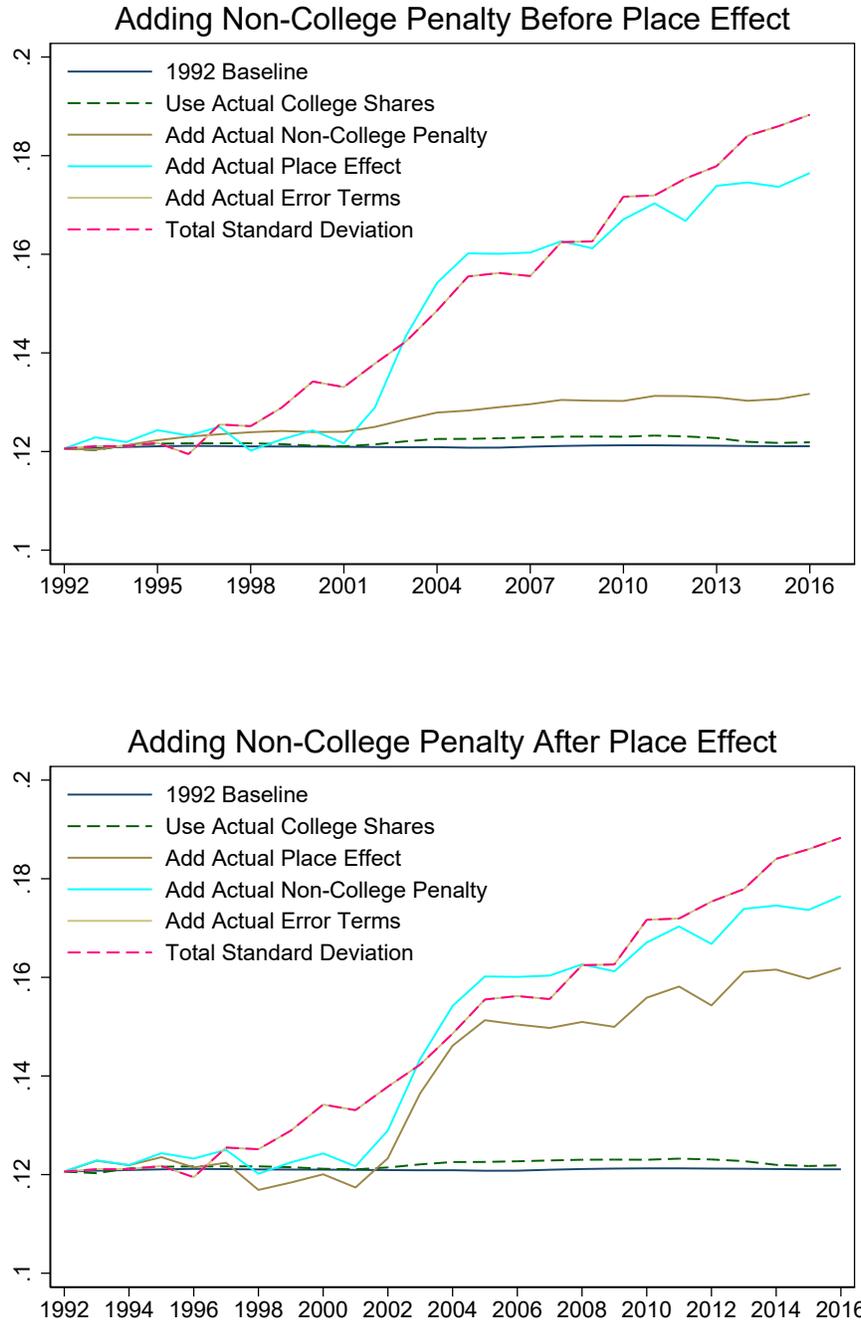


Figure A.4. COUNTERFACTUAL EXERCISE IN PLACE-EFFECT MODEL. *Note:* Each panel displays standard deviations for a series of counterfactual sets of log state-level mortality rates, built up using the “place effect” model outlined in section 3 of this appendix. In this model, the non-college and college residuals of the basic model are essentially collapsed into a single place effect that is shared by both educational groups. The figure confirms that rising dispersion in log state-level mortality rates is driven largely by the rising dispersion in place effects, regardless of the order in which actual model elements are incorporated into the statistical framework.

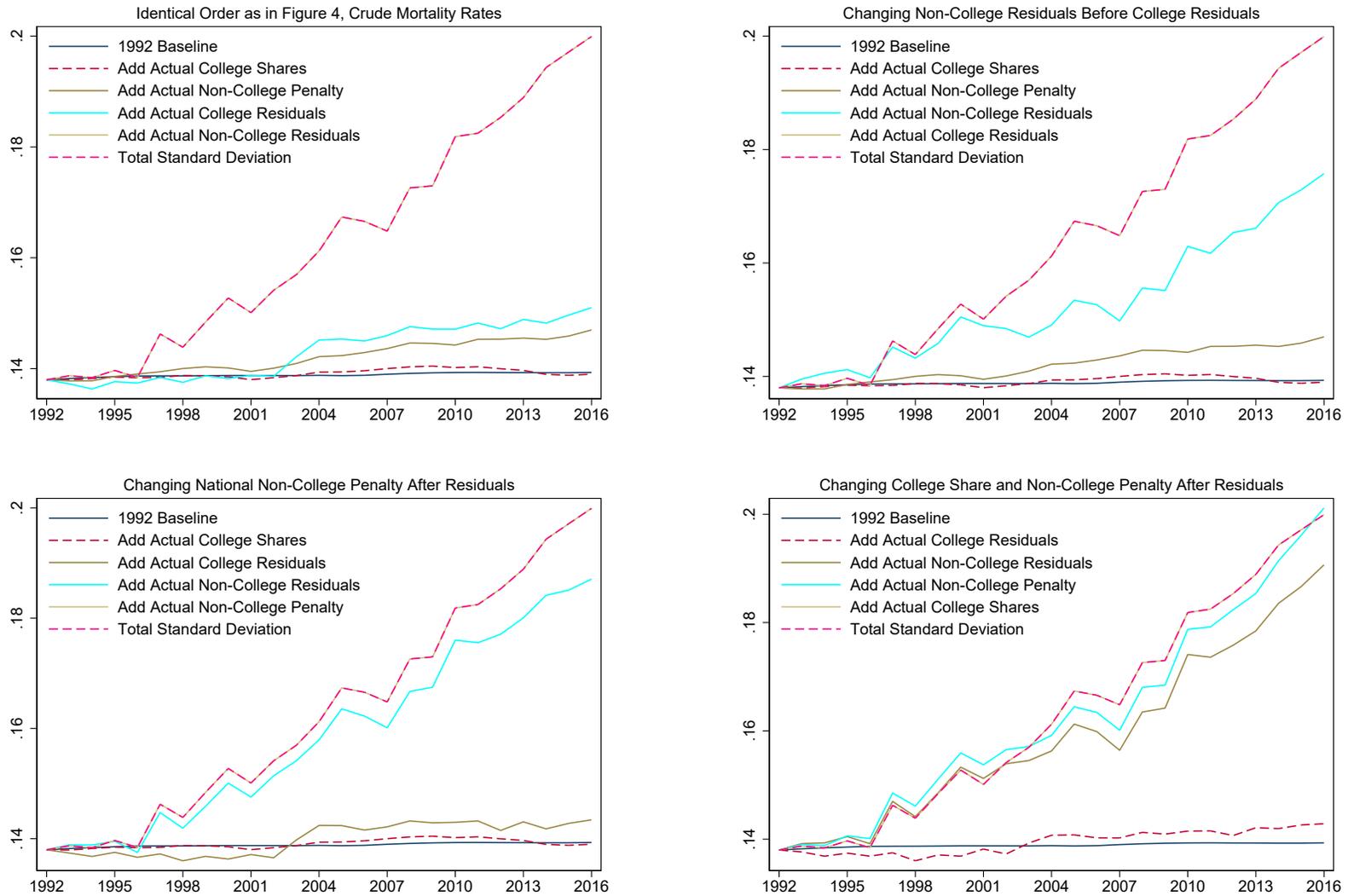


Figure A.5. DECOMPOSITION USING CRUDE RATHER THAN AGE-ADJUSTED MORTALITY RATES *Note:* Each panel displays standard deviations for a series of counterfactual sets of log state-level mortality rates. All panels use crude rather than age-adjusted mortality data (the decomposition is exact only with crude mortality rates). The figure confirms that even with crude rates, rising state-level dispersion is driven largely by rising variance in the college and non-college residuals. This fact remains true regardless of the order in which actual model elements are incorporated into the statistical framework.

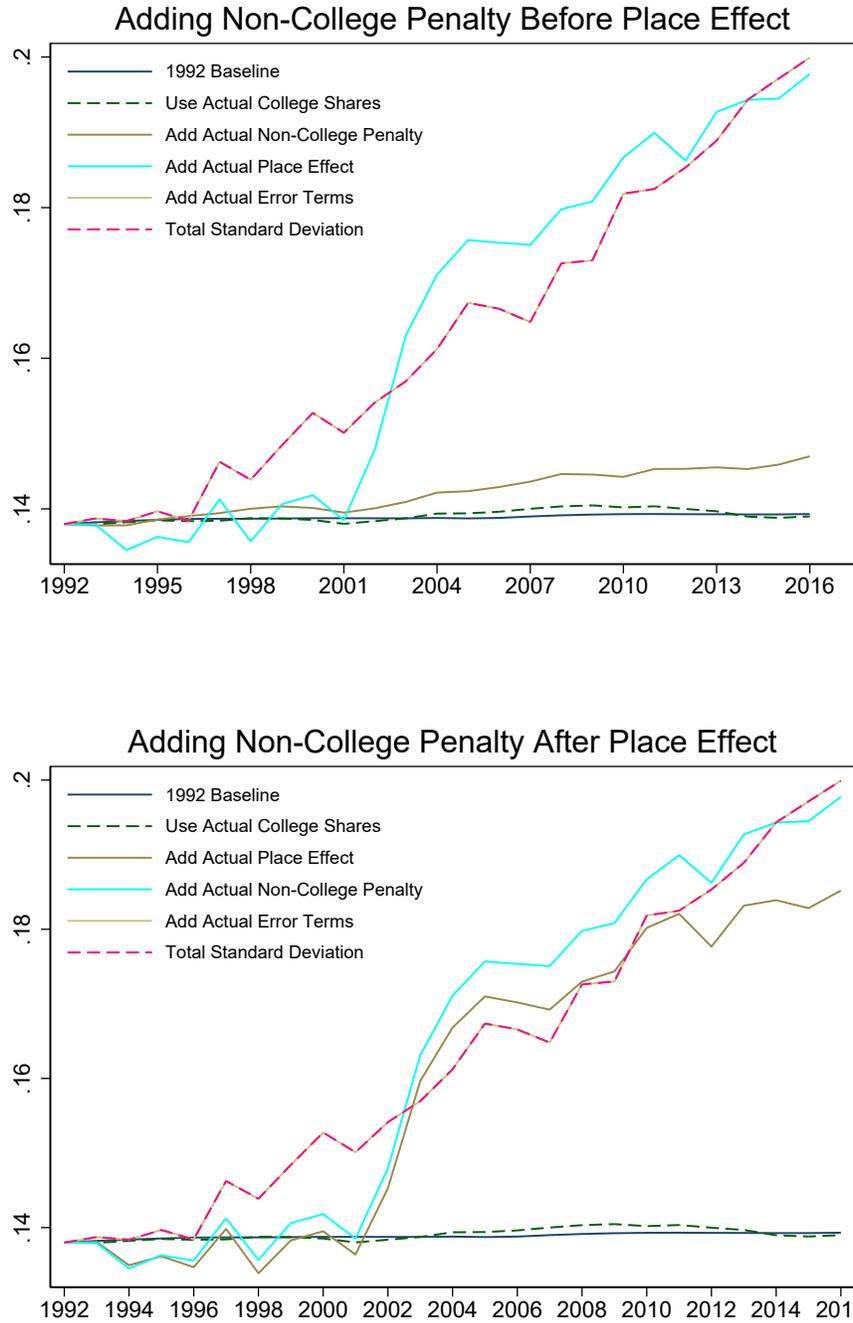


Figure A.6. COUNTERFACTUAL EXERCISE IN PLACE-EFFECT MODEL USING CRUDE MORTALITY. *Note:* These panels show a robustness check for the counterfactual exercise that uses crude mortality rates. (See Figure A.4 for the place effect model using age-adjusted rates.) The panels show that the approximate nature of the decomposition when using age-adjusted mortality rates is not driving our results in the place effect model.

| | (1) | (2) | (3) | (4) | (5) | (6) |
|--|-------------------|------------------------|---------------------|------------------------|----------------------|-----------------------|
| Year | 1992 | 1992 | 2000 | 2000 | 2016 | 2016 |
| Log of Real Per Capita Personal Income (\$2012) | -0.173 (0.180) | 0.527*** (0.124) | -0.466** (0.141) | 0.304*** (0.0859) | -0.977*** (0.131) | -0.260* (0.113) |
| Poverty Share | | 0.0182*** (0.00496) | | 0.0103 (0.00528) | | 0.00626 (0.00904) |
| Smoking Rate | | 1.062* (0.520) | | 1.346** (0.462) | | 2.885*** (0.492) |
| Obesity Rate | | 0.536 (0.966) | | 0.824 (1.103) | | 0.185 (0.761) |
| Manufacturing Share | | -0.355 (0.267) | | -0.574 (0.389) | | -1.081 (0.549) |
| Good-Prescription Rate | | -0.00422 (0.00885) | | -0.0116 (0.00867) | | -0.00933 (0.00656) |
| Risky-Prescription Rate | | 0.0120** (0.00364) | | 0.0172*** (0.00371) | | 0.0107** (0.00340) |
| R-squared | 0.025 | 0.790 | 0.202 | 0.815 | 0.578 | 0.893 |
| <i>N</i> | 44 | 44 | 44 | 44 | 44 | 44 |

Table A.1. STATE-LEVEL REGRESSIONS OF LOG AGE-ADJUSTED MIDLIFE MORTALITY RATES IN SELECTED YEARS. *Note:* Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Comparing columns (1), (3), and (5) demonstrates the increasing correlation of mortality rates with income over time. However, columns (2), (4) and (6) show that income is likely standing in for other effects, as the income coefficient is made significantly less negative after adding additional behavioral, labor market, and healthcare-quality variables. Also note the rising correlation of the smoking rate with mortality; this pattern suggests that smoking could be a “sentinel” measure reflecting changes in both behavior and state-level policies. The number of observations is fewer than 50 due to limited data availability of some covariates.

Appendix References

- Bureau of Economic Analysis. 2021. “Personal Consumption Expenditures Price Index.” United States Department of Commerce. Accessed August 5, 2021, retrieved from FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/series/PCEPI>.
- Bureau of Economic Analysis. n.d.a. “Personal Income, Population, Per Capita Personal Income, Disposable Personal Income, and Per Capita Disposable Personal Income (SAINC1).” United States Department of Commerce. Accessed August 5, 2021, retrieved from <https://apps.bea.gov/itable/itable.cfm>.
- Bureau of Economic Analysis. n.d.b. “Total Full-Time and Part-Time Employment by Industry, SAEMP25.” United States Department of Commerce. Accessed February 2, 2021, retrieved from <https://apps.bea.gov/itable/itable.cfm>.
- Centers for Disease Control and Prevention. 2001. “Comparability of Cause of Death Between ICD–9 and ICD–10.” United States Department of Health and Human Services. Accessed August 6, 2021, retrieved from https://ftp.cdc.gov/pub/health_statistics/NCHS/Datasets/Comparability/icd9_icd10/Comparability_Ratios_tables.xls.
- Centers for Disease Control and Prevention. 2018. “2018 General Equivalence Mappings.” United States Department of Health and Human Services. Accessed August 6, 2021, retrieved from <https://www.cms.gov/Medicare/Coding/ICD10/Downloads/2018-ICD-10-CM-General-Equivalence-Mappings.zip>.
- Centers for Disease Control and Prevention. 2021. “Behavioral Risk Factor Surveillance System (BRFSS).” US Department of Health and Human Services. Accessed August 6, 2021, retrieved from https://www.cdc.gov/brfss/annual_data/annual_data.htm.
- Centers for Disease Control and Prevention, National Center for Health Statistics. 2000. “Compressed Mortality File 1968-1978 on CDC WONDER Online Database.” US Department of Health and Human Services. Data are compiled from Compressed Mortality File CMF 1968-1988, Series 20, No. 2A, 2000. Accessed August 4, 2021, retrieved from <https://wonder.cdc.gov/cmfi-icd8.html>.
- Centers for Disease Control and Prevention, National Center for Health Statistics. 2003. “Compressed Mortality File 1979-1988 and 1989-1998 on CDC WONDER Online Database.” US Department of Health and Human Services. Data are compiled from Compressed Mortality File CMF 1968-1988, Series 20, No. 2A, 2000 and CMF 1989-1998, Series 20, No. 2E, 2003. Accessed August 4, 2021, retrieved from <https://wonder.cdc.gov/cmfi-icd9.html>.

- Centers for Disease Control and Prevention, National Center for Health Statistics. 2017a. “Compressed Mortality File 1999-2016 on CDC WONDER Online Database.” US Department of Health and Human Services. Data are from the Compressed Mortality File 1999-2016 Series 20 No.2U, 2016, as compiled from data provided by the 57 vital statistics jurisdictions through the Vital Statistics Cooperative Program. Accessed August 4, 2021, retrieved from <https://wonder.cdc.gov/cmfi-icd10.html>.
- Centers for Disease Control and Prevention, National Center for Health Statistics. 2017b. “Restricted-Use Vital Statistics Data 1989–2017.” US Department of Health and Human Services. Data are as compiled from data provided by the 57 vital statistics jurisdictions through the Vital Statistics Cooperative Program.
- Centers for Disease Control and Prevention, National Center for Health Statistics. 2020. “Underlying Cause of Death 1999-2019 on CDC WONDER.” US Department of Health and Human Services. Data are from the Multiple Cause of Death Files, 1999-2019, as compiled from data provided by the 57 vital statistics jurisdictions through the Vital Statistics Cooperative Program. Accessed July 10, 2021, retrieved from <https://wonder.cdc.gov/ucd-icd10.html>.
- Flood, Sarah, Miriam King, Renae Rodgers, Steven Ruggles, and J. Robert Warren. 2020. “Integrated Public Use Microdata Series (IPUMS), Current Population Survey: Version 8.0 Basic Monthly.” Accessed April 28, 2020; retrieved from <https://cps.ipums.org/cps/>.
- Munson, Jeffrey, Nancy E. Morden, David C. Goodman, Luca F. Valle, and John E. Wennberg. 2013. *The Dartmouth Atlas of Medicare Prescription Drug Use*. Lebanon, NH: The Dartmouth Institute for Health Policy and Clinical Practice. Accessed December 14, 2020, retrieved from dartmouthatlas.org/downloads/tables/DAP_prescription_drug_tables_2010.zip.
- National Cancer Institute, Surveillance Research Program. 2021. “Surveillance, Epidemiology, and End Results (SEER) Program Populations (1969-2019).” National Institutes of Health. Accessed August 6, 2021, retrieved from <https://seer.cancer.gov/popdata/download.html>.
- United States Mortality Database. 2021. University of California, Berkeley. Accessed July 10, 2021, retrieved from <https://usa.mortality.org/>.
- US Census Bureau. 2006. “A Half-Century of Learning: Historical Statistics on Educational Attainment in the United States: 1940 to 2000.” United States Department of Commerce. Accessed August 18, 2020, retrieved from <https://www.census.gov/data/tables/2000/dec/phc-t-41.html>.

US Census Bureau. 2010–2018. “Educational Attainment (S1501) from the American Community Survey.” United States Department of Commerce. Accessed August 18, 2020, retrieved from <https://data.census.gov/cedsci/table?q=S1501>.

US Census Bureau. 2019. “National Counties Gazetteer Files.” United States Department of Commerce. Accessed August 6, 2021, retrieved from https://www2.census.gov/geo/docs/maps-data/gazetteer/2019_Gazetteer/2019_Gaz_counties_national.zip.

US Census Bureau. 2021a. “Annual Estimates of the Population for the U.S. and States, and for Puerto Rico.” United States Department of Commerce. Accessed August 6, 2021, retrieved from FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/release/tables?rid=118&eid=259194>.

US Census Bureau. 2021b. “Historical Poverty Tables: People and Families: 1959 to 2019.” United States Department of Commerce. Accessed August 6, 2021, retrieved from <https://www.census.gov/data/tables/time-series/demo/income-poverty/historical-poverty-people.html>.