NBER WORKING PAPER SERIES

THE EFFECT OF HEALTH INSURANCE ON MORTALITY: POWER ANALYSIS AND WHAT WE CAN LEARN FROM THE AFFORDABLE CARE ACT COVERAGE EXPANSIONS

Bernard Black
Alex Hollingsworth
Leticia Nunes
Kosali Simon

Working Paper 25568
http://www.nber.org/papers/w25568

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2019

We thank Marcus Dillender, William Evans, Ted Joyce, Robert Kaestner, Amanda Kowalski, Sarah Miller, Dan Sacks, Ben Sommers, Coady Wing, Engy Ziedan, and participants in workshops at Bar-Ilan University Conference on Law and Big Data (May 2018); Georgia State University, Economics Department (August 2018); Hebrew University of Jerusalem, Economics Department (Nov. 2018); Interdisciplinary Center Herzliya (May 2018); ASSA 2018; Stanford University Health Economics Seminar (May 2018); ASHEcon (2018); APPAM (2017); Tel Aviv University, Economics Department (Nov. 2018), Texas A&M (May 2017); and the UK Institute for Fiscal Studies (Nov. 2018). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Effect of Health Insurance on Mortality: Power Analysis and What We Can Learn from the Affordable Care Act Coverage Expansions
Bernard Black, Alex Hollingsworth, Leticia Nunes, and Kosali Simon
NBER Working Paper No. 25568
February 2019
JEL No. I1

ABSTRACT

A large literature examines the effect of health insurance on mortality. We contribute by emphasizing two challenges in using the Affordable Care Act (ACA)'s quasi-experimental variation to study mortality. The first is non-parallel pretreatment trends. Rising mortality in Medicaid non-expansion relative to expansion states prior to Medicaid expansion makes it difficult to estimate the effect of insurance using difference-in-differences (DD). We use various DD, triple difference, age-discontinuity and synthetic control approaches, but are unable to satisfactorily address this concern. Our estimates are not statistically significant, but are imprecise enough to be consistent with both no effect and a large effect of insurance on amenable mortality over the first three post-ACA years. Thus, our results should not be interpreted as evidence that health insurance has no effect on mortality for this age group, especially in light of the literature documenting greater health care use as a result of the ACA. Second, we provide a simulation-based power analysis, showing that even the nationwide natural experiment provided by the ACA is underpowered to detect plausibly sized mortality effects in available datasets, and discuss data needs for the literature to advance. Our simulated pseudo-shocks power analysis approach is broadly applicable to other natural-experiment studies.

Bernard Black
Northwestern University
Pritzker School of Law
375 East Chicago Ave
Chicago, IL 60201
bblack@northwestern.edu

Alex Hollingsworth
School of Public and Environmental Affairs
Indiana University
375C
1315 E. Tenth St.
Bloomington, IN 47405
hollinal@indiana.edu

Leticia Nunes
Department of Economics
Fundação Getulio Vargas
leticiafcnunes@gmail.com

Kosali Simon
School of Public and Environmental Affairs
Indiana University
Rm 443
1315 East Tenth Street
Bloomington, IN 47405-1701
and NBER
simonkos@indiana.edu

A Problem Set version of code to demonstrate shock-based power calculation approach is available at https://github.com/hollina/health_insurance_and_mortality

I. Introduction

The relationship between health insurance and mortality is at the center of much empirical inquiry in the health economics literature. Since the first rigorous study of this relationship through the RAND Health Insurance Experiment, researchers have studied this question using varying study designs and populations, finding mixed results on the existence and strength of any relationship; a recent literature review found over 200 studies published on the topic (Gaudette et al., 2016). Many papers in this literature focus on mortality as an extreme, but readily measurable outcome. Most earlier studies, including the RAND Experiment, studies of Medicare, and the more recent Oregon Health Insurance Experiment find no statistically significant impacts of health insurance on overall mortality for the general adult population (Levy and Meltzer 2008; Finkelstein and McKnight 2008; Finkelstein et al., 2012), [1] but several more recent studies report mortality reductions from state or federal insurance expansions for adults (e.g. Sommers, Long, and Baicker, 2014). A separate literature finds health and mortality gains from health insurance for children (e.g., Currie and Gruber, 1996a,b; Wherry and Meyer, 2015; Brown, Kowalski, and Lurie 2017). Health insurance expansions have also been shown to have substantial improvements in access to health care and in health status (Currie and Gruber 1996a, Sommers et al 2013, Simon et al 2018 JPAM,). Other studies examine the effect of insurance on non-physical health outcomes, such as mental health stress levels and financial health, finding improvements (e.g., Hu et al, 2016; Baicker et al. 2013).

The Affordable Care Act produced substantial insurance expansions for the low-income, non-elderly adult population (e.g. Kaestner et al., 2015; Wherry & Miller, 2016; Frean et al., 2017; Simon et al., 2017; Courtemanche et al., 2017). These expansions provide a new opportunity to study the link between health insurance and mortality, and to examine issues of statistical power for these studies and, more generally, natural experiment studies of low-frequency outcomes. Our study examines this relationship using mortality microdata from 1999-2016. We use both difference-in-differences (DiD) and triple-difference/age discontinuity approaches to study the effect of state Medicaid expansions, and ACA expansion more generally, on mortality during the

[1] These are studies of the effect on mortality of health insurance, not health care. For example, Finkelstein and McKnight (2008) observe that "part of the explanation for [finding no mortality effect could be that], prior to Medicare, elderly individuals with life-threatening, treatable health conditions sought care even if they lacked insurance, as long as they had legal access to hospitals."

first three post-ACA years (2014-2016). We exploit heterogeneity in assignment to "treatment" (health insurance) and potential treatment effect heterogeneity along several dimensions: healthcare amenable vs. non-amenable causes of death; specific major causes of death (cancer, heart disease, diabetes, and respiratory disease); and sociodemographic factors at the individual (gender, race/ethnicity, and education) and the county (baseline percent uninsured and percent in poverty) levels. Our triple-difference/age-discontinuity design compares the near-elderly (age 55-64) to the young-elderly (ages 65-74), who were already covered by Medicare. We focus on the near-elderly, both because they are more likely than younger persons to have health conditions for which healthcare is important for survival, and because focusing on this age band makes the above and below-65 groups more comparable. Our age-discontinuity approach is similar to the Finkelstein and McKnight (2008) study of Medicare, except their treatment group is our control group. We obtain similar results in analyses using broader age ranges (age 45-64, or all non-elderly adults).

We do not find a statistically significant pattern of results consistent with Medicaid expansion causing mortality changes, but we also cannot rule out large effects in either direction. We note here that prior evidence on the effects of insurance expansion on mortality lead one to expect modest effects of incremental insurance expansions on mortality. Reasons for modest overall effects include: many of those in greatest need of healthcare are already insured; and many uninsured persons already receive substantial healthcare. One reason for our "null result" is that the first stage on insurance coverage is weak: our principal identifying variation (the relative change in uninsurance rates for Medicaid expansion versus non-expansion states) amounts to a very small fraction of the population. The average increase in health insurance coverage attributable to Medicaid expansion over 2014-2016 is only around 1.1% for persons aged 50-64, and only around 4% even when we hone in on low-educated populations; precise income measures used to determine ACA eligibility are unavailable in mortality data. A second reason for failure to reject the null of no effect is a high level of "noise" - substantial background variation in mortality, and mortality trends, across states and demographic groups. A third reason is that mortality is a low-frequency outcome. We note too that effects of health insurance on mortality are more likely to emerge over a long time frame.

Our second contribution is to highlight that observational studies can often benefit from performing and reporting power analyses. We use a simulation-based power analysis, in which we impose treatment effects of varying sizes on actual data during the pre-treatment period, and assess whether ACA expansion effects on mortality of plausible size can be detected with our data. We conclude that even the nationwide natural experiment provided by the ACA is severely underpowered to detect effects on mortality of plausible size in county-level death certificate data. To reliably detect effects of insurance coverage on mortality, one would need very large-sample panel data on individuals, which is not currently available. Such data could include information on health, income, and insurance status, which would allow the study to focus on subsamples with a larger first stage and/or higher sensitivity of health and mortality to healthcare use. Even with such hypothetical data, it is likely that only fairly large effects of health insurance on mortality could be reliably detected.

We estimate power using our pre-treatment period data (pre-2014) by first applying a pseudo-shock to health insurance rates at the beginning of 2011 as if the ACA expansion had occurred then. We choose pseudo-treated states at random, and then apply pseudo treatment effect (mortality shocks) of different sizes to the group of pseudo-treated states (by randomly removing deaths from our mortality data). We repeat this process 1,000 times. We then assess the likelihood that these pseudo shocks we introduce in 2011-2013 would be detected, using methods similar to our actual specifications. This approach (applying simulated treatment effects to actual data, drawn from a period when no effect should exist) can be applied to a wide-variety of research settings, including both structural and non-parametric work; The Appendix provides sample Stata code for implementing our power analysis using publicly available mortality data from CDC Wonder.

The minimum detectable effect (MDE) – the minimum reduction in amenable mortality for all persons aged 55-64 years in expansion states, detectable at the 95% confidence level (two-tailed test), 80% of the time (a standard threshold for a study to be considered adequately powered) as a result of a state expanding Medicaid is about 0.018. Together with a 0.011 first-stage, this implies that the MDE is roughly a 160% drop in amenable mortality among the newly insured.[2] The DD and triple difference models have similar power. Power does not improve when we examine subgroups: non-parallel trends remain common, the first-stage remains modest, and the gain in

[2] This estimate assumes that baseline amenable mortality rate is the same for those who differentially gained insurance in expansion states as for the general population, controlling for observable covariates. If the baseline mortality rate for the newly, differentially insured was twice that of the overall population, the MDE would be half as large, thus 80% rather than 160%.

power from a higher first-stage and a higher base mortality rate is more than offset by smaller sample size.

By comparison to the very large MDE from the natural experimental variation available in our study, the historic introduction of sulfa drugs reduced maternal mortality by 24-36% (Thomasson and Treber, 2008; Jayachandran et al., 2012). Finkelstein and McKnight (2008) found no significant effect of the introduction of Medicare on mortality for those aged 65-74 years (point estimate after 5 years = -0.15%; 95% CI [-3.9%, +3.6%]); Card, Dobkin, and Maestas (2004) use an age-discontinuity design and find no reduction in mortality at age 65 (point estimate +0.5%, 95% CI [-3.3%, +4.3%]); the RAND Experiment found no significant overall effect of health insurance on mortality but found a 10% reduction in mortality for a subsample of persons with vulnerable health; and the Oregon Experiment found no significant effect, with a point estimate of -13% but a wide 95% confidence interval (95% CI [-39%, +13%]).[3] Large effects are also unlikely because prior research finds that the uninsured already consume substantial healthcare -- about 80% as much as the insured (e.g., Black et al., 2017). Our prior expectation, considering the near-zero estimates and confidence intervals in the largest prior studies (Finkelstein and McKnight, 2004; Card, Dobkin and Maestas, 2004), the substantial healthcare consumed by the uninsured, the imperfect safety net that already covers some vulnerable populations (e.g., the elderly and the disabled), and the availability of emergency care regardless of insurance status (Card, Dobkin, and Maestas, 2009), were that any effect of the 2014 insurance expansion on mortality was unlikely to exceed 10% for the newly insured, and that any effect would likely appear only over time.

Combining this past literature with a power analysis perspective, we expect that if significant effects of expanding health insurance eligibility on general adult mortality are found, these are likely to greatly overstate actual magnitudes. Reasons to re-examine results from low-powered studies include: they may draw from the right tail of a probability distribution; failure to adequately balance treated and control units or address non-parallel trends; specification searches; and "file-drawer bias" (the tendency for insignificant results to remain unpublished). McCrary, Christensen and Fanelli (2016) propose a minimum *t*-statistic around 3 to correct for file-drawer bias alone.

[3] Sulfa drugs: See Jayachandran et al. (2012), table 1. Medicare adoption: See Finkelstein & McKnight (2008), App. A. Oregon Experiment: See Finkelstein et al. (2012), Table IX. Medicare age discontinuity: see Card et al (2004) table 11, RAND experiment: see Brook et al. (1983) Table 7.

Power analyses are common in the design (*ex ante*) stage of a randomized trial; researchers use them to ensure that the trial does not "fail" to find a true effect due to inadequate sample size. They are rare, however, for DiD and other observational studies. Ioannidis et al. (2017) and McCloskey (1985) criticize the failure of economics researchers to conduct power analyses. DiD and other shock-based, observational studies with panel data would often benefit from assessing plausible effect sizes and conducting power analyses, ideally in an explicit "design stage" (with outcomes hidden; see Rubin, 2008). Conducting these analyses can reduce the chance of inadvertently publishing false positive results or results with inflated magnitudes (Button et al., 2013; Gelman and Carlin, 2014).[4]

For example, we find non-parallel pre-treatment trends between treated and control states; mortality among those aged 55-64 drops fairly substantially in treated states over 2009-2013 relative to control states (Figure 2). The triple difference design cannot fully address this problem, because we also find non-parallel within-state trends for persons aged 55-64 compared to those aged 65-74, which vary across subgroups. DD and triple difference regression estimates ignore these non-parallel trends. For example, we find implausibly large, statistically significant effects of ACA expansion on mortality for blacks and Hispanics, in both DD and triple difference specifications (Appendix Table A2). The power analysis and parallel trends examination (for which a long pre-treatment period can be important) reduce the likelihood that we would inadvertently interpret these significant coefficients as robust results.[5]

We note several limitations of our work. First, our analysis should not be interpreted as evidence that health insurance does *not* affect mortality or health, either overall or for particular diseases or subgroups. Second, studying mortality with ACA-induced variation in health insurance is marginal in three senses: (i) those previously uninsured (implying average lower demand for health insurance; see Kowalski's (2018) evidence on better health among new enrollees in Massachusetts reform than existing enrollees) may experience lower marginal gains from insurance than the already insured; (ii) prior policy interventions already provide emergency care and some healthcare access for vulnerable populations; and (iii) access to health insurance

[4] We discuss below the limited prior examples we have found on use of a simulated power analysis in applied economics research; none involve imposing a simulated treatment effect on actual data.

[5] As we were finalizing this draft, we became aware of Borgschulte and Vogler (2019), who find a post-ACA drop in mortality attributable to Medicaid expansion for both amenable and non amenable causes of death.

does not equate to access to healthcare, as even the uninsured consume substantial healthcare, so that some insurance-induced healthcare could be at the "flat" (or even the downslope) of the marginal benefit curve. We also study a relatively short post-shock time frame, yet any effects of health insurance on mortality may appear only over a longer time frame. However, our simulations suggest that longer-term effects on mortality, with plausible effect sizes, cannot be reliably detected with currently available datasets. Moreover, concern with non-parallel trends during the treatment period increases as one moves further away from the shock. Thus, additional years of data, using existing sources, are unlikely to allow a convincing longer-term effect to emerge.

In Part II we summarize the prior literature on the relationship between health insurance and mortality. This literature presents a mixed picture. There is no consistent evidence for statistically significant effects of insurance on mortality for the general adult population. There are some effects for specific vulnerable populations such as those with HIV, but not for others, such as those with a disability. Part III provides an overview of the conceptual concerns that inform our analysis. Part IV summarizes the ACA insurance expansions. Part V describes our data and presents summary statistics. Part VI summarizes our empirical approach. Part VII presents our results. Part VIII presents our power analysis, highlights the limited sources of identifying variation and the risk of false positives, and assesses which data and sample sizes might provide adequate power. Part IX concludes.

II. Prior Research

A. The Effect of Health Insurance on Health and Mortality

Our first contribution, on whether Medicaid expansion predicts lower mortality, fits into a large literature that examines the connection between health insurance and health status. This literature spans experimental and quasi-experimental settings, and examines morbidity and mortality, physical and mental health, elderly and non-elderly adults, pregnant women, children, infants, short- and long-run effects, and specific diseases and demographic subpopulations.

For our first aim, we focus on the effect of health insurance on mortality in the general adult population.[6] Historically, the first rigorous evidence on how health insurance affects health

---

[6] In early research using a natural experiment, Currie and Gruber (1996a,b) find that Medicaid expansions in the late 1980s and early 1990s reduced infant mortality by 8% and all-cause child mortality by 5%.However, Howell et al (2010) find that the effects of Medicaid expansion on child and infant mortality are limited to accidental deaths, not disease-related deaths – a puzzling result, since emergency care regardless of insurance has been required since 1996

and mortality comes from the RAND Health Insurance Experiment (Brook et al., 1983; Keeler, 1985; Newhouse, 1993) which provided experimental exposure to varying degrees of insurance generosity; none of the study subjects was fully uninsured. Brook et al. (1983) found no significant overall effect on mortality for the full sample (of persons aged 14 to 61, followed for 3-5 years (point estimate -0.02; 95% CI [-0.05, +0.02]), but found 10% lower mortality for high-risk individuals who received generous insurance. The RAND HIE also found some improvements in blood pressure for low-income populations receiving generous insurance, but otherwise found limited evidence that generous insurance led to improved health.

Finkelstein and McKnight (2008) study Medicare's introduction in 1965, which remains the largest health insurance policy change in US history. Their first stage is around 75%, because private insurance for the elderly was uncommon pre-Medicare (Finkelstein, 2007). Finkelstein and McKnight (2008) find a 40% drop in out-of-pocket medical expenditures, but no discernible mortality effects over a 10-year period (point estimate after 5 years = -0.15%; 95% CI [-3.9%, +3.6%]). Finkelstein and McKnight observe that these results may be due to the fact that prior to Medicare, those with life-threatening but treatable conditions likely sought care even if they were uninsured.

Card, Dobkin, and Maestas (2004) exploit the age-65 discontinuity in coverage using more recent data from 1989-1998; they find no significant effect of turning 65 on population mortality (point estimate +0.5%, 95% CI [-3.3%, +4.3%]).[7] Their first stage is around 8% for the full sample (Table 3) and 14% for a low-education subsample. In a related study that speaks to possible mechanisms, Card, Dobkin, and Maestas (2009) find a drop in mortality at age 65 among those admitted to hospital through the ED for severe, non-deferrable reasons for which individuals would seek care at the ED whether insured or not: having insurance through Medicare increases treatment intensity by around 3% and results in a 1% absolute (20% relative) reduction in 7-day mortality and a 3% relative reduction in 1-year mortality.

under the Emergency Medical Treatment and Active Labor Act (EMTALA) and was widely available pre-EMTALA. Wherry and Meyer (2015) examine the long-run impact of eligibility expansions for children using a regression discontinuity design and find lower mortality for nervous system diseases and cancer, rather than for accidents, among black but not white children. These studies, while pointing in different directions, suggest that there is important heterogeneity based on both cause of death and race.

[7] The overall mortality results are included in the 2004 NBER working paper but not later published papers (Card, Dobkin, and Maestas, 2008, 2009).

Doyle (2005) studies a subpopulation with strong need for emergency medical care (victims of auto accidents who are alive when they reach the hospital) and finds higher adult mortality rates for uninsured persons in Wisconsin during 1992-1997. He finds that being uninsured increases in-hospital mortality by 39%, relative to other auto accident victims (1.5 more deaths per 100, relative to a mean of 3.8 deaths per 100) (point estimate 0.015, 95% CI [0.003, 0.027]), which he attributes to differences in treatment intensity, rather than pre-accident differences in health; in this sense, the paper also speaks to a specific channel involving in-hospital treatment intensity for emergency care for severe traumatic injury.[8]

Levy and Meltzer (2004, 2008) review the literature and conclude that, consistent with Finkelstein and McKnight (2008) and Card, Dobkin, and Maestas (2004), the literature presents evidence at most of modest health benefits from general adult health insurance expansions. They note potential exceptions for specific vulnerable populations, but conclude that "for most of the population at risk of being uninsured (adults of ages 19 to 50), we have limited reliable evidence on how health insurance affects health." (Levy and Meltzer 2008, p.404).

In addition to the RAND Experiment, two other randomized experiments deserve attention. Weathers and Stegman (2012) find no significant mortality effect for adults receiving Social Security Disability Insurance when they receive health insurance immediately rather than after the usual 2-year waiting period, even when given assistance in navigating the health insurance system (point estimate for odds ratio 1.28, 95% CI [0.71,1.85]. However, their sample of 2,000 persons is small, and thus confidence bounds are wide. They do find that those receiving insurance have higher self-reported health. The second is the Oregon Experiment, involving Medicaid expansion for adults, administered through a lottery among those who applied. Finkelstein et al. (2012) and Baicker et al. (2013) find limited changes in mortality or measures of physical health after 2 years. They find increased healthcare use, increased diabetes detection and care (but not lower blood sugar levels), reduced financial strain, and less depression. Their first stage on health insurance coverage is strong at around a 25% relative rise in coverage for those in the treatment group; this difference shrinks rapidly, however, and is only half as large after 16 months. Their point estimate for mortality reduction is large, at -13%, but with a wide 95% CI [-26%, +13%]. Thus, both

[8] Another example of health insurance affecting health among a uniquely vulnerable population is Goldman et al. (2001), who use state HIV policies and Medicaid generosity as instruments for insurance status; they find that 6-month mortality falls by 71% as a result of gaining insurance.

experiments find statistically insignificant effects for relatively vulnerable populations (the disabled for Weathers and Stegman (2012), and poor adults who signed up for the Medicaid lottery and later enrolled if eligible for the Oregon Experiment).

In contrast, several recent papers on insurance expansions for nonelderly adults find large effects of health insurance on mortality rates. Sommers, Baicker and Epstein (2012) considers Medicaid expansion for non-elderly adults in three states (Arizona, Maine, and New York) that expanded Medicaid in the early 2000s compared to neighboring non-expansion states; Sommers, Long and Baicker (2014) and Powell (2018) consider the Massachusetts insurance expansion in 2006. McClellan (2017) considers the ACA mandate that requires employers to cover young adults under their parents' employment-based insurance policies until age 26, and Dunn and Shapiro (forthcoming) considers the effect of Medicare Part D prescription drug coverage for elderly adults.

B. Power analyses and prior use of simulated power in economics research

Our second contribution focuses on the value of conducting and reporting a power analysis in an observational study. We perform a power analysis in a form that is generally usable for DiD studies with reasonably long panels by using simulation, in which one imposes treatment effects of varying sizes on actual data during the pre-treatment period.

Ex ante power analyses, before research is carried out, are often used in randomized trial designs to assess feasibility and determine necessary sample size,[9] as well as in grant applications for observational studies.[10] However, even when performed at an early stage in a research project, power analyses are rarely reported in published research. It is rarer still to find simulated power analyses. The exceptions we found include Hsiang et al. (2015) and Croke et al. (2016) from economics and Hannon et al. (1993) from bird ornithology. Of these only Hannon et al. (1993)

[9] For example, after making assumptions about the mean and sampling distribution of a potential treatment effect, a researcher designing an RCT could use a standard formula to estimate the minimum number of subjects needed to detect an effect of that size at a 5% significance level 80% probability – termed 80% power. This ex ante power analysis is helpful in ruling out study designs that are underpowered given realistic assumptions, and can allow researchers to assess the needed sample size, and to enhance power by changing the research design.

[10] The National Institutes of Health (NIH) require reviewers of grant applications to evaluate how statistical power has been addressed and advice to potential grant applicants is to aim for at least 80% power (NIH, 2016; Gerin et al.,. 2017).

modify observed data to discern power, while Hsiang et al. (2015) and Croke et al. (2016) create synthetic data that is designed to proxy for real world variables of interest.[11]

Some have argued that power analysis should not be done after results are available (Hoenig and Heisey, 2001; Senn, 2002); citing concerns that a lack of power will be used to justify insignificant findings, which could be due to lack of a treatment effect. Conversely, Gelman and Carlin (2014) point out in low-powered studies which *find* a statistically significant effect, the estimated effect size will often have the wrong sign or have magnitude far larger than the true effect; this implies a need for power analysis in studies which find significant effects.[12]

A growing literature documents the prevalence of underpowered studies in a number of fields, including neuroscience, psychology, medicine, and economics (Button et al., 2013; Maxwell, 2004; Ioannidis, 2005; Ioannidis et al., 2017). Related early work in this vein by economists includes the lament by McCloskey and Ziliak that power analyses are rarely conducted (McCloskey, 1985; McCloskey and Ziliak, 1996; Ziliak and McCloskey, 2004). Ioannidis et al. (2017) estimate that the median statistical power in a large set of economics articles is 18%, which is far lower than the 80% standard used in experimental design. The authors determine power by relying on meta-analyses of these articles, and comparing the weighted effect size from the meta-analysis to a weighted standard error. Their approach, however, cannot be used to assess power in a single study.[13] In addition, Banerjee et al. (2015) review six randomized trials assessing microcredit and find that most suffer from low power due to a limited take up rate.

Single-study power analyses can be either closed form (based on an assumed data generating process) or simulation-based; the simulation can involve either artificial data (from an assumed data generating process) or actual data, to which a treatment effect is added. A study of bird nest visitation by Hannon et al. (1993), the earliest simulated power analysis we found, is similar in spirit to our own in that the authors apply a simulated treatment effect to actual data.

---

[11] Hannon et al. (1993) are also the only researchers who conduct a power analysis on their own results. Hsiang et al. (2015) and Croke et a. (2016) run power analysis on studies by others.

[12] Gelman (2018) and Button et al. (2013) note a technical concern: power analysis should not be based upon the estimated treatment effect size since noise in the estimated effect size will cause error in the estimated power; an estimated effect that exceeds the true effect would lead to estimated power that exceeds actual power.

[13] Zhang and Ortmann (2013) and Gallet and Doubouliagos (2017) use similar approaches to estimate power for a series of related studies. Zhang and Ortmann report median power of 25% in experimental papers using the dictator game. Gallet and Doubouliagos report that 59% of studies examining the impact of healthcare spending on life expectancy have adequate power.

The authors modify their outcome variable (nest visitation) using draws from the binomial distribution, gradually increasing (or decreasing) the probability of visitation. For each modified sample, they draw 50 bootstrapped samples, re-estimate their statistical model, and report power for each imposed effect size as the percentage of times the imposed effect is statistically significant among the bootstrapped samples.

In contrast, Hsiang et al. (2015) estimate power using synthetic data. They generate the dependent variable (likelihood of conflict) using a normal distribution with a fixed mean and standard deviation; they impose a treatment effect by varying the mean to indicate a "treatment effect." For each imposed effect size, they analyze the synthetic data using their preferred specification and report power as the percent of times a statistically significant result is found at the 95% confidence level. Croke et al. (2016) examine a meta-analysis done by Taylor-Robinson et al. (2015) on the impact of mass administration of deworming drugs on childhood health. Croke et al. (2016) demonstrate that the meta-analysis is under-powered by using a simulation similar to Hsiang et al. (2015).

An advantage of entirely synthetic data is that there will be no pre-treatment trends or treatment effect unless one is imposed. However, fully synthetic data involves large sacrifices, similar to those noted for closed form power analyses by Burlig et al. (2017); one must implicitly impose structure on the variance-covariance matrix, for which the true structure may not be known. For example, in a panel data setting, values could be autocorrelated across time, pre-treatment trends could be non-parallel in complex ways (as we find for our data), and unobserved covariates could predict both treatment and outcome. As Stigler (1977) points out, real data are rarely drawn from a "perfect distribution." Our approach, of applying a simulated treatment effect by modifying existing data during the pre-treatment period, does not guarantee a distribution centered around the null when we impose a zero treatment effect (the data can exhibit an "accidental" effect), but it preserves both the obvious and more subtle relationships present in the actual data that can affect power, and lets us take accidental effects into account in estimating power. We have yet to find a prior example of this exact approach other than Hannon et al. (1993). However, similar procedures are suggested in the online appendix of Burlig et al. (2017), § D.2 and by Gelman and Carlin (2014).

III. Conceptual Concerns

We study the end result (mortality) of a process that starts with policy changes to eligibility for free or subsidized health insurance. To assess the plausible magnitude of any treatment effect and the challenges in measuring that effect using available datasets, one must keep in mind the chain of causation between policy changes and health or mortality. Because large-scale datasets available to researchers do not adequately measure morbidity, many studies (including ours) focus on mortality. However, mortality records are generally not linkable at the individual level to other information, including pre-ACA insurance status (which one could use to exclude the always insured from the sample, thus increasing the first stage)[14] or income (which determines eligibility for Medicaid and subsidized private insurance).

Several concepts inform our analysis and the interpretation of our results. One is the existence of prior policies that provide vulnerable populations with health insurance, or with healthcare regardless of health insurance status. These include health insurance and healthcare for the elderly and disabled through Medicare or Medicaid; pregnant women through Medicaid; many children through the Children's Health Insurance Program; persons needing emergency care through the Emergency Medical Treatment and Active Labor Act (EMTALA); persons with specific high-cost health conditions (AIDS through the Ryan White Act and end-stage renal disease under Medicare); those who suffer workplace or automobile injuries; and those with access to public hospitals, publicly supported clinics, or the charity care provided by nonprofit hospitals. Thus, further health insurance expansions will affect principally populations and medical conditions outside these groups.

A second concept that informs our analysis is selection into coverage for a new program, such as the ACA Medicaid expansion, including selection effects for both take-up of new coverage and crowd-out of other coverage. The less policymakers are practically or politically able to target groups likely to be uninsured and promote a high take-up rate, the less likely it is that studies like ours will have sufficient statistical power to find detectable effects on health or mortality. For example, the ACA changes eligibility but does not directly provide insurance. As in any "intent-to-treat" (encouragement) experimental design, we can estimate a treatment effect only for the

[14] An analogy: The Oregon Experiment achieved a 25% first stage because insurance was offered only to persons who were previously uninsured and had applied for the Medicaid lottery.

"compliers" with the encouragement. Multiple selection effects are possible, including that those who sign up: (i) may be more health-conscious in other ways; (ii) may have greater healthcare needs (e.g., Kenney et al., 2012); (iii) may be more likely to use additional healthcare once insured; and (iv) may be more compliant with medical advice than the "never-takers" who do not sign up. Thus, estimates for compliers may differ from those for never takers or always takers (the already insured). For example, Kowalski (2018) reconciles differences in the effects of the Oregon experiment and the Massachusetts health insurance expansion on emergency department visits on the basis of better initial health for the Massachusetts complier populations.

Third, there could be substantial treatment heterogeneity even among the compliers, with health insurance improving health for some, but being neutral for others ("flat of the marginal benefit curve" medicine) or even detrimental due to overtreatment (e.g., opioid addiction as an unintended effect of pain treatment). Yet the available data limits our ability to study specific subpopulations.

A fourth concern is heterogeneous health insurance quality. In many states, Medicaid insurance is considered to be of lower quality than commercial insurance (Polsky et al., 2015).

Fifth, health insurance is only one factor potentially affecting trends in health and mortality. Other factors can vary by age and ethnic group (e.g., Case and Deaton, 2015, find rising mortality in middle-age for less-educated whites, but not other groups), and by state (as we find below). Differing trends complicate any effort to define a suitable control group.

These concerns, taken together, highlight the complex relationship between health insurance and health outcomes, and anticipate the limitations of the available data and policy shocks.

IV. **Data**

We measure mortality using the confidential version of the Compressed Mortality File (CMF), which contains records on approximately 2.6 million deaths a year.[15] This dataset is compiled by the National Center for Health Statistics (NCHS) and contains individual death records from the National Death Index, with county-level geographic identifiers.[16] Other data in the mortality files

[15] The public-use version of this data can be found at http://wonder.cdc.gov/mortSQL.html, but that version suppresses death counts in county-years with 10 or fewer deaths in any query.

[16] http://www.cdc.gov/nchs/data_access/cmf.htm#data_availability. We do not use data prior to 1999 because that is the first year in which death certificates began using ICD10 codes.

include (1) race, ethnicity, and gender; (2) year of death; (3) age at death (which we collapse into 5yr-age groups, e.g., 55-59, 60-64, etc., because county population, which we use as the denominator for measuring mortality rates, is available only for these groups); and 4) primary cause of death (4 digit ICD-10 code). We use data from 2009-2013 as the pre-treatment period and 2014-2016 as the treatment period for our main DD analysis, but use longer periods for selected analyses. We conduct county-level analyses, using county population (from the U.S. Census Bureau) as weights, to produce state-level and national estimates that are representatives of the respective populations. To examine the first-stage health insurance estimates that correspond to our mortality analyses, we use information on uninsurance rates from the Census Bureau's Small Area Health Insurance Estimates (SAHIE).[17]

V. ACA Insurance Expansions and Identifying Variation

In 2014, the two main insurance expansions under the ACA took place, with Medicaid expansions occurring in 27 states (including the District of Columbia) on or soon after January 1, 2014, in three more states on or soon after January 1, 2015, and in two more in late 2015 or the beginning of 2016. "Standard" expansion included coverage for all non-elderly adults with family income less than 138% of the federal poverty level (FPL). Of these 32 expansion states, 10 had conducted significant expansions prior to 2014 and are not included in our main specifications. The "treated" states for our principal DD analyses are the remaining 22 "Full Expansion States"; the control group consists of the 19 "Non-Expansion States"; we also treat the five late-expansion states as part of the control group during pre-expansion years. A number of other studies of Medicaid expansion also focus on the Full-Expansion States (e.g., Wherry and Miller, 2016). Table 1 lists the states in each expansion group, as well as the change in percent uninsured in each state from 2013-2015 for persons between the ages of 50 and 64; Appendix Table A-1 provides additional details on each state's expansion status.

The second major way in which the ACA expanded coverage was by creating "marketplaces" with private insurance subsidies for those with income between 138% and 400% of the FPL in expansion states, and 100-400% of the FPL in Non-Expansion States and WI (which expanded Medicaid only to 100% of the FPL). Our study design exploits mainly variation in

[17] Source: https://www.census.gov/data/datasets/time-series/demo/sahie/estimates-acs.html. SAHIE data is available for ages 50-64, rather than the 55-64 age group we study in our main analyses, but first-stage magnitudes should be similar.

Medicaid expansion, but we also provide estimates that use both sources of variation provided by the ACA by comparing areas that received different shocks to uninsurance rates due to differing pre-ACA characteristics.

There is ample evidence that the proportion of uninsured adults fell, and that the sources of payment for hospitalizations shifted toward more Medicaid and less self-pay. However, the uninsured population fell in both Expansion and Non-Expansion States. As Table 1 shows, the population-weighted drop in uninsurance rates from 2013 to 2015 for the 50-64 age group averaged 7.1% in Full-Expansion States versus 5.4% in Non-Expansion States; the difference between the two groups is 1.7%.[18]

This small difference in secular uninsurance declines between treatment and control groups poses a major challenge to any effort to use Medicaid expansion to estimate the effect of health insurance on mortality. The "first stage" of the encouragement design is only modestly higher for particular subgroups who were more likely to be affected by Medicaid expansion, for whom we still find first stages of 5% or less (Appendix Tables A-3 and A-4).

Although the ACA unambiguously reduced uninsurance rates, causal effects on healthcare delivery appear more modest and uneven across types of care (e.g., Mazurenko et al., 2018). The Oregon Experiment (Taubman et al., 2014) found a 40% increase in ED visits among the newly Medicaid eligible, and Ghosh et al. (2017) find that ACA Medicaid expansion predicts a nearly 20% increase in prescription drug use. In contrast, there is no evidence that the ACA Medicaid expansion led to a significant rise in ED visits in expansion states (Pines et al., 2016; Wherry and Miller, 2016). Both from this evidence and from prior studies of the effect of health insurance on mortality discussed above, we expect the effect of receiving health insurance on mortality during our study period to be modest.

[18] Here, we use uninsurance rates for persons aged 50-64 as the closest available match in the Small Area Health Insurance Estimates (SAHIE) data on uninsurance rates to our principal treatment group of those aged 55- 64. The drop in uninsurance rates was somewhat larger for the entire adult population. See Appendix. If one weighs states equally, rather than by population, the drop in uninsurance rates is 6.4% versus 4.4% (a difference of 2.0%). But the apparent gain in first-stage strength is offset by greater reliance on small states, for which mortality rates are noisier; moreover, this approach answers a different question: 'how is the average US state affected, rather than how is the average newly insured person affected?'

VI. Empirical Approach

A. Effect of Health Insurance on Mortality

To investigate the effect of Medicaid expansion on mortality, we use several DD specifications: (i) a "simple DD" specification, which assumes a one-time change in mortality rates; (ii) a "leads-and-lags" model, which allows for a separate treatment effect in each year, both before and after Medicaid expansion, and lets us assess whether pre-treatment trends are parallel; and (iii) a "triple difference" model, in which the third difference is persons aged 55-64 versus persons in the same county aged 65-74. Treatment is recorded in event time, relative to the year in which each expansion state expanded Medicaid. For states that expand on a date other than January 1 of year *t*, we treat year *t* as post-expansion if expansion occurred in the first half of the year; we treat year *t* as pre-expansion otherwise (see Table 1 for details). All models use county-level data, county and year fixed effects (FE), county population weights, standard errors clustered at the state level, and data from 2009 through 2016.[19] The simple DD model is:

$$Y_{jt} = \alpha + \beta Post_{st} + \partial X_{jt} + \tau_t + \vartheta_j + \varepsilon_{jt} \qquad [1]$$

Here, *i* indexes individuals; *j* indexes county; *s* indexes state; t indexes time in years, the dependent variable; Y_{jt} is *ln*((deaths)/100,000 persons)+1); we add 1 to the mortality rate to avoid dropping county-years with zero deaths.[20] We limit the sample to Full- and Non-Expansion States to form a stronger comparison. The predictor variable of interest is Post =1 for Full Expansion States in post-expansion years (2014 and 2015 for the 17 states that fully expanded Medicaid in 2014; 2015 for the 3 states that expanded in 2015). The covariate vector X_{jt} includes the following county-level demographic characteristics: % male; % Black; % White, % Hispanic; % aged 0-19, 20-34, 35-44, 45-54, 55-64, 65-74, 75-84, and 85+; managed care penetration (Medicare Advantage beneficiaries as % of all Medicare beneficiaries); % disabled (% of Medicare beneficiaries receiving SSDI benefits); % in poverty; unemployment rate; median household income; mean per-capita income; % with diabetes; % obese; % physically inactive; % smokers; active practicing non-

---

[19] A small number of small, rural counties experienced boundary changes over the study period, which are reflected at different times in different datasets. To handle this problem, we merged some counties (see the Appendix for details).

[20] We use a log-linear model for convenience, so that the regression coefficients are interpretable as (approximate) fractional changes in mortality. We obtain similar results with a linear model, with Y_{jt} = (deaths)/100,000 persons.

federal physicians/1,000 persons.[21] We convert all amounts to 2010 dollars.[22] In some specifications, we use a narrower set of covariates or no covariates, partly to assess whether our results are sensitive to including observable, time-varying, county-level factors, and also because expansion could affect some covariates. We include county and year fixed-effects ($\tau_t and\ \vartheta_{jt}$) in all models to control for potential unobserved covariates that vary across counties but are fixed over time, and for determinants of mortality that are constant across counties but vary over time.

Appendix Table A-2 provides a covariate balance table showing mean values for each covariate by state, averaged over the pre-reform period of 2009-2013. As expected, there are differences in a number of covariates. Expansion states differ from non-expansion states in a number of ways, including age structure (more weighted towards middle ages), race (more White), poverty (less poor), health status (less diabetes, more physical activity), health care access (more physicians per capita) and health insurance (less uninsured). To address covariate imbalance, we also implement an inverse propensity score weighting approach in which we compute ATT weights and use ATT*population weights.[23] Results with these weights, presented in the Appendix, are consistent with those we report in the text.

We principally study mortality due to healthcare-amenable causes (Nolte and McKee, 2003), but also provide some estimates for non-amenable and total mortality. The concept of amenable mortality seeks to capture deaths from conditions that are potentially preventable with timely care; examples include heart disease, stroke, cancer, diabetes, and infections.[24]

To study the time pattern of any apparent treatment effect, and to assess whether pre-treatment trends differ between Full- and Non-Expansion States, we use a leads-and-lags model in event time, with the first expansion year set to zero, following Equation (2):

---

[21] We take population data from the Census Bureau at http://www.census.gov/popest/. We use mid-year inter-censal estimates for 1999 and 2001-2009, and post-census estimates for 2011-2015. We obtain physician counts (interpolating from adjacent years for 2009 due to missing data), unemployment rate, median household income, percent in managed care (interpolating from adjacent years for 2006-2007 due to missing data), and percent disabled from the Area Health Resource File (AHRF) at http://arf.hrsa.gov/. County per-capita personal income comes from the Bureau of Economic Analysis at http://www.bea.gov/regional/. Data on health variables comes from the Centers for Disease Control at https://www.cdc.gov/dhdsp/maps/atlas/index.htm.

[22] Source: www.bls.gov/cpi/. We use the annual average consumer price index for all urban consumers.

[23] To generate propensity scores, we average the covariates over the pre-treatment period (2009-2013). We then run a logit regression, which predicts whether a county is in a Full- or Non-Expansion State, using all variables in Table A-2 to generate the fitted propensity p. ATT weights are calculated as (p/(1-p)).

[24] We implement the concept of amenable mortality using the ICD-10CM causes of death tabulated in Sommers, Long, and Baicker (2014), App. 1, last column. This definition is somewhat broader than the Nolte and McKee definition.

$$Y_{jt} = \alpha + \sum_{k=-5}^{2}(\beta_k * D_{jt}^{k}) + \partial X_{jt} + \tau_t + \vartheta_j + \varepsilon_{jt} \quad [2]$$

Here, k indexes "event time" in years relative to Medicaid expansion. $D_{jt}^{k} = 0$ for Non-Expansion States for all t and k. For Full-Expansion States, $D_{st}^{k} = 1$ for the k^{th} year relative to the adoption year, and 0 otherwise. For states that expanded Medicaid on January 1, 2014, $D_{st}^{1} = 1$ for 2014 and 2 for 2015. Thus, β_1 provides the estimated population average treatment effect for the first expansion year, while β_{-1} is the estimated effect one year before adoption, and so on. We adjust the coefficients by subtracting β_{-3} from each, so that reported $\beta_{-3} \equiv 0$.

We find evidence that states have differing mortality trends during the pre-treatment period, which casts doubt on the parallel trends assumption required for valid DD analysis. To address these sources of differing trends, we use a further source of within-state variation: mortality trends among those who are 65 or older (and thus always insured) can potentially control for the otherwise unobserved state-specific factors that generate non-parallel trends. We thus also use a triple-difference/age-discontinuity specification (similar to Finkelstein and McKnight, 2008), where the third difference is mortality among persons between the ages of 65 and 74, who are eligible for Medicare and should not be affected by Medicaid expansion, and limit the sample to persons between the ages of 55 and 74, thus comparing mortality trends for the 55-64 age group to thosein the 65-74 age group. The triple-difference specification, analogous to simple DiD, is:

$$Y_{jt} = \alpha + \beta Post_{st} * Under65_{st} + \beta Post_{st} + \beta Under65_{st} + \partial X_{jt} + \tau_t + \vartheta_j + \varepsilon_{jt} \quad [4]$$

Heterogeneity/Robustness

We also seek to strengthen the first stage (the fraction of county population that gains insurance due to Medicaid expansion) and to investigate potential heterogeneous treatment effects, by estimating a model that interacts the double difference with an indicator for counties with high uninsurance rates in 2013, prior to Medicaid expansion. High2013 indicator equals 1 for the counties with the highest uninsurance rates in 2013, such that together they contain 20% of the population of our treated and control states (or demographic subsamples), and 0 for the counties with the lowest uninsurance rates in 2013, containing another 20% of this population; we remove from the sample counties with moderate uninsurance rates (containing 60% of the U.S. population). We thus compare high-uninsurance counties to low-uninsurance counties. The regression equation is:

$$Y_{jt} = \alpha + \beta Post_{st} x High2013_j + \beta Post_{st} + \partial Xjt + \tau_t + \vartheta_j + \varepsilon_{jt} \quad [5]$$

We similarly compare counties with high poverty rates in 2013, containing 20% of the sample population, to counties with low poverty rates, also containing 20% of this population. This approach exploits variation from the ACA overall, rather than just the Medicaid expansion component.

We also estimate separate models for subsamples stratified on covariates that may predict uninsurance rates or response to health insurance, for which we also have mortality data: education, gender, and race/ethnicity. For example, lower-educated subgroups will have larger first stages and higher mortality rates, and thus will (before the offsetting effect of reduced sample size) could be more likely to produce detectable mortality changes.

B. Power Analysis

The power of a statistical test is the probability that the test will correctly reject a false null hypothesis, at a given confidence level. For a regression coefficient, power is normally taken to be the likelihood that the coefficient will be found to be significantly different from zero, at that confidence level. We conduct a simulation-based power analysis by artificially introducing treatment effects of different sizes into the data in the pre-treatment period, and then assessing (over 1,000 iterations) how often our DD and triple-difference regression models can detect these effects at the 90%, 95%, 99%, and 99.9% confidence levels (using two-tailed tests). The goal of this analysis is to determine the minimum effect of health insurance on amenable mortality that is reliably detectable with our data and research design.

The alternative of a closed-form power analysis requires fully modeling the data generating process, including parameterizing the error term for both variance and covariance terms, and is especially hard to construct with panel data in which observations can be correlated over time (Burlig et al., 2017). A simulation using entirely artificial data has similar problems. We therefore use simulation methods applied to real data. For example, our simulation approach builds in "noise" from non-parallel trends in the actual data; with a closed-form analysis we would have to model the level and form of these trends. Our use of regression weights and clustered standard errors further contributes to the difficulty in producing a tractable and credible form for an analytic power calculation. Simulation, applied to real data, avoids these challenges and lets us use the same research design and econometric specification as the main analysis (Burlig et al., 2017).

Our simulation proceeds as follows. We exclude all data from the post-treatment period and use data from 2007-2013 rather than the 2009-2015 period used in our actual analyses. We then do the following 1,000 times: we randomly assign a pseudo-expansion status to 20 of the 41 states in our final study (that either fully expanded or did not expand Medicaid). Thus, in each draw, 20 states are pseudo-treated and 21 are pseudo-control. In each case, we assume that the expansion occurred in 2012, giving us two years of post-expansion data for each pseudo-treated state.

For each randomly drawn set of pseudo-treated states, we impose a pseudo treatment effect of a reduction in amenable mortality (from 0% to 6%, in 0.25% increments) for all persons aged 55-64 living in a pseudo-treated state. We do this by randomly removing deaths from each pseudo-treated county-year using draws from a binomial distribution. For example, if a county-year has 100 healthcare-amenable deaths and the imposed treatment effect is 0.5%, we remove each death with probability 0.005. The expected number of remaining deaths is then 99.5, but the actual number must be a whole number and could be 100, 99, 98, etc. Each imposed treatment effect is randomly distributed across the pseudo-treated states and across counties in each state. Thus, as in this example, it is unlikely that any pseudo-treated county will have its mortality rate decrease by exactly 0.5%, but the pseudo-treated counties will still experience the imposed treatment effect on average (subject to sample variation).

Once we have introduced the artificial shocks, we run the DD model in eqn. (1) and save the regression coefficient and standard error. The percentage of times a result is found to be statistically significant for a given effect size and significance level is the power for that effect size and significance level; a common threshold for a study to be deemed adequately powered is 80% power at a 95% confidence level. We similarly assess power using the DDD model in eqn. (4). In addition to statistical power, we also report three measures based upon Gelman and Carlin (2014) that inform the plausibility of any significant results obtained, given the study's underlying power: the percentage of times a significant, estimated treatment effect has the wrong sign (opposite from the imposed effect; that is, a *higher* mortality in expansion states); in the subset of cases where a significant effect is found, the mean ratio of the estimated treatment effect to the true (imposed) effect (the exaggeration ratio); and the percentage of significant treatment effect estimates that have the correct sign and an exaggeration ratio below 2 (which we term a "believable" coefficient).

VII. Principal Results

We present full-sample results in this section, principally for adults aged 55-64 some limited results for adults in a broader 45-64 age group. We first present univariate results, and then results from DD and triple difference models. See the Appendix for similar results for all non-elderly adults. In the Appendix, we assess whether we could obtain a better match between treated and control states, and thus tighter confidence bounds, using synthetic control methods. We conclude that we cannot rely on these methods for inference due to poor pre-treatment fit.[25]

A. Univariate Graphical Evidence

In Figure 1, we display trends in amenable mortality for the four state groups, for the full time period with available data (1999-2015). We aggregate data to the state-group level using population weights, and show amenable mortality rates per 100,000 persons aged 55-64; Appendix Figure A-1 shows data for persons aged 18-64. Several features of Figure 1 are important. First, there are substantial differences in mortality rates across the state groups, although these are smaller between our principal comparison groups—the Full-Expansion vs. Non-Expansion States.

Second, Figure 1 shows clear evidence of non-parallel pre-treatment trends. Unless these differences are absorbed by the regression covariates (for our data, we show below they are not) or by our third difference (they partly are), any DD analysis is suspect. More specifically, over 2010-2016, mortality continues to decline in the Mild-Expansion and Substantial-Expansion states, but levels off in the Full-Expansion States and rises in the Non-Expansion States. We also find non-parallel univariate trends for all non-elderly adults (Appendix Figure A-1).

If one simply compares the post-treatment average difference in mortality rates for Non-Expansion versus Full-Expansion States to a similar post-treatment average difference—as a simple DD regression does—it would appear that Medicaid expansion has a large, immediate effect in reducing mortality. In fact, mortality rates for these two state groups diverge principally during the pre-treatment period. There is little additional divergence during 2014-2016. The simple DD coefficient is misleading, because it ignores the non-parallel pre-treatment trends. One value of the power analysis presented below is to protect against finding spurious significance due to non-parallel trends. The power simulation during the pre-treatment period treats pre-treatment trends as a source of additional noise, which reduces power.[26]

---

[25] See Appendix Figures A-2 and A-3.

[26] A common robustness check, which provides some protection against DD results being driven by non-parallel, pre-treatment trends, is to add linear unit-specific trends to a DD regression. This can be effective in some cases, but requires a long pre-treatment period to estimate the linear trends and assumes a simple parametric (linear) form for those "trends."

[FIGURE 1 about here]

B. Covariate Balance

Appendix Table A-2 provides a covariate balance table showing means, and the normalized difference in means, between Full- and Non-Expansion states for the pre-expansion period of 2009-2013. There are meaningful differences between the two state groups on a number of covariates, as well as on mortality (see Figure 1) and uninsurance rates. In light of these differences, we reran the analyses reported below with ATT*population weights instead of population weights. Results are similar to those we present; see the Appendix. We use the simpler, population-weighted results as our main specification, as they are more transparent.

C. Leads-and-Lags Results

We turn next to leads-and-lags graphs, using equation (3). Figure 2, Panel A, provides annual point estimates and 95% CIs over 2004-2015, for amenable mortality among persons aged 55-64. There is, as expected, strong evidence for non-parallel pre-treatment trends, with relative mortality improving in Full-Expansion States over 2007-2013. There is also no evidence of a change in relative mortality in the first two expansion years. In Appendix Figures A-3 and A-4, we provide leads-and-lags graphs for total mortality and non-amenable mortality, these also show no evidence of a significant treatment effect.

[FIGURE 2 around here]

The likelihood of finding credible evidence of causal effects weakens further when we compare the coefficient magnitudes in Figure 1 to plausible effect magnitudes for the full populations of the treated states, given the small first stage shown in Table 1. Based on the prior research discussed in Part II, even a 10% effect of health insurance on mortality within two years would be large. Yet a 10% reduction in mortality for the treated (newly insured), with a roughly 1% first-stage (percent of the population treated), implies an average mortality reduction for all persons aged 55-64, and thus a DD coefficient of 0.001 (0.1%). It is apparent from Figure 1 that this reduction would be undetectable; it would be far lower than the annual 95% CIs, and far lower than year-to-year relative changes in mortality in the pre-treatment period, which can be up to 20 times as large (0.02 from year -2 to year -1).

If we take 0.02 as the minimum detectable effect with one year of data and 0.001 as a large but perhaps plausible effect size coefficient, Figure 2 suggests that our study is underpowered by a factor of 20 (equivalently, the ages 55-64 population needs to be 400 times larger). Adding one or two more years of data (which should be possible in the near term) would help, but would not be adequate to overcome this issue. We present a formal power assessment below, which is consistent with this qualitative discussion.

In Figure 2, Panel B, we present a similar figure for amenable death rates for those aged 65-74 to provide background for our triple-difference regression estimates. There is again evidence of non-parallel trends, with mortality dropping in Full- versus Non-Expansion states in the pre-treatment period. This suggests that the third difference (where we use 65-74 year olds as a within-state control) can limit the non-parallel trends we saw in Figures 1 and 2A.

Figure 2, Panel C provides triple-difference leads-and-lags results: annual point estimates and CIs are for Full- versus Non-Expansion States and for the 55-64 versus 65-74 age groups. Non-parallel trends are muted, but standard errors are larger than in Panel A. Moreover, there are still large year-to-year swings in relative mortality in the pre-treatment period, with a jump of around 0.02 from 2006 to 2007, and a similar jump from 2009-2010. Figure 2C shows dips in relative mortality in Full-Expansion States in 2014 and 2015, but the magnitude is both much larger than the plausible causal effect of around 0.001 and too small to be statistically convincing, given the year-to-year variation we observe in the pre-treatment period.

In Panel D, we present results from an age discontinuity specification that compares persons aged 55-64 to those aged 65-74 in the same state. This specification exploits both sources of ACA insurance expansion, leading to a stronger first stage, and can be applied in both Full- and No-Expansion states. We find, however, strongly non-parallel pre-treatment trends (rising relative mortality for those aged 55-64, compare Case and Deaton, 2015). These trends are similar in both Full- and No-Expansion States (Appendix Figure A-4). There is no evidence of a post-ACA change in this long-term trend.

Our overall assessment is that the triple difference specification in Panel C is the best available in limiting the extent of non-parallel pre-treatment trends. It remains suspect, however, because it depends on non-parallel trends in the three relevant double differences tending to offset each other in the pre-treatment period, with no basis for confidence that they would continue to do so in the treatment period.

D. DD and Triple-Difference Regression Results

We next turn to regression analysis. Table 2 shows results from DD regressions, following eqn. [1], with county and year FE and county population weights, separately for our principal treatment group (ages 55-64) and the placebo group (ages 65-74). It also shows triple-difference results, following eqn. [4]. While both DD and triple-difference specifications are suspect because of parallel trends problems, non-parallel pre-treatment trends are less severe for the triple difference; thus our discussion focuses on those results. We show separate results for healthcare-amenable mortality, non-amenable mortality, and total mortality. Even-numbered columns include the covariates noted above. We present results for the 55-64 age group both because we expect the effects of health insurance to be higher for this group than for younger persons, and because we need to study a limited age band to pursue the triple-difference approach. In the Appendix we estimate DD models that include younger ages for the treated population, with similar results. We caution that these regressions assume flat pre-treatment trends, but we in fact observe a declining trend. Given this trend, DD results will be biased toward finding a post-expansion drop in mortality.

In Table 2, in regressions with covariates, we find a statistically significant 2.1% post-expansion fall in amenable mortality for those aged 55-64, with no significant change in non-amenable mortality. However, in addition to assuming parallel trends, these results are fragile. First, the coefficient on the Full-Expansion dummy is far too large to be credible. Given our roughly 1.2% first stage, it implies an impossible 175% (2.1%/1.2%) reduction in amenable mortality among those who gain health insurance. Second, for the placebo group (ages 65-74) and the placebo-outcome (non-amenable mortality), we observe a large, statistically significant *rise* in mortality. Third, the triple-difference decline in mortality is far smaller, at 0.7% (although still implausibly large) and is not close to statistical significance. Note too that the standard errors for amenable mortality are around 0.007 with covariates and rise to 0.009 in the triple-difference specification. This implies a minimum detectable effect of around 0.014 to 0.018, which implies a 120-150% drop in amenable mortality for compliers. This is further evidence the research design is severely underpowered.[27]

[27] If we expand the age range for the treated group to 45-64 instead of 55-64, the insignificant negative triple-difference point estimate in Table 2 switches sign; see Table 4. Moreover, by broadening the age range, we weaken

[TABLE 2 around here]

E. Evidence on Heterogeneous Effects

We conducted extensive additional analyses of the effects of ACA-induced insurance variation on mortality, focusing on vulnerable subgroups or particular causes of death. These subgroups can potentially provide a stronger first stage, a stronger second stage, or both. However, moving to subgroup analysis also reduces sample size. We consider subgroups based on gender, race/ethnicity, education level, specific cause of death, and county poverty and pre-ACA uninsurance rates. We present and discuss these results in the Appendix.

Our search for evidence of a significant effect of Medicaid expansion on mortality for particular subgroups also proves to be underpowered. The discouraging conclusions we drew for the general adult population—no evidence of a statistically significant effect, and far too little power to detect effects of plausible magnitude—do not change. Reduced power due to a smaller sample outweigh any gains from a larger first stage or a higher base mortality rate.

Indeed, given the problems we found for the full sample, with both non-parallel trends and low power, these analyses have the flavor of beating (or, perhaps, seeking to revive) a dead horse. We find no success here. Most regression coefficients are insignificant. When significance is found (for Non-Hispanic Blacks and for Hispanics, see Appendix Table A-3), there are other factors that cut against a causal interpretation, including non-parallel pre-treatment trends and coefficients of implausible magnitudes given the weak first stages.

VIII. Power Analysis

We return to our conceptual framework of the chain of events by which insurance expansions may affect mortality, and discuss the conditions under which studies of the ACA using death certificate data could establish a connection between health insurance and mortality.

A. An Illustrative Example

Suppose first that out of 100,000 individuals aged 55-64, half became newly insured. By how much would the likelihood of death within 2 years have to change for us to find that change to be statistically significant? The annual amenable mortality rate in this group is around 600 per

---

the logic behind using mortality for persons aged 65-74 as a third difference, yet we need that third difference to address non-parallel pre-treatment trends.

100,000 per year (Appendix Table App-2), if insurance were to reduce the probability of death by 25% among the newly insured, then insuring 50,000 individuals among 100,000 individuals would reduce the expected number of annual deaths by 75 (0.5*0.25*600) to 525. In expectation, a DD regression should show a 25% reduction in mortality rate.[28] But there will also be random variation in mortality. If mortality events are independent, the expected standard deviation (σ) of mortality/100,000 persons will be around 24,[29] and the expected *t*-statistic will be 3.07.

Now assume that there is random "external" variation in state-level mortality rates, with a standard deviation of around 2% per year (±12 deaths per year. As we show below, this is a reasonable level for our data. If this source of variance is independent of that due to health insurance, expected total variance will be 596 (from random mortality events) + 144 (from external variation) = 740, expected standard deviation will be around 28 and the expected t-statistic will be 2.76 – lower but not dramatically so.[30] The large effect of health insurance swamps the additional "noise" from other sources of variation in mortality.

Now assume that the background noise remains the same, but only 5% of the population is treated, and the mortality reduction for the newly insured is 10% instead of 25%. The expected population average treatment effect is now a reduction in the mortality rate of 3 (0.05*0.1*600) to 597. The standard deviation in the number of expected deaths remains the same, so the expected *t*-statistic will be only 3/28 = 0.11. To bring this *t*-statistic up by a factor of, say, 20 to 2.2, one might initially imagine we would need a sample 400 times as large – 40 million people.

However, as sample size increases, the variance in mortality rate due to independent mortality events falls by the usual factor of $n^{1/2}$. With a hypothetical sample of 40 million, the variance in the mortality rate (per 100,000 persons) would be 594/20 ≈ 30. But the variance due to external state-level mortality shocks will not fall and will dominate expected total variance, which will be 30 + 144 = 174; implying expected (σ = 13.2; *t* = 0.23).

This, in a nutshell, is the power problem we face. With a weak first stage, and a moderate second stage, even a very large sample cannot overcome the confounding effect of external

---

[28] The expected coefficient in a regression, such as those we run, with *ln*(mortality rate +1) as the dependent variable should be around -0.22

[29] This uses the standard formula for the variance of a binomial distribution with probability Var = n*p(1-p). For n 100,000 and p = .006, Var = 596 and σ = $Var^{0.5}$ = 24.42.

[30] Variances due to independent sources add so Var_{tot} = 596 + 124 = 740, and σ_{tot} = $Var_{tot}^{0.5}$ = 27.56.

variation in mortality rates. If that external variation is independent across states, then having more treated and control states will help but only somewhat. For example, if we had 20 treated states and 20 control states, all of equal size, the combined external variance for both groups would be (144/20) + (144/20) = 14.4; expected total variance would be around 44, implying expected (σ = 6.64, t = 0.45). If the treatment effect of health insurance on mortality were felt immediately then more years of data would help, but only somewhat, given that state-level mortality shocks are likely to persist over time. For example, 3 years of data, variance due to random arrival of deaths would fall to $29.7/(3^{1/2}) = 17.1$, but if state shocks are persistent, total expected variance will be 17.1 + 14.4 = 31.5; implying expected (σ = 5.62; t = 0.53). Having a first stage lower than 5% -- as we do -- will only exacerbate matters.

Thus, this example illustrates that a full-sample effect size on the order of a 0.5% reduction in mortality (hence an expected regression coefficient around - 0.005 in the log-linear specification we use) will not be detectable. Our power analysis formalizes this intuition, and shows that for plausible effect sizes, the effect of ACA Medicaid expansion on mortality is too small to be captured using death certificate data, unless that data can be linked to income data and insurance data, thus permitting a much larger first stage. We also show below that given lower power, one should be cautious in interpreting any statistically significant results from studies such as ours, even if parallel trends assumptions appear satisfied.

B. Available First-Stages

An initial question for our power analysis is what first stage one could realistically achieve with better data. Our full-sample first stage is similar to that in other ACA Medicaid expansion studies.[31] From SAHIE data, the first stage for low-income, Medicaid-eligible adults (income < 138% of FPL, age 50-64) is around 5.3%. We also saw above that the first stage for low-educated adults is around 4%.[32] Thus, around 5% is likely as large a first stage as one can achieve without linked individual data on some combination of income, family status (children at home), pre-expansion insurance, and mortality.[33] ACA-derived insurance gains were somewhat smaller

[31] Long et al (2014), using data from 2013-2014, find a 5.8% drop in uninsurance in expansion states vs 4.8% in non-expansion states, between 2013 and 2014, implying a 1.0% first stage. Smith and Medalia (2015) find a 3.4% reduction in uninsurance for all persons aged 0-64 in expansion states vs 2.3% in non-expansion states, hence a 1.1% first stage.

[32] Kaestner et al. (2015) estimate a similar 3% first-stage for low-educated adults, age 19-64.

[33] Wherry and Miller (2016), use income data from the National Health Interview Survey to isolate persons with incomes < 138% of FPL and find a 7% relative increase in insurance rates from 2010 to 2H2014 low-income persons

among the near elderly (on whom we focus) than among younger adults, perhaps because the near-elderly have greater healthcare needs and greater income, which led many to obtain insurance pre-ACA.[34]

We present power calculations below for the aged 55-64 population (around 29M persons, 14M in treated states), and also for our triple-difference specification. The first stage for the closest population for which we have data, persons aged 50-64, is around 1.1% (see Appendix Table A2). A 10% reduction in mortality for the newly insured, as large a near-term effect as we consider plausible, thus corresponds to a 0.012% reduction in mortality for all persons in this age group. The upper end of the 95% CIs from Finkelstein and McKnight (2008) and Card, Dobkin and Maestas (2004) imply an even lower mortality decline, bounded at 0.004%.

To put these numbers in context, Medicaid expansion led to around 170,000 more people gaining health insurance in Full-Expansion States (0.0012 * 14.1M) relative to non expansion states. If the mortality of the newly insured would have been similar to all persons in this age range but for Medicaid expansion, about 0.6% would have died each year (about 1,000 persons), and a 10% reduction in mortality would save around 100 lives annually. We cannot directly measure the relative mortality of the uninsured with our mortality data, but Black et al. (2017) provide evidence from the Health and Retirement Study that mortality for uninsured persons in the HRS population (initial age 50-61, so similar to the group we study) was similar to mortality for insured persons.[35]

The power challenge is to find statistically significant evidence for a fall in mortality of 100 persons (or less), in a combined treated and control population of around 29M, with 170,000 annual deaths. As we show below, that challenge cannot be met without individual level data on

---

aged 19-64; compare the 5% increase from 2013 to 2014 we find using SAHIE data. Simon et al. (2017) combine income data with childless status and find a 10% increase for childless adults age 19-64, with incomes < 100% of FPL and no children at home in 2014-2015, relative to a 2010-2013 baseline.

[34] Appendix Figure A-25, reproduced from the American Community Survey (ACS), shows the ACA-related change in uninsurance rates by age.

[35] Black et al. (2017), Table 2 calculates mortality differences in the manner most appropriate for these comparisons; the uninsured (aged 50-61) have higher mortality than the privately insured, but lower mortality than the publicly insured, leading to similar overall mortality between insured and uninsured over two- and four-year observation periods. To put these estimates in the context of prior literature, Galea et al 2011 reports that mortality for poor non-elderly adults is 75% higher than for the non-poor but does not report mortality differences for poor uninsured vs poor insured, which is the relevant comparison for our study. Kronick (2009) finds a 1.20 mortality hazard ratio for the uninsured versus the privately insured over a 14-year followup period after controlling for income (but does not compare the uninsured to the publicly insured).

personal characteristics (income, family status, pre-ACA insurance and health status), sufficient to greatly increase the first stage, linked to mortality data. Even with that data (not currently available), one would need a very large sample of newly insured persons and similar controls.

C. Full Sample Power Simulation Results

To investigate the minimum effect that our main DD and triple-difference specifications can detect, we perform the power exercise outlined in Section VI B. Figure 3 illustrates the results from our power simulation, using the amenable mortality rate for all persons aged 55 to 64 as the dependent variable. The simulation uses data from 2007-2013, and a pseudo-shock applied on January 1, 2012, to states chosen at random from our actual treated and control states.

Panel A shows DD results and Panel B shows triple-difference results, using the same regression models as in Table 2. The DD results indicate that to achieve 80% statistical power (finding a significant effect at least 80% of the time), the minimum detectable population average treatment effect size at the 95% confidence level is a mortality reduction of 1.8% for the DD and for the triple-difference simulation. Below, we focus on the triple-difference results, which we prefer because they are less subject to concern with non-parallel trends. A 1.8% fall in overall amenable mortality, given the roughly 1.1% first stage, implies that Medicaid expansion would have to reduce the average amenable mortality rate of all newly insured persons by (.018)/(.011) = 163%. If we apply a stricter significance standard, to account for specification error, specification searches, and file-drawer bias, the minimum detectable effect will be substantially higher – Figure 3 also shows power curves for the 99% and 99.9% and confidence levels.

The minimum detectable effect can also be framed in terms of lives saved. The 1.80% reduction in mortality needed for 80% power and 95% confidence translates into about .0180 * 14.1M * .006 = 170,000 = 1,522 annual deaths – almost 20 times the maximum plausible effect.

[Figure 3 about here]

The power analysis assumes that the underlying mortality rate of the newly Medicaid insured is similar to other persons aged 55-64. The actual rate could be higher (the newly insured tend to be low income, and thus higher mortality), or lower (the disabled are already insured, those in poor health could be more likely to already have insurance, and the first stage is lower for men, who have higher mortality rates than women), but is unlikely to be radically different. By comparison, Finkelstein et al. (2012, Table IX) study a likely lower-income, less-healthy population (persons who applied for the Oregon Medicaid expansion lottery), and report annual total mortality for the controls of 0.008, which is similar to the average total mortality rate we find for persons aged 55-64 in both Full-Expansion and Non-Expansion States. Power is also similar if we weight states equally, rather than by population; this increases the first stage to around 2%, but increases noise by giving more weight to smaller states.

"Power" also has peculiar properties, in the situation we face, where plausible effect sizes are small relative to those one can reliably detect. This implies both that: (i) the estimated effect is likely to greatly exceed the true effect; and (ii) there is an important risk that the estimated effect has the wrong sign (opposite from truth). Gelman and Carlin (2014) therefore recommend reporting two measures of plausibility in addition to power, the wrong-sign-likelihood and the exaggeration-ratio. Ioannides et al. (2017) report evidence that much economics research and thus prone to these concerns. We illustrate these problems in Figure 8.

In Figure 4, Panel A, we show the ratio of the magnitude of the estimated effect (when found to be statistically significant) to the "true" magnitude, imposed in the simulation. For population effect sizes under 1% (recall that a 10% mortality reduction for the newly insured implies a population effect around 0.1%) the exaggeration ratio is high – an effect which is large enough to be statistically significant is likely to be far from truth. In Panel B we show the proportion of statistically significant results that have the wrong sign. This proportion is also appreciable for the smaller population effect sizes. As we increase the imposed population effect size, the wrong-sign problem shrinks, and is negligible for effect sizes s above 1%; the exaggeration ratio also shrinks, but more slowly.

[Figure 4 around here]

As we discussed in Section A, one important source of "noise," captured in the power simulations but assumed away in DD regressions, is non-parallel mortality trends across states. We illustrate that concern in Figure 5. For this figure, we use a DD model, continue to use data from 2007-2013, apply a pseudo-shock to amenable mortality on January 1, 2012, but this time to one state at a time, treating all others as controls. We show a scatter plot of the DD estimates for each state of the change in amenable mortality, from regressions otherwise similar to those used for Table 2, versus *ln*(state population in 2012). We also superimpose a regression line showing the best linear fit between the point estimates and *ln*(population).

It is apparent from Figure 5 that for single states, it is common to find pseudo-treatment effects of 2% or more, with a fair number of states showing pseudo-effects of 4% or more, and Montana and Mississippi showing pseudo-effects around 6%. There is also a tendency for larger states to have better mortality trends than smaller states over 2012-2013, shown by the negative slope of the best-fit line.

[Figure 5 around here]

D. Power for Vulnerable Subgroups

We also conducted power analyses for the demographic, education, and cause of death subsamples discussed above, and report results in the Appendix. Power is generally similar to, or lower than, that shown in Figure 3. Smaller sample size, which reduces power, offsets the effect of the modestly larger first stages, which are all we can achieve. And the effect of non-parallel trends, in reducing power, remains.

E. What Data Would Be Needed for Reasonable Power?

We turn in this section to a different question – what combination of a stronger first stage and a reduction in amenable mortality for the newly insured would be detectable with reasonable power, if we could use a richer dataset, with data on mortality linked to data on income and family status (to determine eligibility for expanded Medicaid coverage) and pre-ACA insurance status (to exclude the always-insured from the sample). This hypothetical data would improve the first stage and bring it toward (or even above) the 5% one could obtain by studying only adults with incomes < 138% of FPL, or the 10% in Simon et al. (2017) for childless adults with incomes < 100% of FPLs. We consider the triple-difference design, which has similar power to DD and does a better, although imperfect job, of addressing non-parallel trends.

In this scenario, we imagine that we can identify in each county both a treated subsample and a similar control subsample, both aged 55-64. For example, if the treated subsample is childless adults with income < 138% of FPL, the within county control subsample could be childless adults with incomes from 138% to 250% of FPL. We assume hypothetical first stages varying from 1% to 15% and hypothetical second stages varying from 0% to 10%. For, say, a 5% first stage and a 10% second stage, we assign "insurance due to Medicaid expansion" to 5% of the persons in a "5% first stage" subsample of each expansion county, and then remove 10% of the amenable mortality deaths from the treated persons in this subsample (thus applying an overall mortality reduction to the subsample of .005). We again use data from 2007-2013 and a pseudo-treatment at Jan. 1, 2012, and assess whether we could detect this mortality effect if we did not know which specific individuals within this subsample would have gained insurance due to this pseudo-treatment. Since the treated and control samples are drawn at random from the same county and age range, they have the same expected mortality rates, by construction.[36]

We assume that with the hypothetical data, (i) researchers can identify the subsample members, and (ii) *all* effects of Medicaid expansion on uninsurance rates are concentrated in the subsample we consider. Thus, in our 5% first stage/10% second stage example, we assume that the entire Medicaid-expansion-related relative drop in uninsurance –170,000 persons in Full-Expansion States -- comes from this subsample. This defines the subsample size at 170,000/.05 = 3.4M treated persons, and a similar number of controls.

In Figure 10, we show power curves only for the 95% significance level. We vary (i) the assumed first stage (we show curves for 1%, 3%, 5%, 10%, 15%, and 20% first stages) and (ii) the imposed mortality reduction for the newly insured (from 0% to 10%) for the 5% significance level. With this hypothetical richer data, we need a smaller number of avoided deaths to be able to reliably detect a treatment effect. For example, with a 10% first stage, we could reliably detect mortality reductions of 2.4% or more in this subsample, or around 1,563 annual deaths. This is only slightly below the number of deaths we could detect in the full sample; thus, this hypothetical study remains severely underpowered. Recall that with a 10% second stage, we expect around 100 fewer annual deaths among those who actually gain insurance.

F. Implications of Power Analysis for Other Studies

While our exact simulation approach for understanding the minimum detectable effect is specific to our dataset and research design, a similar approach can be used in many other studies. We offer here four examples of why we believe power analyses such as ours, including an assessment of the minimum detectable effect and whether that effect size is plausible, can be broadly valuable in quasi-experimental research.

[36] For small subsamples, there are many county-years with zero deaths in smaller counties. The log transform we use (y_{it} = *ln*((deaths)/100,000 persons)+1) can produce substantial bias when there are many zero-death observations but most non-zero death rates are large (because we multiply the fractional rate by 100,000), which can lead to underestimating statistical power. We therefore use a linear model in conducting power analysis for specifications that examine small sub-groups, and thus have many county-level observations with zero deaths.

First, our power analysis can be usefully compared to the results in Finkelstein et al. (2012), who study the Oregon Health Insurance Experiment. With a sample of 75,000 people and a roughly 25% first stage among people who signed up for the Oregon Medicaid lottery, who were randomly offered Medicaid or assigned to control, the study reports a large point estimate for the near-term effect of receiving Medicaid on mortality of around 13%, but a *t*-statistic only around 0.5. This implies that the sample was undersized, even for that large point estimate, by a factor of around 16 –a sample of 1.2M people (with 300,000 newly insured) would be needed to reliably find a 13% effect on mortality – and a sample of 8M people (with 2M newly insured) to find a 5% effect.[37] Yet, from SAHIE data, the number of people in Full-Expansion states aged 50-64, with income < 138% of FPL is around 3.4M, and the first-stage for this group is around 5.3%, hence around 180,000 newly insured due to Medicaid expansion relative to non expansion states. Thus, even if we could link mortality and income data at the individual level, and focus on the income range eligible for Medicaid (< 138% of FPL), power to detect mortality effects would be low.

Second, our analysis of power to detect the effect of health insurance on non-elderly adult mortality has direct implications for other DD studies of the effect of insurance expansions on adult mortality. We provide a back of the envelope calculation here, for example for SLB (2014), who report a statistically significant near-term decline in adult mortality following the "Romneycare" health insurance expansion in Massachusetts in 2006. Massachusetts has a moderate sized population (6.55M in 2017; 14th among all states). Kolstad and Kowalski (2012) find a first stage insurance gain of 5.6%. The DD effect estimate in SLB – a 4.5% drop in amenable mortality by two years after reform –implies an 80% drop in amenable mortality for compliers.

To assess power, we build on Kaestner's (2016) replication of SLB (2014), in which he finds that their results are insignificant, using randomization inference to estimate confidence intervals.[38] We used Kaestner's code to compute the minimum effect size in their analysis with p < .05 (95% confidence). This minimum effect is 6.9%. The minimum detectable mortality decline for the newly insured, implied by this minimum effect size, is 6.9%/5.6% = 123%.

---

[37] By comparison, the population in 2013 aged 50-64 with income < 138% of FPL was around 3.4M in Full-Expansion states and 4.1M in No-Expansion states, but the first-stage for this group is around 5%, well below the 25% in the Oregon Health Insurance Experiment.

[38] We thank Robert Kaestner for providing his Stata code, which we used in our analysis.

In two more examples, we turn to recent work by two of us, in separate projects. Soni et al. (2018a) report that Medicaid expansion predicts a 2.4% relative drop in the fraction of people with cancer who are uninsured. They cannot measure the drop in uninsurance among those with undiagnosed cancer, whose baseline uninisurance rate is likely higher. Soni et al. (2018b) report a 6.4% increase in diagnoses of early-stage cancer, but do not discuss plausible effect sizes or minimum detectable effects. What first stage would be needed among those with undiagnosed cancer to make a 6.4% increase in early diagnoses plausible? A back of the envelope calculation using their reported 95% CI suggests a standard error of around 2% and thus an MDE for early-stage cancer diagnoses of around 4%.

Pines et al. (2016) find no evidence that Medicaid expansion predicts a significant increase in ED visits; their point estimate is a 0.6% drop in expansion states, relative to non-expansion states. They do not discuss the first stage (the relative drop in ED visits by uninsured persons), but from their Appendix, one can determine that the first stage is around 6.7%. Twice their standard error is .018, and .018/.067 = 0.27. This implies that if the only reason for change in ED visit rates were gaining insurance, the 95% CI around their point estimate implies a [-36%, +18%] change in ED visits by the newly insured. There is still no evidence of a higher visit rate by the newly insured, and the upper end of the 95% CI is still well below the +40% point estimate from the Oregon Health Insurance Experiment, but it one cannot rule out a fairly large increase in ED visits by the newly insured.

X. Discussion

In this paper, we examine the relationship between mortality and health insurance, principally using the DD research design used in many prior ACA studies. This design exploits the natural experiment created by variation between those states that expanded Medicaid insurance and those that did not. We also exploit variation that results from counties having varying uninsurance or poverty levels prior to 2014. We focus on persons aged 55-64 years, whose mortality rates are the most likely to be affected by health insurance. We study effects of the first three years after expansion by type of mortality (healthcare amenable vs non amenable), demographics (gender, race, and ethnicity), education level, cause of death, and residence in counties most likely to gain from the ACA expansion).

We find large confidence intervals with no statistically significant evidence of an ACA-induced decline in mortality in Medicaid expansion states. Instead, there are important non-parallel pre-treatment trends, and standard errors are far too large to allow detection of effects of plausible sizes. We confirm lack of power through a formal, simulation-based power analysis.

While it is possible that the mortality effect of the ACA health insurance expansion variation we study may materialize with more time, other factors make it unlikely they too could be statistically detected; lengthening the study period would increase likelihood that other sources of variation, including cross-border moves, the instability of insurance status over time, and the underlying causes of the non-parallel pre-treatment trends we observe, will pose challenges for credible causal inference. Moreover, our power analysis implies that an extra few years would still be insufficient to attain reasonable power, given plausible effect sizes.

We end with a discussion of the data needed to push forward the literature on the health outcome effects of health insurance. Large-scale data sets that include individual-level data on income, insurance, baseline health status, and mortality are essential. Income and prior insurance information would permit a substantially larger first stage. Baseline health data would provide a more sensitive second stage, and might also permit analysis limited to health-vulnerable subpopulations, provided that one still has reasonably sized samples. At the same time, given the power concerns we identify, studies of the health effects of health insurance should include efforts to assess the first stage, estimate reasonable magnitudes for treatment effects, and conduct a power analysis.

References

Baicker, Katherine, Sarah Taubman, Heidi Allen, Mira Bernstein, Jonathan Gruber, Joseph Newhouse, Eric Schneider, Bill Wright, Alan Zaslavsky, and Amy Finkelstein, 2013. The Oregon Experiment – Effects of Medicaid on Clinical Outcomes, 368 (18) *New England Journal of Medicine* 1713-1722.

Banerjee, Abhijit, Dean Karlan, and Jonathan Zinman (2015) "Six Randomized Evaluations of Microcredit: Introduction and Further Steps," *American Economic Journal: Applied Economics*, Vol. 7, No. 1, pp. 1–21.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan (2004) "How Much Should We Trust Differences-In-Differences Estimates?," *Quarterly Journal of Economics*, Vol. 119, No. 1, pp. 249–275.

Black, Bernard, Jose Espin-Sanchez, Eric French, and Kate Litvak (2017). The Long-term Effect of Health Insurance on Near-Elderly Health and Mortality 3 *American Journal of Health Economics* 33-59 (2017).

Borgschulte, Mark, and Jacob Vogler (2010), Did the ACA Medicaid Expansion Save Lives, working paper, University of Illinois.

Brook, Robert, John Ware, William Rogers, Emmett Keeler, Allyson Davies, Cathy Donald, George Goldberg, Kathleen Lohr, Patricia Masthay, and Joseph Newhouse., 1983. Does free care improve adults' health? Results from a randomized controlled trial. *New England Journal of Medicine.* 309(23): p. 1426-1434.

Brown, David, Amanda Kowalski and Ithai Lurie (2017). Long term impacts of childhood Medicaid expansions on outcomes in Adulthood. Working paper.

Burlig, Fiona, Louis Preonas, and Matt Woerman (2017), Panel Data and Experimental Design. Working paper University of California Berkeley.

Button, Katherine S., John P. A. Ioannidis, Claire Mokrysz, Brian A. Nosek, Jonathan Flint, Emma S. J. Robinson, and Marcus R. Munafo (2013) "Power failure: why small sample size under-mines the reliability of neuroscience," Nature Reviews Neuroscience, Vol. 14, No. 5, pp. 365– 376.

Card, David, Carlos Dobkin, and Nicole Maestas 2004. The impact of nearly universal insurance coverage on health care utilization and health: evidence from Medicare. Natl. Bur. Econ. Res. Work. Pap. 10365.

Card, David, Carlos Dobkin, and Nicole Maestas 2008. The impact of nearly universal insurance coverage on health care utilization: evidence from Medicare. *Am Econ Rev*; 98: 2242–2258.

Card, David, Carlos Dobkin, and Nicole Maestas (2009), Does Medicare Save Lives? 124 *Quarterly Journal of Economics* 597-636.

Courtemanche, Charles, James Marton, Benjamin Ukert, Aaron Yelowitz, and Daniela Zapata. 2017. Impacts of the Affordable Care Act on health insurance coverage in Medicaid expansion and non-expansion states. J*ournal of Policy Analysis and Management*, Vol. 36, No. 1, 178–210

Croke, Kevin, Joan Hamory Hicks, Eric Hsu, Michael Kremer, and Edward Miguel (2016), "Does Mass Deworming Affect Child Nutrition? Meta-analysis, Cost-Effectiveness, and Statistical Power, "Technical Report 22382, National Bureau of Economic Research, Cambridge, MA.

Currie, Janet and Jonathan Gruber. 1996a, Health insurance eligibility, utilization of medical care, and child health. 111(2) *Quarterly Journal of Economics* 431-466.

Currie, Janet and Jonathan Gruber. 1996b. Saving babies: The efficacy and cost of recent changes in the Medicaid eligibility of pregnant women. *Journal of Political Economy*, 104(6): p. 1263-1296.

Case, Anne, and Angus Deaton, 2015. Rising Morbidity and Mortality in Midlife Among White, non-Hispanic American in the 21st Century, PNAS, at http://www.pnas.org/cgi/doi/10.1073/pnas.1518393112.

Doyle, Joseph, 2005. Health insurance, treatment and outcomes: using auto accidents as health shocks. *Review of Economics and Statistics*, 87(2): p. 256-270.

Dunn, Abe and Adam Shapiro, forthcoming. Does Medicare Part D Save Lives? *American Journal of Health Economics*.

Finkelstein, Amy. 2007. The Aggregate Effects of Health Insurance: Evidence from the Introduction of Medicare, 122 *Quarterly Journal of Economics* 1-37.

Finkelstein, Amy and Robin McKnight. 2008. What Did Medicare Do? The Initial Impact of Medicare on Mortality and Out of Pocket Medical Spending, 92 *Journal of Public Economics* 1644-1668.

Finkelstein, Amy, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph Newhouse, Heidi Allen, Katherine Baicker, and the Oregon Health Study Group. 2012. The Oregon Health Insurance Experiment: Evidence from the First Year, 127 *Quarterly Journal of Economics* 1057-1106.

Frean, Molly, Jonathan Gruber and Benjamin Sommers. 2017. Premium subsidies, the mandate, and Medicaid expansion: Coverage effects of the Affordable Care Act. 53 *Journal of Health Economics* 72-86.

Galea, Sandro, Melissa Tracy, Katherine J. Hoggatt, Charles Dimaggio, and Adam Karpati. 2011. Estimated Deaths Attributable to Social Factors in the United States. *American Journal of Public Health* 101 (8): 1456–65.

Gallet, Craig A. and Hristos Doucouliagos (2017) "The impact of healthcare spending on health outcomes: A meta-regression analysis," *Social Science & Medicine*, Vol. 179, pp. 9–17.

Gaudette, E. G. Pauley and J. Zissimopoulos. 2016. Lifetime consequences of early and midlife access to health insurance: a review. Working paper.

Gelman and Carlin. 2014. Beyond Power Calculations. *Perspectives on Psychological Science* 2014, Vol. 9(6) 641–651

Gelman, Andrew (2018) "Don't Calculate Post-hoc Power Using Observed Estimate of Effect Size," *Annals of Surgery*, Vol. XX, No. Xx, p. 1.

Gerin, William, Christine Kapelewski, and Niki L. Page (2017) *Writing the NIH Grant Proposal: A Step-by-Step Guide*, Los Angeles, CA: SAGE Publications, 3rd edition.

Goldman, Dana, Jayantha Bhattacharya, McCaffrey DF, Duan N, Arleen Leibowitz et al. 2001. The effect of insurance on mortality in an HIV+ population in care. *J. Am. Stat. Assoc.* 96(455):883–894

Ghosh, Ausmita, Benjamin Sommers and Kosali Simon. 2017. The Effect of State Medicaid Expansions on Prescription Drug Use. NBER WP 23044 www.nber.org/papers/w23044

Hannon, Susan J., Kathy Martin, Len Thomas, and Jim Schieck (1993) "Investigator Disturbance and Clutch Predation in Willow Ptarmigan : Methods for Evaluating Impact," *Journal of Field Ornithology*, Vol. 64, No. 4, pp. 575–586.

Hoenig, John M. and Dennis M. Heisey (2001) "The Abuse of Power: The Pervasive Fallacy of Power Calculations for Data Analysis," *The American Statistician*, Vol. 55, No. 1, pp. 19–24.

Howell, Embry. et al. 2010, Declining child mortality and continuing racial disparities in the era of the Medicaid and SCHIP insurance coverage expansions. Am J Public Health, 2010. 100(12): p. 2500-6.

Hsiang, Solomon M, Marshall Burke, Edward Miguel, Kyle Meng, and Mark Cane (2015) "Anal ysis of statistical power reconciles drought-conflict results in Africa.", Center for Effective Global Action working paper, at http://escholarship.org/uc/item/77s421cd.

Hu, L., Robert Kaestner, B. Mazumder, Sarah Miller, and A. Wong. 2016. The Effect of the Patient Protection and Affordable Care Act Medicaid Expansions on Financial Well-Being. NBER Working Paper No. 22170. Cambridge, MA: National Bureau of Economic Research.

Ioannidis, John, T.D. Stanley, and Hristos Doucouliagos (2017). The Power of Bias in Economics Research, *Economics Journal* 127: F236-F265.

IntHout, Joanna, John PA Ioannidis, and George F Borm (2016) "Obtaining evidence by a single well-powered trial or several modestly powered trials," *Statistical Methods in Medical Research*, Vol. 25, No. 2, pp. 538–552.

Ioannidis, John P A (2005) "Why Most Published Research Findings Are False," *PLoS Medicine*, Vol. 2, No. 8, p. e124.

Ioannidis, John P. A., T. D. Stanley, and Hristos Doucouliagos (2017) "The Power of Bias in Economics Research," *The Economic Journal*, Vol. 127, No. 605, pp. F236–F265.

Jayachandran, Seema, Adriana Lleras-Muney, and Kimberly Smith., 2010. Modern Medicine and the Twentieth Century Decline in Mortality: Evidence on the Impact of Sulfa Drugs. *American Economic Journal: Applied Economics* 2 (April): 118–146.

Kaestner, Robert. 2016. Did Massachusetts Health Care Reform Lower Mortality? No According to Randomization Inference. Statistics and Public Policy, 3(1): 1-6.

Kaestner, Robert, Bowen Garrett, A. Gangopadhyaya, and C. Fleming. 2015. Effects of the ACA Medicaid Expansion on Health Insurance Coverage and Labor Supply. NBER Working Paper No. 21836. Cambridge, MA: National Bureau of Economic Research.

Kaiser Family Foundation. 2016. Medicaid and the uninsured: Where are states today? Medicaid and CHIP eligibility levels for children and non-disabled adults. The Henry J. Kaiser Family Foundation.

Keeler E. 1985. How free care reduced hypertension in the health insurance experiment. *Journal of the American Medical Association* 254:1926–31.

Kenney, Genevieve, Victoria Lynch, Jennifer Haley, and Michael Huntress. 2012. Variation In Medicaid Eligibility and Participation among Adults: Implications for the Affordable Care Act. *Inquiry* 49:231–53.

Kolstad, Jonathan, and Amanda Kowalski 2012. The Impact of Health Care Reform on Hospital and Preventive Care: Evidence from Massachusetts, 96 *Journal of Public Economics* 909-929.

Kowalski, Amanda. 2018. "Extrapolation using Selection and Moral Hazard Heterogeneity from within the Oregon Health Insurance Experiment". NBER Working Paper 24647, May 2018.

Kronick, Richard. 2009. Health Insurance Coverage and Mortality Revisited. *Health Services Research* 44 (4): 1211–31.

Levy, Helen, and David Meltzer. 2004. What do we really know about whether health insurance affects health? In *Health Policy and the Uninsured: Setting the Agenda*, ed. C. McLaughlin, pp. 179–204.Washington, DC: Urban Inst. Press

Levy, Helen, and David Meltzer, 2008. The impact of health insurance on health. *Annu. Rev. Public Health*, 2008. 29: p. 399-409.

Maxwell, Scott E. (2004). The Persistence of Underpowered Studies in Psychological Research: Causes, Consequences, and Remedies. *Psychological Methods* 9(2), pp. 147–163.

Mazurenko, Olena, Casey Balio, Rajender Agarwal, Aaron Carroll, and Nir Menachemi. 2018. The effects of Medicaid Expansion under the ACA: A systematic review. June. Health Affairs. *Health Affairs* 37, NO. 6 (2018): 944–950.

McCrary, Justin, Garret Christensen, and Daniele Fanelli (2016), Conservative Tests under Satisficing Models of Publication Bias, 11(2) *PLOS One* e0149590.

McClellan, Chandler, 2017. The Affordable Care Act's Dependent Care Coverage and Mortality. *Medical Care* Volume 55, Number 5, May 2017.

McCloskey, D.R., 1985. The Loss Function Has Been Mislaid: The Rhetoric of Significance Tests, American Economic Review 75(2)(Papers and Proceedings) 201-205.

McCloskey, Deirdre N and Stephen T Ziliak (1996) "The Standard Error of Regressions," *Journal of Economic Literature*, Vol. 34, No. 1, pp. 97–114.

Newhouse, Joseph, and Insurance Experiment Group. 1993. *Free for All? Lessons from the RAND Health Insurance Experiment*. Cambridge, MA: Harvard Univ. Press.

NIH (2016) "Reviewer Guidance on Rigor and Transparency: Research Project Grant and Mentored Career Development Applications."

Nolte E, McKee M. 2003. Measuring the health of nations: analysis of mortality amenable to health care. BMJ. 327:1129. [PMID: 14615335]

Pines Jesse, Mark Zocchi, Ali Moghtaderi, Bernard Black, Steven Farmer, Greg Hufstetler, Kevin Klauer, and Randy Pilgrim (2016), The Impact of the 2014 Medicaid Expansion on Hospital-Based Emergency Department Visits, 35 *Health Affairs* 1480-1486.

Polsky, Daniel., Ph.D., Michael Richards, M.D., Ph.D., Simon Basseyn, B.A., Douglas Wissoker, Ph.D., Genevieve M. Kenney, Ph.D., Stephen Zuckerman, Ph.D., and Karin V. Rhodes, M.D. 2015. Appointment Availability after Increases in Medicaid Payments for Primary Care. February 5. N Engl J Med 2015; 372:537-545 DOI: 10.1056/NEJMsa1413299

Powell, David (2018), Imperfect Synthetic Controls: Did the Massachusetts Health Care Reform Save Lives? Working paper, at https://works.bepress.com/david_powell/24/.

Rubin, Donald B. (2008), For Objective Causal Inference, Design Trumps Analysis, 2 *Annals of Applied Statistics* 808-840.

Senn, S. J (2002) "Power is indeed irrelevant in interpreting completed studies," *BMJ*, Vol. 325, No. 7375, pp. 1304–1304.

Simon, Kosali, A. Soni, and J. Cawley (2017). The Impact of Health Insurance on Preventive Care and Health Behaviors: Evidence from the First Two Years of the ACA Medicaid Expansions. *Journal of Policy Analysis and Management*, 36: 390-417.

Smith, J.C., and C. Medalia (2015) U.S. Census Bureau, Current Population Reports, P60-253, Health Insurance Coverage in the United States: 2014. U.S. Government Printing Office, Washington, DC. Available from: https://www.census.gov/content/dam/Census/library/publications/2015/demo/p60-253.pdf.

Sommers, B.D., K. Baicker, and A.M. Epstein, *Mortality and access to care among adults after state Medicaid expansions.* N Engl J Med, 2012. **367**(11): p. 1025-34.

Sommers, Benjamin, Sharon Long, and Katherine Baicker. 2014. *Changes in mortality after Massachusetts health care reform: a quasi-experimental study.* Annals of Internal Medicine, 160(9):585-593.

Sommers, Benjamin. 2017. State Medicaid Expansions and Mortality, Revisited: A Cost-Benefit Analysis . *American Journal of Health Economics*. vol 3 issue 3 p.392-421.

Soni, Aparna, John Cawley, Lindsay Sabik, and Kosali Simon. 2018a. Effect of Medicaid Expansions of 2014 on Overall and Early-Stage Cancer Diagnoses. *American Journal of Public Health* 108, no.2. Feb 1) pp.216-218.

Soni, A., K. Simon, L. Sabik, and S. Sommers. 2018b Changes in Insurance Coverage Among Cancer Patients Under the Affordable Care Act . *JAMA Oncology,* Jan 2018 Vol 4 No 1 p.122.

Stigler, Stephen (1977) "Do robust estimators work with real data?," *Annals of Statistics*, Vol. 5, No. 6, pp. 1055–1098.

Taubman SL, Allen HL, Wright BJ, Baicker K, Finkelstein AN. 2014 Medicaid increases emergency-department use: evidence from Oregon's health insurance experiment. *Science*, 343:263-268

Thomasson, M. a., & Treber, J. (2008). From Home to Hospital: The Evolution of Childbirth in the United States, 1928–1940. *Explorations in Economic History*, *45*(1), 76–99. https://doi.org/10.1016/j.eeh.2007.07.001

Taylor-Robinson, David C., Nicola Maayan, Karla Soares-Weiser, Sarah Donegan, and Paul Gar- ner (2015) "Deworming drugs for soil-transmitted intestinal worms in children: effects on nu- tritional indicators, haemoglobin, and school performance," *Cochrane Database of Systematic Reviews*, Vol. 2015, No. 7.

Weathers, Robert and M. Stegman, 2012. The effect of expanding access to health insurance on the health and mortality of Social Security Disability Insurance beneficiaries. *Journal of Health Economics*. 31(6): p. 863-75.

Wherry, Laura and Bruce Meyer, 2015. Saving Teens: Using a Policy Discontinuity to Estimate the Effects of Medicaid Eligibility. *Journal of Human Resources*, 51(3): 556-588.

Wherry, Laura and Sarah Miller. 2016. Early Coverage, Access, Utilization, and Health Effects Associated With the Affordable Care Act Medicaid Expansions: A Quasi-experimental Study. *Annals of Internal Medicine*, 164 (12): 795-803.

Zhang, L., & Ortmann, A. (2013). *Exploring the Meaning of Significance in Experimental Economics. Australian School of Business Working Paper. No. 2013 ECON 32.* https://doi.org/10.2139/ssrn.2356018

Ziliak, Stephen T. and Deirdre N. McCloskey (2004) "Size matters: the standard error of regressions in the American Economic Review," *The Journal of Socio-Economics*, Vol. 33, No. 5, pp. 527–546.

Table 1. Full Expansion; Substantial Expansion; Mild Expansion, and No-Expansion states, and % Uninsured for Selected Years

Table shows expansion status of each state (including D.C.). Treatment group is Full Expansion states and control group is No-Expansion states, for these states, table shows expansion date if other than Jan. 1, 2014. For "substantial" and "mild" expansion states, table shows year of significant prior Medicaid expansion. Summary rows give either equal weight to all states in each expansion group, or population weight, as indicated. See Appendix Table A-1 for additional details and sources for each state's expansion status.

State	Expansion Date	% uninsured (age 50-64) 2013	2014	2016	change in % unins. (2013-2016)
Full Expansion		**13.7**	**9.4**	**6.4**	**7.3**
Pop. weighted		**13.4**	**9.4**	**6.3**	**7.1**
Alaska	Sep 2015	19.1	17.2	13.8	5.3
Arizona[1]		17.6	13.1	10.0	7.6
Arkansas[2]		16.5	11.3	6.9	9.6
Colorado[3]		13.6	9.2	6.7	6.9
Illinois		14	9.8	6.3	7.7
Indiana	Feb 2015	12.9	11	7.5	5.4
Iowa[4]		7.7	5.9	3.9	3.8
Kentucky		14.4	7.5	4.8	9.6
Maryland		10	7.2	5.2	4.8
Michigan	Apr 2014	11.4	8.2	5.1	6.3
Montana	Jan 2016	18	14.4	9.2	8.8
Nevada		19.8	14.3	10.5	9.3
New Hampshire	Aug 2014	11.6	9.7	6.2	5.4
New Jersey[5]		13.1	10.8	7.1	6.0
New Mexico		19	15	9.8	9.2
North Dakota		9.6	6.9	5.2	4.4
Ohio		12.3	8.3	5.6	6.7
Oregon[6]		15.3	9.6	6.4	8.9
Pennsylvania	Jan 2015	9.5	7.7	5.0	4.5
Rhode Island		11.2	6.2	3.4	7.8
Washington[5]		13.1	8.2	5.7	7.4
West Virginia		14.5	8.9	5.9	8.6
Substantial Expansion		**10.3**	**7.1**	**4.8**	**5.5**
Pop. weighted		**15.0**	**10.3**	**6.4**	**8.6**
California[5]	2010	18.1	12.3	7.4	10.7
Connecticut[5]	2010	9.7	6.2	4.4	5.3
Hawaii[7]	1994	7.3	5.5	4.1	3.2
Minnesota[5]	2010	7.3	5.1	3.6	3.7
Wisconsin[8]	2009	9.1	6.6	4.7	4.4
Mild Expansion		**7.6**	**5.7**	**4.0**	**3.6**
Pop. weighted		**8.6**	**6.8**	**4.6**	**4.0**
Delaware[9]	1996	9.5	7.2	5.2	4.3
Dist. of Columbia[5]	2010	6.7	5.2	3.2	3.5
Massachusetts[10]	2006	3.5	3	2.5	1.0
New York[11]	2001	10.4	8.2	5.4	5.0
Vermont[12]	1996	7.9	5	3.7	4.2
No Expansion		**15.2**	**12.7**	**10.2**	**5.0**
Pop. weighted		**16.5**	**13.7**	**11.0**	**5.4**

State	Expansion Date	% uninsured (age 50-64)			change in % unins.
		2013	**2014**	**2016**	**(2013-2016)**
Alabama		13.4	11.6	9.5	3.9
Florida		22	17.9	13.8	8.2
Georgia		18.2	15.2	12.5	5.7
Idaho		17.1	12.9	11.5	5.6
Kansas		11.8	9.7	7.8	4.0
Louisiana	Jul 2016	17.8	15.6	10.8	7.0
Maine		12.5	11	8.7	3.8
Mississippi		18.8	15.4	13.1	5.7
Missouri		13.3	10.3	8.9	4.4
Nebraska		10.7	8.8	7.8	2.9
North Carolina		15.8	12.6	10.5	5.3
Oklahoma		18.1	15.6	13.9	4.2
South Carolina		17.1	13.7	10.9	6.2
South Dakota		11.3	9.3	8.8	2.5
Tennessee		15	12.7	9.4	5.6
Texas		21	17.4	15.5	5.5
Utah		13	11.3	8.8	4.2
Virginia		12.3	10.8	8.3	4.0
Wyoming		13.3	12.7	11.5	1.8
National		**13.5**	**10.4**	**7.7**	**5.8**
Pop. weighted		**14.6**	**11.2**	**8.3**	**6.3**

Table 2: DD and Triple-Difference Estimates: Effect of Medicaid Expansion on Mortality
County-level regressions, with county and year FE and population weights, of *ln*((mortality/100,000 persons)+1) over 2009-2016 on full-Expansion dummy (=1 for Full-Expansion States in expansion years; 0 otherwise), and covariates (same as in Figure 2, used in even-numbered regressions. Third difference (regressions (5)-(6)) is ages 55-64 versus aged 65-74. Standard errors use state clusters. *.**, *** indicates statistical significance at the 10%, 5%, and 1% levels, respectively; significant results at 5% level or better in **boldface**.

	DD 55-64 years		DD 65-74 years		Triple diff.	
	(1)	(2)	(3)	(4)	(5)	(6)
Healthcare Amenable Mortality						
Full Expansion Dummy	**-0.018** **(0.009)**	**-0.018** **(0.007)**	**-0.013** **(0.006)**	-0.008 (0.006)	-0.002 (0.009)	-0.004 (0.008)
Full Expansion Dummy x Age 55-64 Dummy					-0.002 (0.009)	-0.004 (0.008)
Non-amenable Mortality						
Full Expansion Dummy	0.016 (0.010)	0.010 (0.009)	**0.021** **(0.010)**	**0.020** **(0.009)**	-0.002 (0.013)	-0.006 (0.011)
Full Expansion Dummy x Age 55-64 Dummy					-0.002 (0.013)	-0.006 (0.011)
All Mortality						
Full Expansion Dummy	-0.006 (0.008)	-0.009 (0.006)	-0.003 (0.005)	-0.001 (0.005)	0.000 (0.008)	-0.003 (0.007)
Full Expansion Dummy x Age 55-64 Dummy					0.000 (0.008)	-0.003 (0.007)
County Population Weights	Yes	Yes	Yes	Yes	Yes	Yes
Year and County FE	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	No	Yes	No	Yes	No	Yes
Observations	22,464	22,464	22,464	22,464	44,928	44,928

Figure 1. Time Trends in Amenable Mortality for Persons Aged 55-64

Figure shows amenable mortality rate for persons age 55-64 for Full-Expansion, Substantial Expansion, Mild Expansion, and Non-Expansion States, over 1999-2016, using county population weights. State groups are defined in Table 1. Vertical line separate pre-expansion from expansion period.

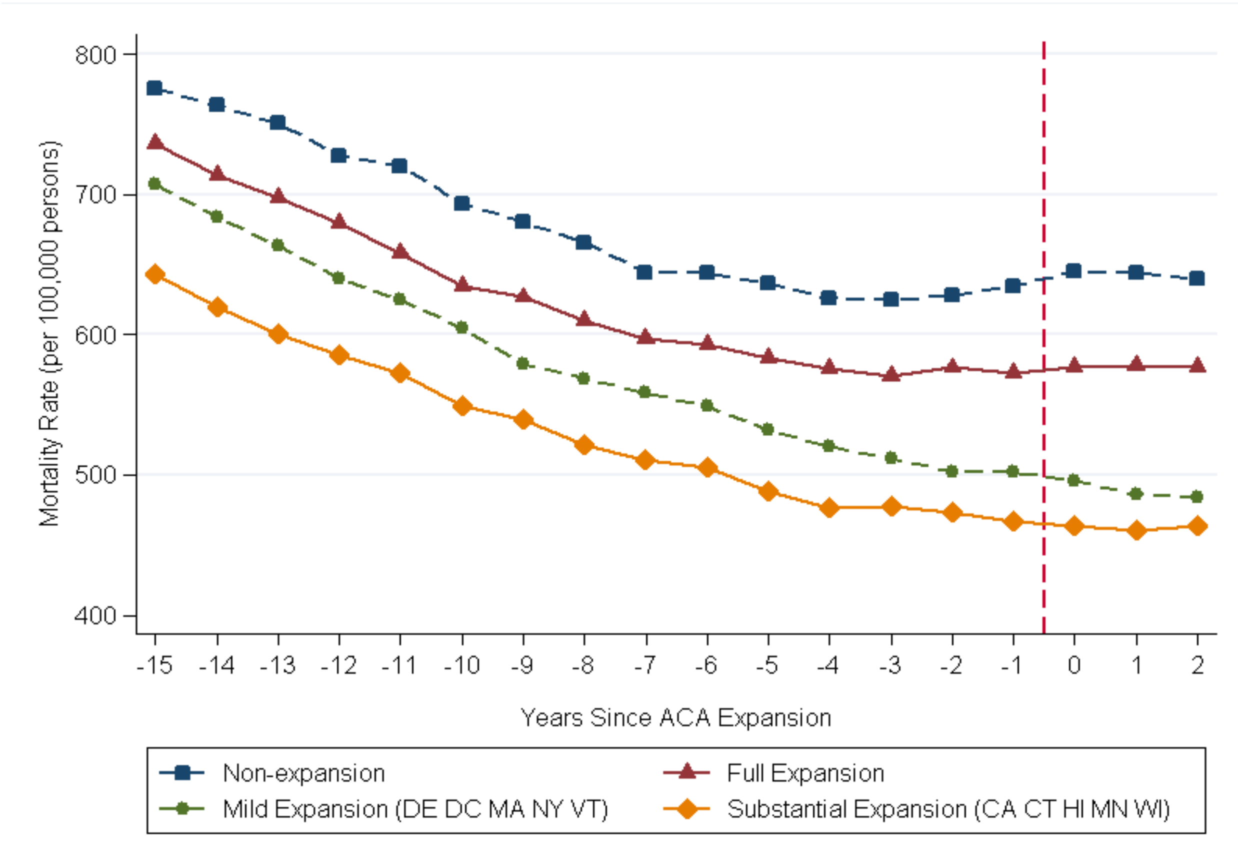
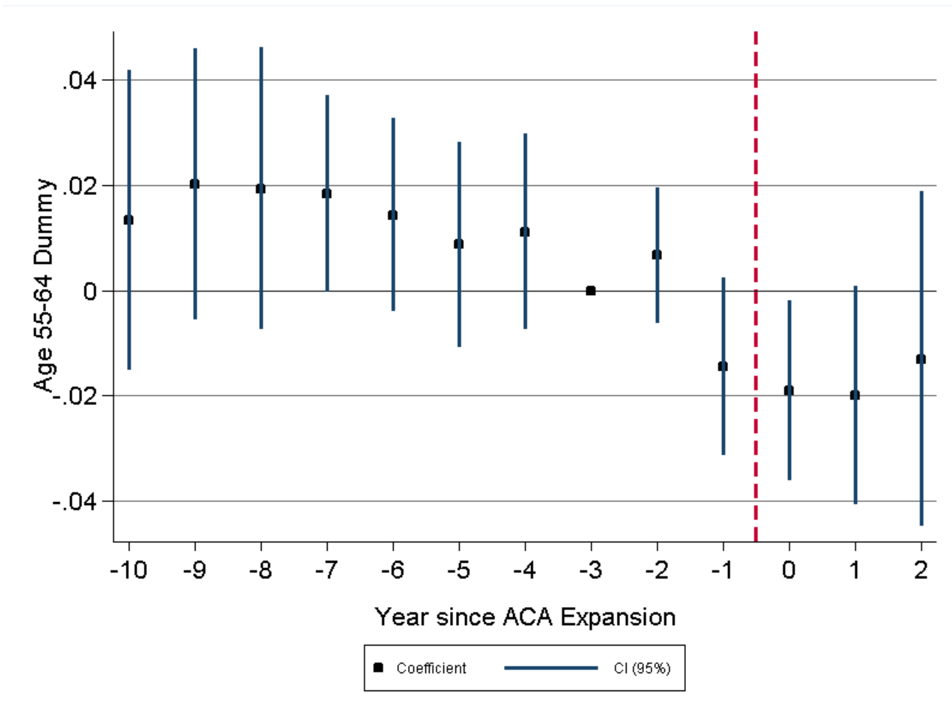


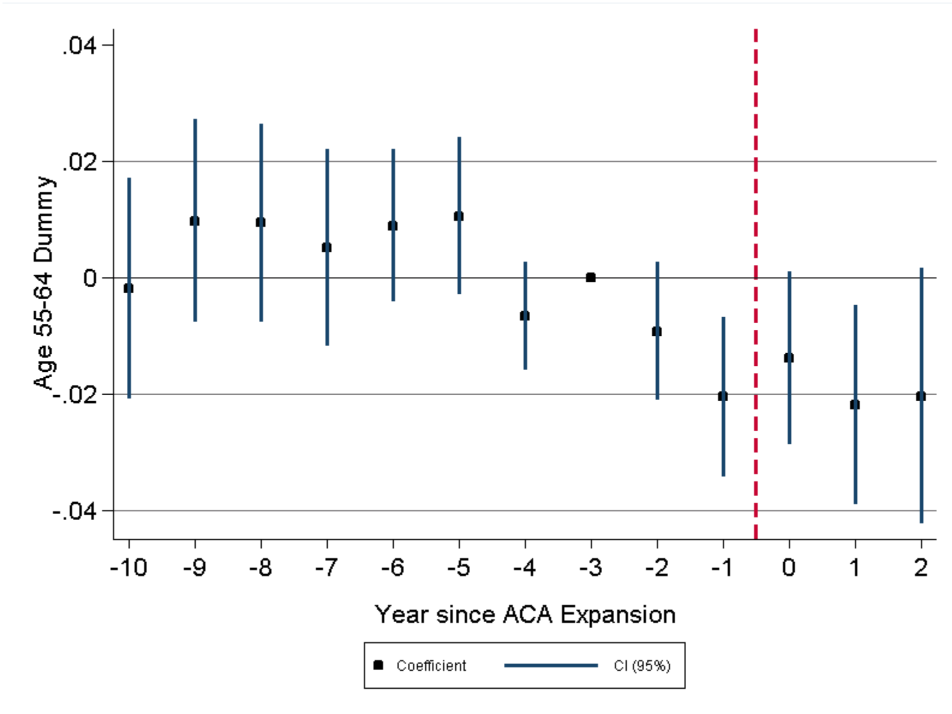
Figure 2. Leads-and-Lags Results for Ages 55-64 and 65-74, Amenable Mortality

Graphs from leads and lags regressions of *ln*((amenable mortality/100,000 persons)+1) for Full-Expansion States versus control group of Non-Expansion States, over 2004-2016 are shown in panels A (age 55-64) and B (age 65-74). Panel C shows triple difference results, with age 55-64 versus 65-64 as the third difference. Panel D shows age discontinuity results, comparing age 55-64 to age 65-74 with state (both Full and No-Expansion States together. Covariates are listed in paper. Regressions include county and year FE, and county-population weights. y-axis shows coefficients on lead and lag dummies; vertical bars show 95% confidence intervals (CIs) around coefficients, using standard errors clustered on state. Coefficient for year -3 is set to zero.

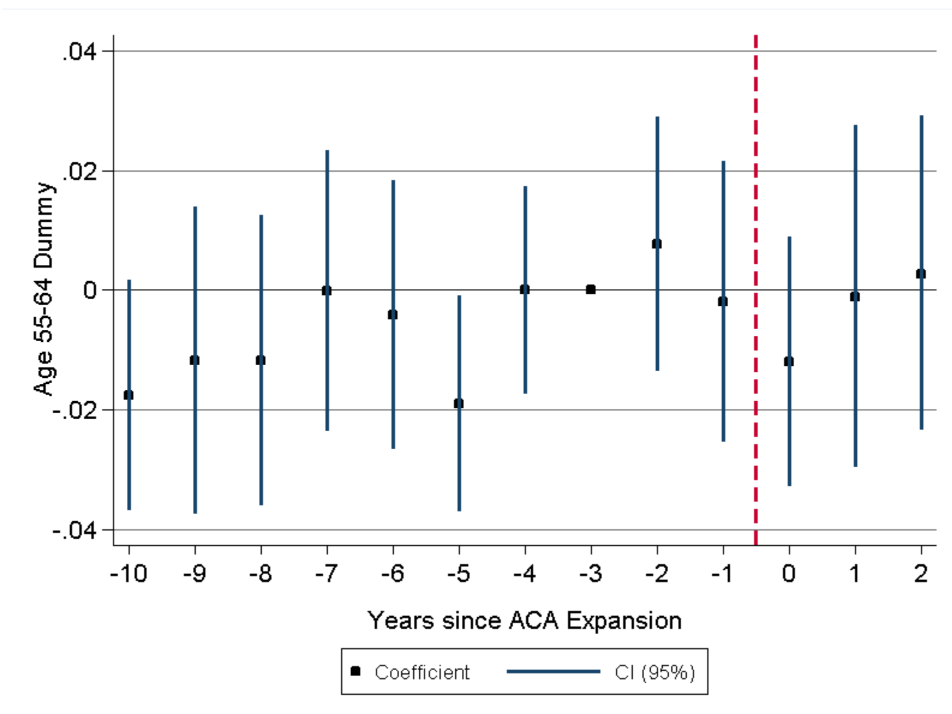
Panel A. Amenable Mortality for Ages 55-64



Panel B. Amenable Mortality for Ages 65-74



Panel C. Triple difference. Leads and lags graphs for amenable mortality for persons age 55-64 in Full-Expansion States, relative to (i) persons age 65-74 in Full-Expansion States, and (ii) persons age 55-64 in Non-Expansion States.



Panel D. Age Discontinuity. Leads and lags graphs for amenable mortality among persons aged 55-64, versus those aged 65-74 old in Full and No-Expansion States.

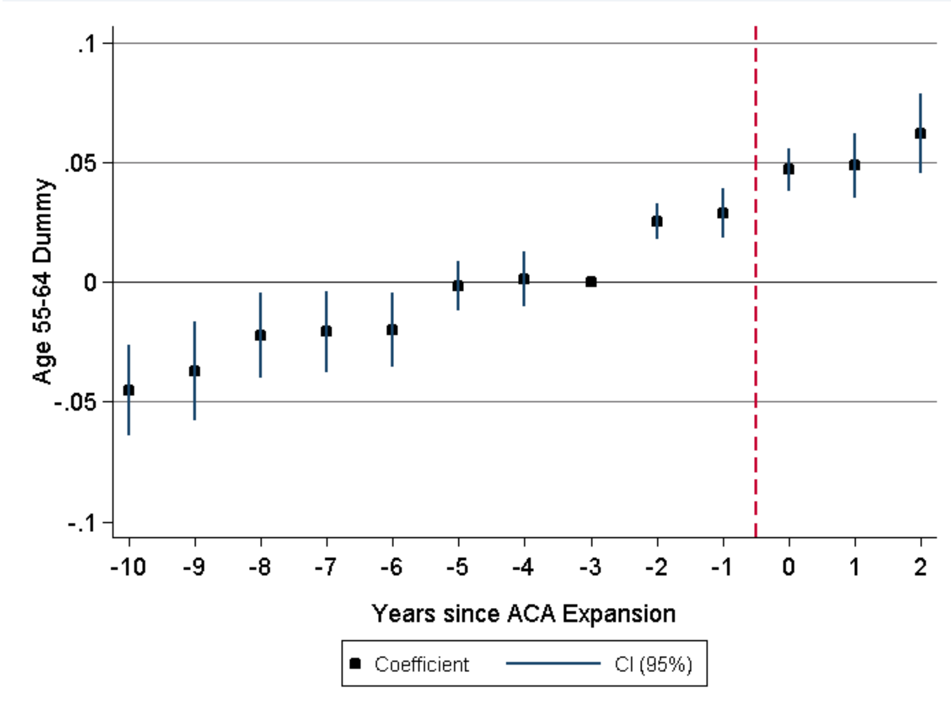
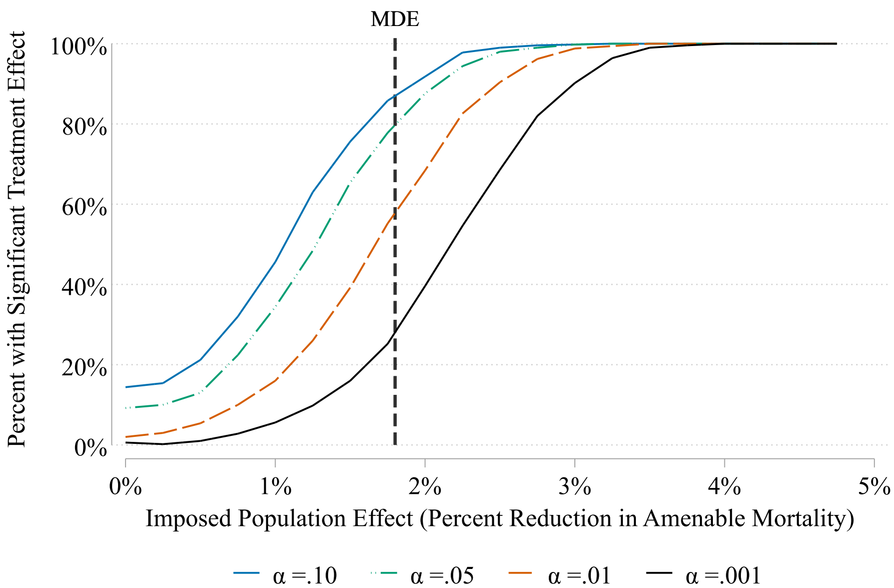


Figure 3: Simulation-Based Power Analysis

Power curves for simulated Medicaid expansion, as of January 1, 2012, applied to persons aged 55-64 during pre-treatment period (2007-2013). Graphs show power (likelihood of detecting a statistically significant effect on amenable mortality, at the indicated confidence levels, for two-tailed test), given imposed "true" population average effect. Curves are based on 1,000 replications of the DD and triple difference regressions models used in Table 2. In each draw, we select 20 pseudo-treated states at random from the combined set of 41 treated and control states, and remove a fraction of the observed deaths at random from the treated states, where the fraction removed corresponds to an assumed true treatment effect, and vary the imposed treatment effect from 0-5% in increments of 0.1%. Curves for α = .10/.05/.01/.001 correspond to 90%/95%/99%/99.9% confidence levels, respectively. Dashed vertical line shows MDE.

Panel A. DD Analysis



Panel B. Triple Difference Analysis

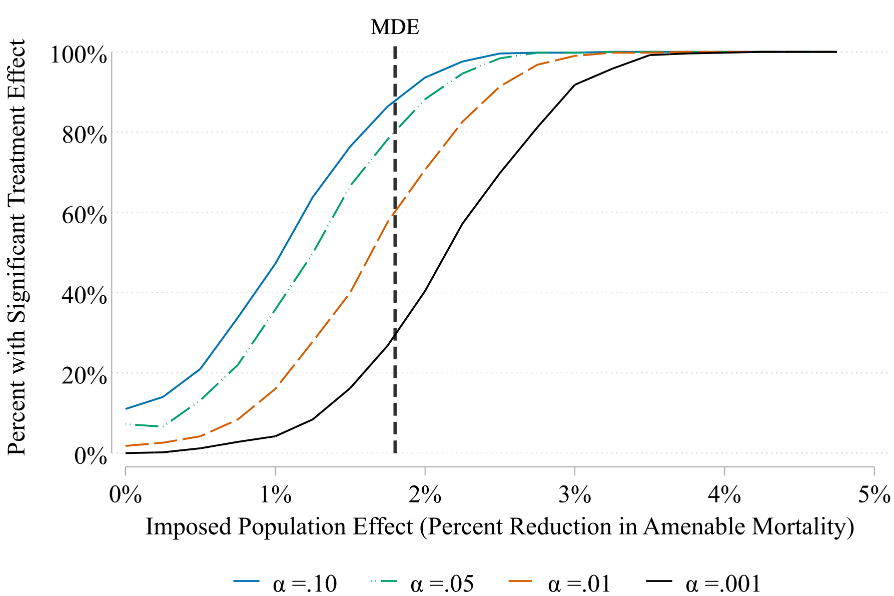
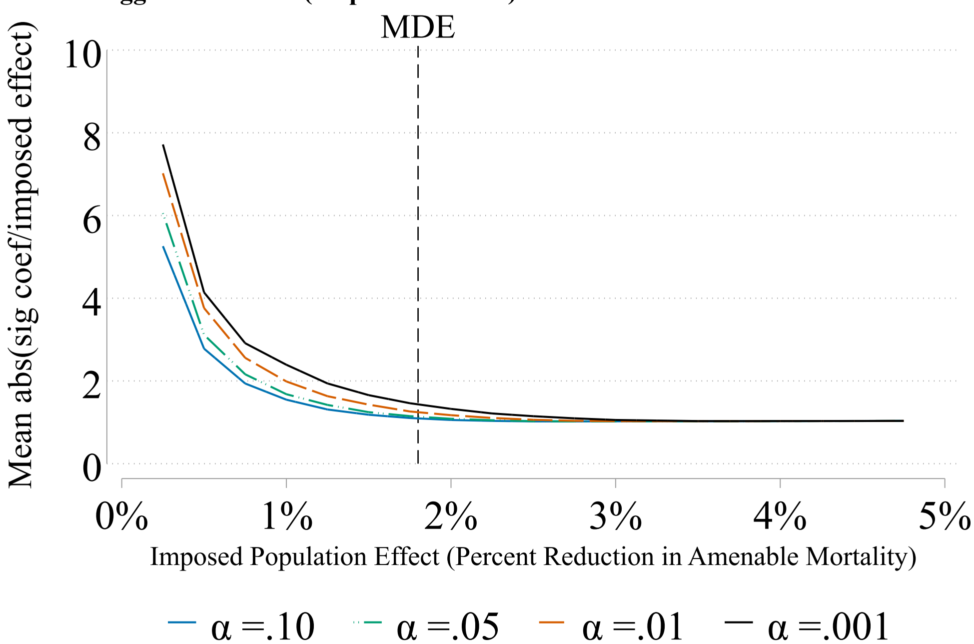


Figure 4. Power Analysis Extensions: Exaggeration Ratio and Likelihood of Wrong Sign

We conduct the same power analyses as in Figure 3 and then plot, for the instances in which a statistically significant effect is found at the indicated confidence levels, the ratio of |estimated effect|/imposed true effect ("exaggeration ratio") (Panel A), and the likelihood that the sign of the estimated effect is opposite from the imposed true effect. Curves for α = .10/.05/.01/.001 correspond to 90%/95%/99%/99.9% confidence levels, respectively. Dashed vertical line shows MDE.

Panel A. Exaggeration Ratio (Triple Difference)



Panel B. Probability that Estimated Effect, if Significant, Has Wrong Sign (Triple Difference)

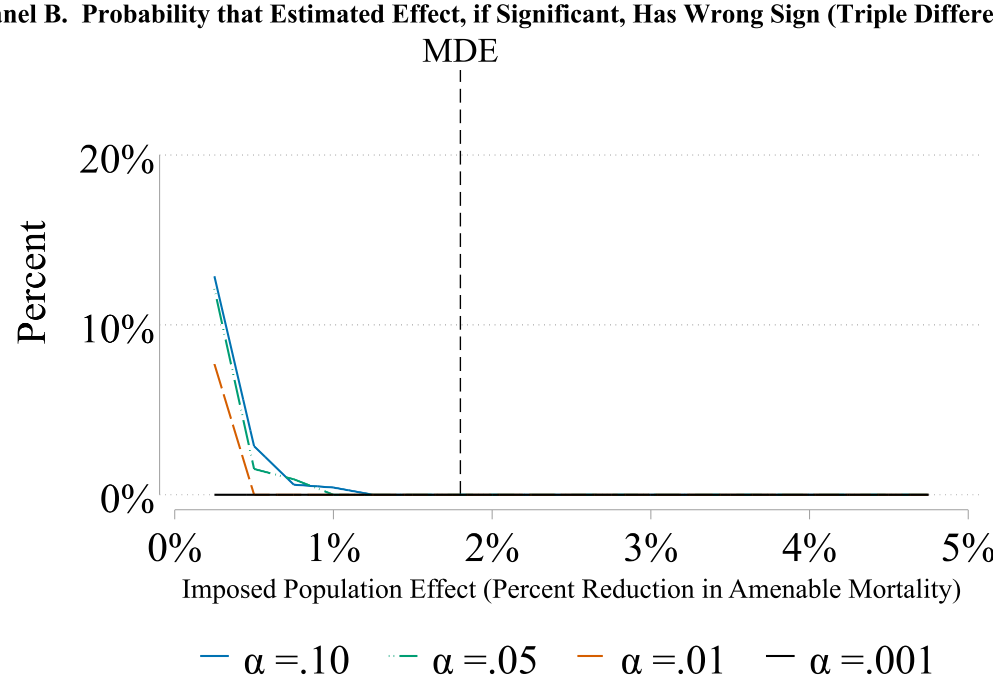


Figure 5. Pseudo-Shocks to Individual States in 2012-2013

Scatter plot of pseudo-treatment effects for individual Full-Expansion and No-Expansion states, using a sample period of 2007-2013 and a pseudo-shock to that state at Jan. 1, 2012, using the remaining Full- and No-Expansion states as a control group. Treatment effects are estimated using the DiD model as in Table 2. Downward sloping line is regression line for regression of pseudo-treatment effect on ln(state population in 2012) and constant term.

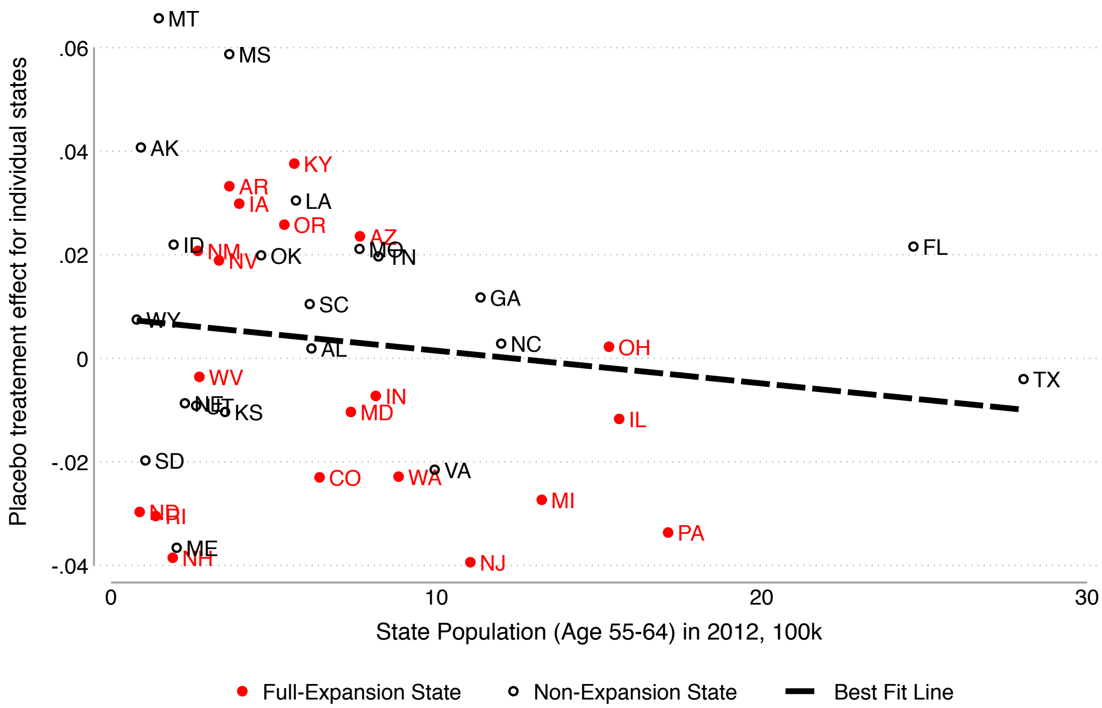
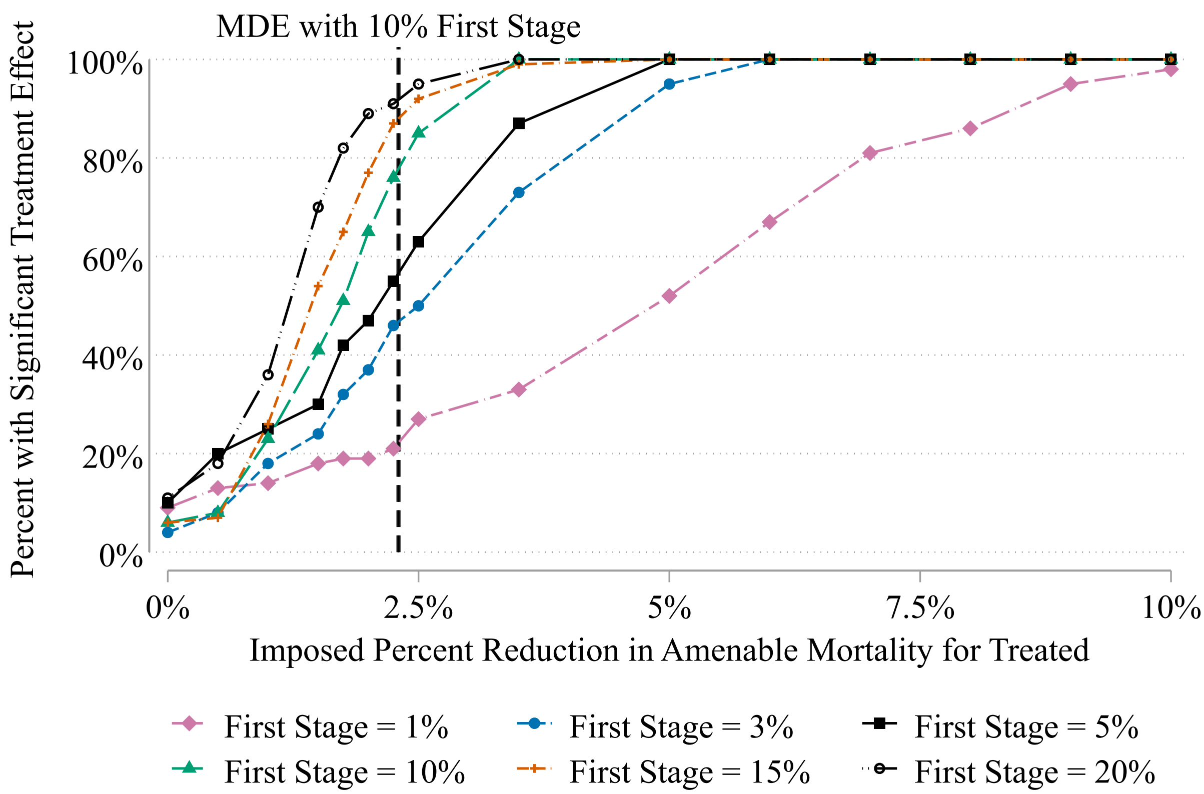


Figure 6. Simulation Based Power Analysis with Known Mortality Status of Decedent

Power curves for simulated Medicaid expansion, as of January 1, 2012, applied to persons aged 55-64 during pre-treatment period (2007-2013). Graphs show power (likelihood of detecting a statistically significant effect on amenable mortality, at the indicated confidence levels, for two-tailed test), given imposed "true" population average effect. Curves are based on 1,000 replications of a triple difference specification. In each draw, we select 20 pseudo-treated states at random from the combined set of 41 treated and control states. We further break each county into a treated and untreated population. We remove a fraction of the observed deaths at random from the treated states and treated portions of each county, where the fraction removed corresponds to an assumed true treatment effect, and vary the imposed treatment effect from 0-5% in increments of 0.1%. Curves for α = .10/.05/.01/.001 correspond to 90%/95%/99%/99.9% confidence levels, respectively. All control variables and standard errors are as in Table 2.



Appendix for

The Effect of Health Insurance on Mortality: Power Analysis and What We Can Learn from the Affordable Care Act Coverage Expansions

Bernard Black
Northwestern University, Pritzker Law School and Kellogg School of Management

Alex Hollingsworth
Indiana University, School of Public and Environmental Affairs.

Leticia Nunes
Fundação Getulio Vargas, Department of Economics

Kosali Simon
Indiana University, School of Public and Environmental Affairs, and NBER

Draft February 2019

Abstract

This Appendix contains additional methods details and results for Black, Hollingsworth, Nunes, and Simon, *The Effect of Health Insurance on Mortality: Power Analysis and What Can We Learn from the Affordable Care Act Coverage Expansions?*

A1. Synthetic Control Results

We sought to assess whether we could obtain a better match between treated and control states, and thus tighter confidence bounds, using synthetic control methods. We used two approaches. In the first, we combined the Full-Expansion States into a single treated unit and used usual synthetic control methods (Abadie, Diamond, and Hainmueller, 2010)[1] to construct a synthetic match using the Non-Expansion States as donor states. We report results in Figure A-2, and report the weights on donor states in Table A-12.

The synthetic control approach minimizes the difference between the pre-treatment mortality rates of the treated states and a weighted combination of the Non-Expansion States. However, the maximum difference between the two series is still sizeable, at around 0.02 in 2007. Moreover, visually, a large gap arises in 2013. Thus, this approach fails to create a close enough match in 2013 for this method to produce a satisfying solution to our concern with non-parallel trends. We were not persuaded that, for our data, the synthetic control approach is an improvement over the triple-difference design.[2]

We also considered an extension of the synthetic control strategy, following Xu (2017). Xu's "generalized synthetic control (gsynth)" method generates a separate synthetic control for each full-expansion state, drawn from the non-expansion states. One can then conduct DD analyses on the resulting treated and control units, and obtain analytical standard errors (which the original method does not provide). This procedure does not allow for weighting different units. We therefore only discuss state-level results.[3] While we cannot exactly replicate our triple difference models using the gsynth method, we constructed an approximation, by using as the treated units each treated state's 55 to 64 year olds, and as the donor pool both every non-expansion state's 55 to 64 year olds and every state's (expansion or not) 65 to 74 year olds. We present results in Appendix Figure A-3. Similar to the simpler synthetic control method presented above, there is a large drop in amenable mortality in Full-Expansion States in 2013; mortality in expansion states then rebounds in 2014. The poor pre-period fit is even more pronounced with county-level data, and is driven by small counties, which have highly varying death

[1] We used code for this approach from Soni (2016).

[2] A further concern with the synthetic control approach is that it gives zero weight to most donor states and assigns positive weights to several very-low-population states (Alaska, Maine, Wyoming) that do not otherwise seem good matches for the Full-Expansion States. Appendix Table A-8 shows the weights on each donor state.

[3] Although we could not directly use population weights within Xu's method, we simulate doing so by repeatedly running his procedure on bootstrapped datasets with draws weighted by population. Results, with both state-level and county-level data, were similar to those we discuss in the text.

rates and are hard to fit even with a large donor pool. We concluded that the gsynth approach cannot be reliably applied to our data

A2. Results for Different Demographic Groups

In this and the next two sections, we assess the effects of Medicaid expansion on mortality for various subgroups. The demographic groups we consider are males, females, non-Hispanic blacks, non-Hispanic whites, and Hispanics. We also consider subgroups based on education and mortality based on cause of death. Our data has limitations for all subgroups except gender. For race and ethnicity, we can obtain estimates of the first stage (change in uninsurance rates) only at the state level, not the county level, due to limitations of the SAHIE data. The DD design does not explicitly use the first stage, but it is central to assessing what coefficient magnitudes are reasonable. For education, population data is available only for broad age groups (45-64 and 65+; 5-year average). For analysis by prior insurance status and by income, we observe percent uninsured and percent below 138% of the FPL threshold for full ACA expansion at the county*year level, but cannot directly study these subsamples because the mortality data does not contain information on income or insurance.

We begin our analysis of demographic subgroups in Figure A-5 with leads-and-lags graphs of the triple differences in amenable mortality for samples subdivided on gender and on race/ethnicity. Most post-expansion point estimates are insignificant. The exception is non-Hispanic Blacks, who show a post-expansion drop in mortality. However, for this subgroup, we observe non-parallel pre-treatment trends even with the triple-difference specification; the post-expansion drop in mortality could merely reflect continuation of those trends. Also, the first stage for non-Hispanic Blacks is not greatly different from that for the population as a whole (Table A-3). Thus, the point estimates in Figure A-5 (around -0.05) are not possible as true effects of Medicaid expansion.

We turn next to DD and triple-difference regression results for amenable mortality for these subsamples, starting with demographic subsamples in Table A-3. The "all" row in Table A-3 is the same as in text Table 2. The first column of Table A-3 shows the first-stage change in uninsurance rates for Full- versus Non-Expansion States, in percent, for persons aged 50-64 (the closest available age match to our main treatment sample). All first stages are small; the largest is for Hispanics at 1.5% (not significant).

In Table A-3, a number of the DD coefficients in column (2) are significant and negative, but significance disappears in the triple-difference specification except for non-Hispanic Blacks. However, as noted above, these estimates are suspect due to non-parallel pre-treatment trends and implausibly large point estimates. We are also wary of assigning too much importance to statistically significant results in particular specifications given the number of estimates we produced, although we did not conduct formal Bonferroni type p-value adjustments.

A3 Variation Based on Education Level

In Figure A-6, we show leads-and-lags graphs for the triple difference in amenable mortality for subsamples stratified on education. Low education predicts poverty and hence eligibility for Medicaid expansion; it may also affect the mortality response to the "treatment" of obtaining Medicaid. Recall that for these subsamples, we study persons aged 45-64, and the triple difference compares these persons to all persons age 65+. We present leads-and-lags graphs for elementary school only; partial high school without graduating; high-school graduate; and some college. There is no evidence of a post-expansion decline in mortality for any subgroup, including the less-than-high-school groups.

In Table A-4, we show regression results by education level. The first row shows full sample results. These differ from text Table 2 due to the broader age range that we use due to data limitations. Note that in our preferred triple-difference specification, the point estimate for overall mortality is now positive (higher mortality) and insignificant, and that Medicaid expansion predicts a significant drop in mortality for the elderly (a placebo group). Both results cast further doubt on whether an effect of Medicaid expansion on mortality can be reliably detected.

The first column shows the relevant first stages. The first stage is close to 4% for persons without a high school degree, but drops to 1.5% for high school graduates with no college, and to 1% for persons with some college**.** However, the non-high-school graduates are only 12% of the 45-64 age group, so the power gained from a stronger first stage is offset by smaller sample size.

The first row shows full sample results. The second through fifth rows show effects for the four education groups, starting with the lowest group, those with only elementary school completion, while the other rows show successively higher education categories. All DD and triple-difference point estimates are insignificant, consistent with the leads-and-lags graphs in Figure 5. The point estimate for three of the four education groups, including the least educated, are positive (opposite from predicted).

A4. Variation by Primary Cause of Death

In Table A-5, we present results by cause of death, for the top 4 causes of death: cancer, diabetes, cardiovascular causes, and respiratory illnesses, and also for HIV. Figure A-7 provides the corresponding leads-and-lags graphs. All of these causes are within the broad category of amenable mortality. First-stage estimates are not available with our data, because we lack data on Medicaid insurance takeup among those with specific diseases. However, Soni et al. (2018a, 2018b) use a DiD design based on Medicaid expansion and report a 2.4% first stage among persons with cancer diagnoses and a 6.4% increase in early-stage cancer diagnoses. Diabetics could plausibly benefit more strongly from Medicaid expansion given the negative correlation between income and diabetes prevalence and evidence from the Oregon Medicaid Experiment that gaining Medicaid insurance predicts increased diabetes diagnosis (Baicker et al., 2013). HIV is another specific condition, for which health insurance has predicted lower mortality in previous studies (Goldman et al., 2001). However, both DD and triple-difference coefficients are insignificant for all causes of death.

A5. Variation by Pre-ACA Uninsurance and Poverty Rates

We turn next to an effort to exploit pre-AC`A uninsurance rates and poverty levels. We cannot measure the second stage (mortality by individual income and insurance status) from the mortality data, so we address this source of heterogeneity indirectly at the county level. The DD specification is the same as above; the third difference for is high-versus-low pre-ACA uninsurance rates in counties. We compare "treated" high-uninsurance counties (the counties with the highest pre-ACA uninsurance rates, defined so that they together contain 20% of the U.S. population) to "control" counties with the lowest pre-ACA uninsurance rates, also containing 20% of the U.S. population; we drop all other counties. This is similar to the analysis in Finkelstein and McKnight (2008), exploiting pre-Medicare variation in insurance levels, and Courtemanche et al. (2017) for the ACA. The third difference for high-vs-low poverty counties is similar: high-poverty counties (the counties with the highest poverty rates, together containing 20% of the US. population) versus low-poverty counties (counties with the lowest poverty rates, also containing 20% of the U.S. population); we drop all other counties. These comparisons rely on all ACA-induced sources of health insurance expansion, rather than Medicaid expansion alone.

We present leads-and-lags graphs for amenable mortality in Figure A-8. Neither graph shows evidence of a treatment effect. Both graphs show signs of a pre-treatment trend toward lower mortality in the last few years prior to ACA expansion, in both high-uninsurance counties and high-poverty counties, which does *not* continue in the post-expansion period and indeed reverses for the high-uninsurance counties.

We present regression estimates in Table A-6, for the full sample and for demographic subsamples. Data are sufficient to let us compute first-stage estimates only for the full sample and for male and female subsamples. The first stage remain quite small. There is no evidence of significant effects of Medicaid expansion on mortality. For the full sample, the coefficients for both subsamples are insignificant. For the comparison of high-vs-low uninsurance counties, the coefficient is positive (opposite from predicted). For the demographic subsamples, five of the 14 coefficients are positive; and the only significant coefficient is also positive.[4]

A6. Alternative Specifications: ATT Weights; All-Non-Elderly Adults; and Total Mortality

In Tables A-7 through A-11, we present results using a number of different specifications. Table A-7 is similar to text Table 2, but uses the following alternative specifications: (i) ATT * population weights (we use population weights in the text); (ii) using linear state trends; (iii) running regressions at the state instead of the county level, with population weights); and running state-level regressions without population weights. All triple-difference coefficients are insignificant. Figure A-9 provides leads-and-lags graphs for amenable mortality with ATT * population weights.

To generate the ATT (average treatment effect on the treated) weights, we first average the covariates over the pre-treatment period (2009-2013). We then run a logit regression, which predicts whether a county is in a Full- or Non-Expansion State, using all variables in Table A-2 to generate the fitted propensities p for each county. ATT weights are calculated as (p/(1-p)).

Figure A-10 presents leads-and-lags graphs for DD and triple differences for total mortality, instead of amenable mortality. Figure A-10 presents leads-and-lags graphs for DD and triple differences for non-amenable mortality.

In Table A-8, we present triple-difference results using these same alternative specifications with each of the demographic subgroups. The significant, negative coefficient for non-Hispanic Blacks survives in several of these specifications, but loses significance in state-level regressions without population weights. All other coefficients are insignificant, except that we find a significant negative coefficient for men in state-level regressions without population weights. The sizeable differences, for several subgroups, between state-level regressions with and without population weights confirm our initial concern that results from this specification are sensitive to outlier results in a few low-population states. Figure A-12 provides leads-and-lags graphs for amenable mortality for demographic subgroups, with ATT * population weights.

In Table A-9, we present triple-difference results with these alternative specifications with each of the education subgroups. All estimated effects are statistically insignificant. Figure A-13

[4] In Table A-6, we use all counties and estimate continuous versions of the comparisons in Table 6 between high and low uninsurance (or poverty) counties, again with insignificant results.

provides leads-and-lags graphs for amenable mortality for education subgroups, with ATT * population weights.

In Table A-10, we present triple-difference results with these alternative specifications with each cause of death. All estimated effects are statistically insignificant. Figure A-14 provides leads-and-lags graphs for amenable mortality by cause of death, with ATT * population weights.

Figure A-15 presents leads-and-lags graphs for the comparison of high-versus low poverty and high-versus low-uninsurance counties, with ATT * population weights. Figure A-16 is similar, but the sample is all non-elderly adults.

In Table A-11, we present triple-difference results using two alternative specifications (ATT * population weights, and comparing all non-elderly adults to all elderly adults), for each of the demographic subgroups. There are some scattered significant coefficients, positive for women and negative for men (with ATT * population weights) and for non-Hispanic Blacks (for the broad age range), but no consistent results across specifications. Figure A-17 presents leads-and-lags graphs for the comparison of amenable mortality for all non-elderly adults.

Across all tables, the scattered significant coefficients that we find are far too large in magnitude to be true causal effects. Indeed, given our standard errors, only implausibly large coefficients would appear to be statistically significant.

Table A-1. Medicaid Expansion States (2014-2016)

This table includes Medicaid expansions through 2016. It is based on combining and reconciling the classification of states as "full expansion," "None," or in-between ("mild" or "substantial" expansion), by Simon, Cawley and Soni (2017), Lou et al. (2018), and Kaiser Family Foundation (2015). Most states could be classified based on their rules for when and to what level they expanded Medicaid for all adults. Arizona required special care; see detailed analysis below. Because our mortality data are annual, we consider New Hampshire to be a 2015 expansion, Alaska to be a 2016 expansion, and Louisiana to be a 2017 expansion, hence beyond our study period.

In the "expansion details" column, "ACA Expansion" means regular expansion to 138% of FPL, on the date stated in the "Effective Date" column. In the "inclusion/exclusion column, C = control (non-expansion), T = treatment (full expansion); other states are excluded. Simon et al. (2017) classify early expansion states as "mild" or "substantial" expansion, based on their assessment of the extent to which enrollment increase with full Affordable Care Act expansion in 2014. This classification of states based on expansion status is also used in Black et al. (2018) ("BHNS"). % change in uninsured enrollees (2013-20156) come from SAHIE estimates for ages 18-64 and considering all income groups.

State	Abbr.	Expansion Details	Effective Date	% change in uninsured enrollees (2013-2016)	Inclusion/ Exclusion	Expansion type	Compare to BHNS
Alabama	AL	None		6.4	C [.]	None	Consistent
Alaska	AK	Medicaid Expansion	09/01/2015	6.8	T [2016]	None	Consistent for 2014-2015 (expanded late 2015)
Arizona[5]	AZ	§ 1115 Waiver (100% FPL, but closed to new enrollees in 2011) ACA Expansion	2000 01/01/2014	9.6	T[2014]	Full	Consistent
Arkansas[6]	AR	§ 1115 Waiver	01/01/2014	12.4	T [2014] Private Option	Full	Consistent
California[7]	CA	§ 1115 Waiver (LA county) § 1115 Waiver (200% FPL) ACA Expansion	01/01/1995 11/01/2010 01/01/2014	13.5	Excluded (Early expansion)	Substantial	Consistent
Colorado[8]	CO	§ 1115 Waiver (to 10% of FPL)	04/01/2012	8.6	T [2016]	Full	Consistent

[5] Arizona used a § 1115 waiver to expand Medicaid coverage to childless adults up to 100% FPL during 2000-2011. In 2011, the state started to phase out that program (transitioning into Medicaid expansion). Which category Arizona belongs in was unclear based on its rules, so we also examined the extent to which Medicaid enrollment increased in 2014. See details below.

[6] Arkansas operated a limited-benefit premium-assistance program for childless adults who worked for small uninsured employers (ARHealthNetworks waiver) prior to the ACA. Arkansas's Medicaid expansion includes a "private option" under which Medicaid-eligible persons receive health insurance from the state insurance exchange, with a small monthly premium.

[7] California expanded Medicaid in 2010-2011, in selected counties.

State	Abbr.	Expansion Details	Effective Date	% change in uninsured enrollees (2013-2016)	Inclusion/ Exclusion	Expansion type	Compare to BHNS
		ACA Expansion	01/01/2014		T [2014]		
Connecticut[9]	CT	State Plan Amendment (56% FPL) ACA Expansion	04/01/2010 01/01/2014	6.4	Excluded (Early Expansion)	Substantial	Consistent
Delaware[10]	DE	ACA Expansion	01/01/1996 01/01/2014	5.1	Excluded (Early Expansion)	Mild	Consistent
District of Columbia[11]	DC	State Plan Amendment (133% FPL) § 1115 Waiver ACA Expansion	07/01/2010 12/01/2010 01/01/2014	4.2	Excluded (Early expansion)	Mild	Consistent
Florida	FL	None		10.4	C [.]	None	Consistent
Georgia	GA	None		7.6	C [.]	None	Consistent
Hawaii[12]	HI	ACA Expansion	08/01/1994 01/01/2014	4.6	Excluded (Early expansion)	Substantial	Consistent
Idaho	ID	None		8.2	C [.]	None	Consistent
Illinois	IL	ACA Expansion	01/01/2014	9.2	T [2014]	Full	Consistent
Indiana	IN	§ 1115 Waiver	02/01/2015	8.5	T [2015]	Full	Consistent
Iowa[13]	IA	§ 1115 Waiver	01/01/2014	5.8	T [2014]	Full	Consistent
Kansas	KS	None		5.2	C [.]	None	Consistent
Kentucky	KY	ACA Expansion	01/01/2014	13.7	T [2014]	Full	Consistent
Louisiana	LA	ACA Expansion	07/01/2016	9.0	C [.]	None	Consistent
Maine	ME	None		4.2	C [.]	None	Consistent
Maryland	MD	ACA Expansion	01/01/2014	5.8	T [2014]	Full	Consistent

[9] Connecticut, elected to enact the Medicaid expansion in 2010 through a state amended plan at 56%. Connecticut expanded its Medicaid program fully in 2014.

[10] In Delaware, childless adults with incomes up to 100% FPL were eligible for Medicaid through the Diamond State Health Plan waiver, effective on 01/01/1996.

[11] DC expanded its Medicaid program at 133% of FPL in 2010.

[12] In Hawaii, childless adults with incomes up to 100% FPL were eligible for the state's QUEST Medicaid managed care waiver program, effective on 08/01/1994.

[13] Under the IowaCare program, childless adults with income below 200% FPL were eligible for health insurance since 2005. However, IowaCare provided limited services in a limited network, so low-income adults in Iowa received a substantial coverage expansion in 2014 (Damiano et al., 2013). During 2014-2015, Iowa residents with income < 100% of FPL were enrolled in Medicaid managed care plans, while those with income of 100-138% of FPL received private insurance obtained through the Iowa health exchange, with premiums waived (a partial "private option"). See https://www.medicaid.gov/Medicaid-CHIP-Program-Information/By-Topics/Waivers/1115/downloads/ia/Market-Place-Choice-Plan/ia-marketplace-choice-plan-state-term-app-06012016.pdf...

State	Abbr.	Expansion Details	Effective Date	% change in uninsured enrollees (2013-2016)	Inclusion/ Exclusion	Expansion type	Compare to BHNS
Massachusetts[14]	MA	"Romneycare" ACA Expansion	04/12/2006 01/01/2014	1.7	Excluded	Mild	Consistent
Michigan	MI	ACA Expansion	04/01/2014	8.5	T [2014]	Full	Consistent
Minnesota[15]	MN	State Plan Amendment (75% FPL) § 1115 Waiver (200% FPL) ACA Expansion	03/01/2010 08/01/2010 01/01/2014	5.6	Excluded (Early Expansion)	Substantial	Consistent
Mississippi	MS	None		7.3	C [.]	None	Consistent
Missouri	MO	§ 1115 Waiver (St. Louis County Only) (200% FPL) None	07/01/2012	5.7	C [.]	None	Consistent
Montana	MT	ACA Expansion	01/01/2016	11.5	T [2016]	None	Consistent for 2014-2015 (expanded in 2016)
Nebraska	NE	None		4.1	C [.]	None	Consistent
Nevada	NV	ACA Expansion	01/01/2014	11.2	T [2014]	Full	Consistent
New Hampshire[16]	NH	§ 1115 Waiver	08/15/2014	7.0	T [2015]	Full	Consistent
New Jersey[17]	NJ	§ 1115 Waiver (23% FPL) ACA Expansion	04/01/2011 01/01/2014	7.4	T [2014]	Full	Consistent
New Mexico	NM	ACA Expansion	01/01/2014	13.8	T [2014]	Full	Consistent
New York[18]	NY	§ 1115 waiver ACA Expansion	10/01/2001 01/01/2014	6.7	Excluded (Early expansion)	Mild	Consistent
North Carolina	NC	None		7.4	C [.]	None	Consistent

[14] Massachusetts implemented reforms to expand insurance coverage to low-income adults in 2006.

[15] Minnesota conducted early expansion in 2010 two ways. Persons with income $\leq$ 75%FPL were insured through Medical Assistance Medicaid, funded through a State Plan Amendment, persons with income from 75~200% of FPL were insured through MinnesotaCare, funded through a § 1115 Waiver, which had limited benefits and cost-sharing.

[16] New Hampshire implemented a "private option" (mandatory purchase of subsidized private insurance, instead traditional Medicaid, in 2016. See https://www.medicaid.gov/Medicaid-CHIP-Program-Information/By-Topics/Waivers/1115/downloads/nh/health-protection-program/nh-health-protection-program-premium-assistance-appvl-amend-req-06232015.pdf.

[17] New Jersey's expansion in 2011 only extended to 23% FPL; we therefore treated it as a full expansion state.

[18] In New York, childless adults up to 78% FPL were eligible for the Medicaid (Home Relief) waiver program and childless adults up to 100% FPL were eligible for the Family Health Plus waiver program (Heberlein et al., 2011).

State	Abbr.	Expansion Details	Effective Date	% change in uninsured enrollees (2013-2016)	Inclusion/ Exclusion	Expansion type	Compare to BHNS
North Dakota	ND	ACA Expansion	01/01/2014	6.0	T [2014]	Full	Consistent
Ohio	OH	ACA Expansion	01/01/2014	8.1	T [2014]	Full	Consistent
Oklahoma	OK	None		5.3	C [.]	None	Consistent
Oregon	OR[19]	ACA Expansion	01/01/2014	12.2	T [2014]	Full	Consistent
Pennsylvania	PA	ACA Expansion	01/01/2015	6.2	T [2015]	Full	Consistent
Rhode Island	RI	ACA Expansion	01/01/2014	10.5	T [2014]	Full	Consistent
South Carolina	SC	None		8.1	C [.]	None	Consistent
South Dakota	SD	None		2.9	C [.]	None	Consistent
Tennessee	TN	None		6.8	C [.]	None	Consistent
Texas	TX	None		7.5	C [.]	None	Consistent
Utah	UT	None		6.9	C [.]	None	Consistent
Vermont	VT[20]	§ 1115 Waiver ACA Expansion	01/01/1996 01/01/2014	4.7	Excluded (Early expansion)	Mild	Consistent
Virginia	VA	None		5.3	C [.]	None	Consistent
Washington[21]	WA	§ 1115 Waiver (133% FPL) ACA Expansion	01/03/2011 01/01/2014	11.1	T [2014]	Full	Consistent
West Virginia	WV	ACA Expansion	01/01/2014	12.8	T [2014]	Full	Consistent
Wisconsin[22]	WI	New eligibility for BadgerCare but not ACA Expansion	2009	5.5	Excluded	Substantial	Consistent
Wyoming	WY	None		3.6	C [.]	None	Consistent

[19] In 2008, Oregon enacted a small Medicaid expansion for low-income adults through a lottery among applicants. However, less than one-third of the 90,000 people on the waitlist were selected to apply for Medicaid in 2008 (Baicker et al., 2013), some of the denied applicants were then enrolled in 2010. We treat Oregon as full expansion due to the small size of this earlier expansion.

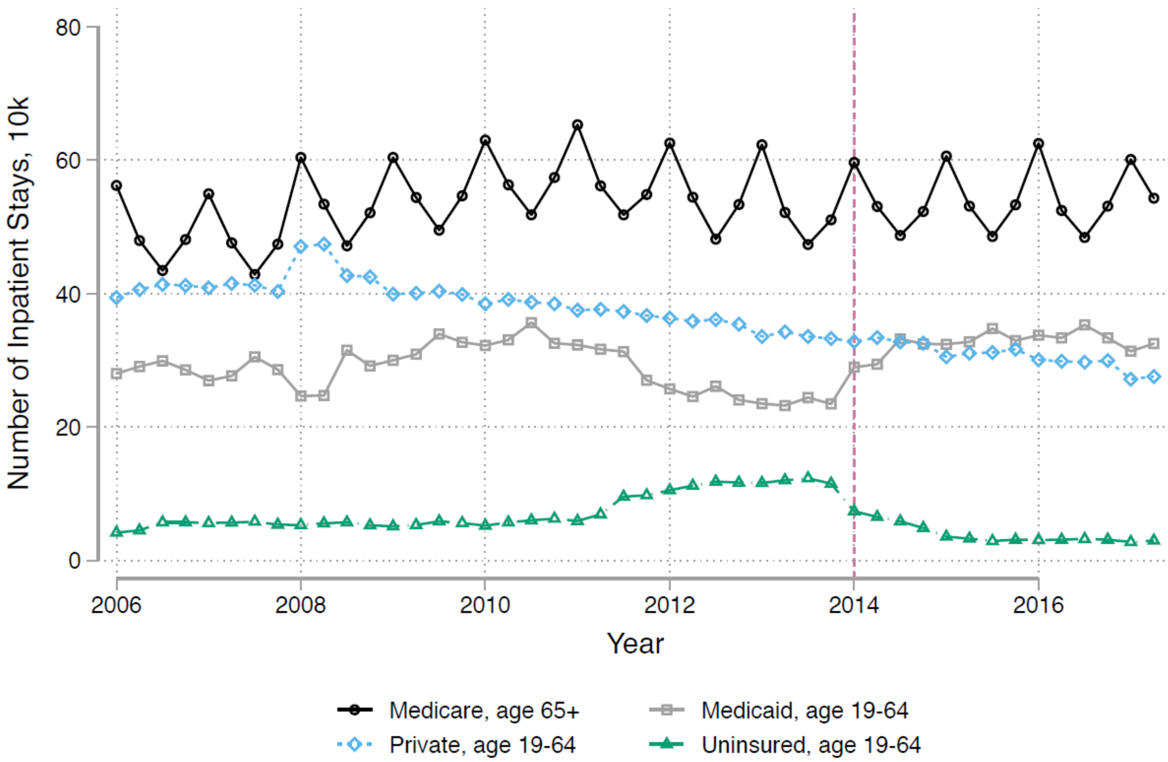
[20] In Vermont, childless adults up to 150% FPL were eligible for Medicaid equivalent coverage through the Vermont Health Access Plan waiver program (Heberlein et al., 2011). Vermont Health Access Plan (Sec. 1115 waiver) was approved in 1995 and effective in 1996.

[21] Washington's early expansion was limited to prior state plan enrollees (Sommers et al., 2013).

[22] Wisconsin received federal approval to offer Medicaid to childless adults below 100% FPL through the BadgerCare program as of 2009 (Gates & Rudowitz, 2014); it did not formally adopt ACA expansion in 2014 and kept the income threshold at 100% FPL.

Arizona Details for Table A-1

Arizona had a S.1931 program providing Medicaid up to 106% FPL for parents. It also had a limited program for childless adults, under a § 1115 waiver, starting in 2001, which was closed to new entrants since 2011.[23] Whether to treat Arizona as a full expansion state or an early expansion state turns on how many childless adults were still covered at the ACA onset in 2014, given churn in eligibility. The tail off in hospital admissions with Medicaid payment, and jump at the start of 2014 (with uninsured admissions showing the opposite pattern), persuades us that Arizona should be treated as a regular expansion state.



Source: Author reproduction of HCUP figure using HCUP Fast Stats at https://www.hcup-us.ahrq.gov/faststats/StatePayerServlet?state1=AZ.

[23] Source: https://www.kff.org/medicaid/fact-sheet/proposed-changes-to-medicaid-expansion-in-arizona/.

References for Table A-1

Black, B., Hollingsworth, A., Nunes L., & Simon, K., *The Effect of Health Insurance on Mortality: What Can We Learn from the Affordable Care Act Coverage Expansions* (working paper 2018).

Damiano, P., Bentler, S.E., Momany, E.T., Park, K.H., & Robinson, E. (2013). Evaluation of the IowaCare program: Information about the medical home expansion. The University of Iowa Public Policy Center, at: https://ir.uiowa.edu/cgi/viewcontent.cgi?article=1074&context=ppc_health

Gates, A., & Rudowitz, R. (2014). Wisconsin's BadgerCare program and the ACA. Kaiser Family Foundation. At: https://www.kff.org/medicaid/fact-sheet/wisconsins-badgercare-program-and-the-aca/.

Heberlein, M., Brooks, T., Guyer, J., Artiga, S., & Stephens, J. (2011). Holding steady, looking ahead: Annual findings of a 50-state survey of eligibility rules, enrollment and renewal procedures, and cost-sharing practices in Medicaid and CHIP, 2010–2011. Kaiser Commission on Medicaid and the Uninsured.

Sommers, B. D., Buchmueller, T., Decker, S. L., Carey, C., & Kronick, R. (2013). The Affordable Care Act has led to significant gains in health insurance and access to care for young adults. *Health Affairs* 32: 165-174.

Kaiser Family Foundation. (2018). Status of State Action on the Medicaid Expansion Decision, at https://www.kff.org/health-reform/state-indicator/state-activity-around-expanding-medicaid-under-the-affordable-care act/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D.

Kaiser Family Foundation. (2012). Status of State Action on the Medicaid Expansion Decision. How is the Affordable Care Act Leading to Changes in Medicaid Today? State Adoption of Five New Options, at https://kaiserfamilyfoundation.files.wordpress.com/2013/01/8312.pdf.

Lou, Q., Moghtaderi, A., Markus, A., Dor, A. (2018), Impact of Medicaid Expansion on Total Revenue of Community Health Centers by Funding Sources. Working paper.

Simon, Kosali, Aparna Soni and John Cawley (2017), The Impact of Health Insurance on Preventive Care and Health Behaviors: Evidence from the First Two Years of the ACA Medicaid Expansions, *Journal of Policy Analysis and Management* 36: 390-417.

Wishner, JB., Holohan, J., Upadhyay, D., McGrath, M. (2015). Medicaid Expansion, the Private Option, and Personal Responsibility Requirements: The Use of § 1115 Waivers to Implement Medicaid Expansion Under the ACA. Urban Institute, at https://www.urban.org/sites/default/files/publication/53236/2000235-Medicaid-Expansion-The-Private-Option-and-Personal-Responsibility-Requirements.pdf.

Table A-2. Covariate Balance for Full-Expansion and Non-Expansion States

Table shows summary statistics for county-level covariates and mortality for Full-Expansion and Non-Expansion states during pre-expansion period (means over 2009-2013), using county population weights. *t*-statistics use two-sample *t*-test for difference and robust standard errors with state clusters. Normalized difference is a sample-size independent measure of the difference between two means, scaled by standard deviation): $ND_j = (\bar{x}_{jt} - \bar{x}_{jc}) / [(s_{jt}^2 + s_{jc}^2)/2]^{1/2}$. State groups are defined in Table A-1. Mortality rates are per 100,000 persons. Dollar amounts are in 2010 $.

	Full-Expansion States (1)	Non-Expansion States (2)	Difference t-stat (3)	Normalized Difference (4)
% age 0-19	23.36	24.35	1.11	-0.30
% age 18-34	22.74	23.42	1.40	-0.15
% age 35-44	12.94	13.11	0.71	-0.11
% age 45-54	14.53	13.98	2.32	0.40
% age 55-64	12.56	11.81	2.05	0.36
% age 65-74	7.56	7.48	0.16	0.04
% age 75-84	4.38	4.19	0.52	0.13
% age 85+	1.94	1.66	1.53	0.32
% Male	49.21	49.13	0.47	0.04
% White	82.91	77.43	2.19	0.36
% Black	11.42	18.16	2.61	-0.49
% Other Races	5.67	4.41	1.35	0.15
% Hispanic	11.44	16.33	0.87	-0.38
% In Poverty	14.67	16.89	2.75	-0.36
% Managed Care Penetration	24.55	22.99	0.42	0.15
% Disabled (ages 18-64)	16.31	17.57	1.29	-0.20
Mean Per Capita Income	40,208	37,537	1.72	0.31
Median Household Income	51,691	47,122	1.81	0.44
Unemployment Rate, 16+	8.84	8.28	1.12	0.20
% with Diabetes	8.85	9.72	2.45	-0.46
% Physically Inactive	22.89	24.70	1.85	-0.40
% Obese	27.95	29.11	1.16	-0.28
% Smoker	21.96	21.71	0.27	0.06
Physicians/1,000 people	3.10	2.65	2.88	0.27
% Uninsured (ages 18-64)	18.68	24.96	3.36	-1.09
Amenable Mortality (all ages)	510.52	481.21	0.90	0.18
Amenable Mortality (ages 55-64)	575.22	623.78	1.86	-0.24
Non-amenable Mortality (all ages)	345.28	341.33	0.20	0.04
Non-amenable Mortality (ages 55-64)	278.85	309.76	2.50	-0.30

Table A-3: DD and Triple-Difference Estimates: Different Demographic Groups (ages 55-64)

First column shows annual averages over 2009-2016 for number of deaths and population in millions. Of the full sample (28.8M people), 14.5M were in expansion states. Second column shows mortality rate for persons aged 55-64 for indicated groups. Third column shows first-stage DD estimates of change in uninsurance rates (in percent) from 2013 to 2016 for indicated demographic subsamples, for persons aged 50-64, from regression of percent uninsurance on Full Expansion dummy, with state and year FE and state population weights, using state-level SAHIE data (best available), and same covariates as the DD and triple difference regressions. Remaining columns show coefficients from DD or triple difference regressions on Full-Expansion dummy or, for triple difference column, full-expansion dummy * age 55-64 dummy, from county-level regressions with county-and year FE and population weights, similar to Table 2, for *ln*((amenable mortality/100,000 persons)+1) over 2009-2015. Standard errors use state clusters. *.**, *** indicates statistical significance at the 10%, 5%, and 1% levels, respectively; significant results at 5% level or better in **boldface**.

Demographic Subsamples	**Ann. Deaths (Pop. in M) (1)**	**Mortality rate (2)**	**First stage (%) 50-64 yrs (3)**	**DiD 55-64 yrs (4)**	**DiD 65-74 yrs (5)**	**Triple diff. (6)**
All Amenable	174,379	605.3	**1.113****	**-0.018****	-0.008	-0.004
	(28.8)		**(0.452)**	**(0.008)**	(0.006)	(0.008)
Male	105,465	759.8	0.692	-0.018*	-0.004	-0.004
	(13.9)		(0.747)	(0.010)	(0.008)	(0.010)
Female	68,914	461.7	0.936	**-0.020****	-0.016*	0.004
	(14.9)		(0.705)	**(0.009)**	(0.009)	(0.012)
White (Not Hispanic)	129,542	589.8	**1.130****	-0.015*	-0.011*	-0.003
	(22.0)		**(0.490)**	(0.008)	(0.007)	(0.009)
Black (Not Hispanic)	32,217	917.0	0.994	-0.031*	0.020	**-0.055*****
	(3.5)		(0.852)	(0.016)	(0.015)	**(0.017)**
Other	3,619	321.6	-	-0.050	-0.039	-0.035
	(1.1)		-	(0.060)	(0.052)	(0.078)
Hispanic	9,086	398.2	1.484	**-0.161*****	-0.092	-0.055*
	(2.3)		(1.228)	**(0.057)**	(0.057)	(0.029)
Not Hispanic	165,293	623.1	-	**-0.018****	-0.008	-0.005
	(26.5)		-	**(0.008)**	(0.006)	(0.007)
Pop. Weights			Yes	Yes	Yes	Yes
Covariates			Yes	Yes	Yes	Yes

Table A-4: DD and Triple-Difference Estimates: by Educational Attainment (ages 45-64)

First column shows annual averages over 2009-2016 for number of deaths and population in millions. Second column shows mortality rate for persons aged 55-64 for indicated groups. Third column shows first-stage DD estimates of change in uninsurance rates (in percent) from 2013 to 2016 for indicated education-levels, for persons aged 45-64, from regression of percent uninsurance on Full Expansion dummy, with state and year FE and state population weights. Remaining columns show coefficients from DD or triple difference regressions on Full-Expansion dummy or, for triple difference column, full-expansion dummy * age 45-64 dummy, from county-level regressions with county and year FE and population weights, similar to Table 2, for *ln*((amenable mortality/100,000 persons)+1) among persons with indicated education levels, over 2009-2015. Standard errors use state clusters. *.**, *** indicates statistical significance at the 10%, 5%, and 1% levels, respectively; significant results at 5% level or better in **boldface**.

Education Subsample	Ann. Deaths (Pop. in M) (1)	Mortality Rate (2)	First stage (%) 45-64 yrs (3)	DiD 45-64 yrs (4)	DiD 65+ yrs (5)	Triple diff. (6)
All Amenable	252,285	422.1	1.048	-0.012	**-0.020*** **	0.014
	(59.77)		(0.738)	(0.008)	**(0.006)**	(0.009)
Elementary School	14,776	565.4	3.747	0.047	0.014	0.066
	(2.61)		(2.530)	(0.046)	(0.058)	(0.048)
High School Incomplete	33,698	768.6	**3.912*** **	-0.009	-0.003	-0.011
	(4.38)		**(1.449)**	(0.061)	(0.064)	(0.036)
High School Complete	110,019	607.2	1.533	-0.021	-0.032	0.010
	(18.12)		(0.939)	(0.040)	(0.037)	(0.014)
Some College	86,793	250.5	0.468	-0.015	-0.026	0.013
	(34.65)		(0.572)	(0.035)	(0.031)	(0.011)
Population Weights			Yes	Yes	Yes	Yes
Covariates			Yes	Yes	Yes	Yes

Table A-5: DD and Triple-Difference Estimates: by Cause of Death (age 55-64)

First column shows annual averages over 2009-2016 for number of deaths and population in millions. Second column shows mortality rate for persons aged 55-64 for indicated groups. Remaining columns show coefficients from DD or triple difference regressions on Full-Expansion dummy or, for triple difference column, full-expansion dummy * age 45-64 dummy, from county-level regressions with county and year FE and population weights, similar to Table 2, for *ln*((amenable mortality/100,000 persons)+1) among persons with indicated primary cause of death, over 2009-2016. Standard errors use state clusters. *.**, *** indicates statistical significance at the 10%, 5%, and 1% levels, respectively; significant results at 5% level or better in **boldface**.

By Cause of Death	**deaths (pop. In M)**	**DiD 55-64 yrs**	**DiD 65-74 yrs**	**Triple diff.**
	(1)	(2)	(3)	(4)
All Amenable	174,379	**-0.018****	-0.008	-0.004
	(28.81)	**(0.008)**	(0.006)	(0.008)
Cancer	87,170	-0.003	0.003	-0.004
	(28.81)	(0.006)	(0.006)	(0.009)
Diabetes	14,394	-0.024	0.001	-0.007
	(28.81)	(0.019)	(0.025)	(0.020)
Cardiovascular	70,677	-0.010	-0.009	0.006
	(28.81)	(0.010)	(0.010)	(0.010)
Respiratory	16,442	-0.030	-0.017	-0.010
	(28.81)	(0.020)	(0.013)	(0.023)
HIV	1,282	-0.058	0.005	-0.051
	(28.81)	(0.037)	(0.038)	(0.060)
Pop. Weights		Yes	Yes	Yes
Covariates		Yes	Yes	Yes

Table A-6: Triple Difference Estimates: Separating Counties by Baseline Health Uninsurance or Poverty Levels (age 55-64)

First column shows annual averages over 2009-2016 for number of deaths and population aged 55-64 in millions, for sample of high-versus low- uninsurance counties. Second and fourth columns column shows full-sample and by gender first stages; we lack the data to compute first stages for the other subsamples. Remaining columns show coefficients from triple difference, county-level regressions with county and year FE and population weights, similar to Table 2, over 2009-2016, for amenable mortality for full sample and indicated subsamples. Third difference in column (3) is between the counties with the highest uninsurance rate in 2013, containing 20% of the U.S. population, and the counties with the lowest uninsurance rate in 2013, containing 20% of the U.S. population. Third difference in column (5) is similar but is between the counties with lowest versus highest poverty rates in 2013. Standard errors use state clusters. *.**, *** indicates statistical significance at the 10%, 5%, and 1% levels, respectively; significant results at 5% level or better in **boldface**.

Sample	**Deaths (pop. in M) (1)**	**First Stage (%) 50-64 yrs (2)**	**Triple diff. Uninsurance (3)**	**First Stage (%) 50-64 yrs (4)**	**Triple diff. Poverty (5)**
All	66,329	1.221	0.003	0.720	0.000
	(11.9)	(0.653)	(0.020)	(0.789)	(0.013)
Male	40,750	0.593	-0.020	0.408	-0.024
	(5.8)	(0.657)	(0.028)	(0.721)	(0.018)
Female	26,103	**1.829*** **	0.050*	0.912	**0.037*** **
	(6.1)	**(0.679)**	(0.028)	(0.791)	**(0.014)**
White (Not Hispanic)	51,198		-0.017		-0.015
	(9.1)		(0.018)		(0.010)
Black (Not Hispanic)	11,970		-0.001		-0.073*
	(1.4)		(0.059)		(0.040)
Other	1,496		-0.083		-0.005
	(0.4)		(0.137)		(0.107)
Hispanic	3,421		0.279		0.082
	(0.9)		(0.267)		(0.103)
Not Hispanic	60,879		0.003		-0.005
	(10.4)		(0.021)		(0.015)
Pop. Weights		Yes	Yes	Yes	Yes
Covariates		Yes	Yes	Yes	Yes

Table A-7. Estimated Effect of Medicaid Expansion on Amenable Mortality: Different Specifications

Table 2 in the text shows DD and triple-difference estimates for county-level regressions, with county and year FE and population weights, of ln[(amenable mortality/100,000 persons)+1] over 2009-2016 on full-expansion dummy (=1 for Full-Expansion States in expansion years; 0 otherwise), and covariates. Third difference is ages 55-64 versus ages 65-74. This table provides results for principal coefficients of interest, from regressions in which we vary this specification as follows: Panel A reproduces our results from text Table 2; Panel B uses ATT*population weights instead of only population weights; Panel C adds linear state trends; Panel D reports results from regressions at state- instead of county-level (with population weights); and Panel E reports results from state-level regressions without weights. Standard errors use state clusters. *, **, *** indicates statistical significance at the 10%, 5%, and 1% levels, respectively; significant results at 5% level or better in **boldface**.

	DiD 55-64 years		**Triple diff.**	
	(1)	**(2)**	**(3)**	**(4)**
Panel A. Main Specification (from text Table 2)				
Full Expansion Dummy	-0.018*	**-0.018****		
	(0.010)	**(0.008)**		
Full Expansion Dummy x Age 55-64 Dummy			-0.002	-0.004
			(0.009)	(0.008)
Panel B. With ATT x Population Weights				
Full Expansion Dummy	-0.014	-0.015*		
	(0.013)	(0.009)		
Full Expansion Dummy x Age 55-64 Dummy			-0.014	-0.013
			(0.009)	(0.012)
Panel C. With Linear State Trends				
Full Expansion Dummy	-0.006	-0.009		
	(0.008)	(0.008)		
Full Expansion Dummy x Age 55-64 Dummy			-0.001	-0.003
			(0.009)	(0.008)
Panel D. State-Level (with Pop Weights)				
Full Expansion Dummy	**-0.020****	-0.011*		
	(0.009)	(0.007)		
Full Expansion Dummy x Age 55-64 Dummy			-0.006	-0.009
			(0.008)	(0.010)
Panel E. State-Level (No Weights) Specification				
Full Expansion Dummy	**-0.018****	-0.009		
	(0.009)	(0.009)		
Full Expansion Dummy x Age 55-64 Dummy			-0.009	-0.015
			(0.010)	(0.011)
Covariates	No	Yes	No	Yes

Table A-8: Triple-Difference Estimates by Demographic Group: Different Specifications

Table 3 in the text shows DD and triple-difference estimates for different demographic groups, from county-level regressions, with county and year FE and population weights, of ln[(amenable mortality/100,000 persons)+1] over 2009-2016 on full-expansion dummy (=1 for Full-Expansion States in expansion years; 0 otherwise), and covariates. Third difference is ages 55-64 versus ages 65-74. This table provides triple difference results for principal coefficients of interest, from regressions in which we vary this specification as follows: using ATT*population weights; adding linear state trends; and running regressions at state- instead of county-level, with and without population weights. Standard errors use state clusters. *, **, *** indicates statistical significance at the 10%, 5%, and 1% levels, respectively; significant results at 5% level or better in **boldface**.

	Triple Difference Results				
Subsamples	**Main Specification (1)**	**ATT x Pop weights (2)**	**with Linear State Trends (3)**	**State-Level w. pop. weights (4)**	**State-Level unweighted (5)**
All Amenable	-0.004	-0.013	-0.003	-0.009	-0.015
	(0.008)	(0.012)	(0.008)	(0.010)	(0.011)
Male	-0.004	-0.003	-0.003	-0.015	**-0.034****
	(0.010)	(0.022)	(0.010)	(0.012)	**(0.014)**
Female	0.004	-0.022	0.005	0.006	-0.003
	(0.012)	(0.015)	(0.011)	(0.013)	(0.014)
White (Not Hispanic)	-0.003	-0.010	-0.002	-0.015	-0.002
	(0.009)	(0.010)	(0.009)	(0.011)	(0.011)
Black (Not Hispanic)	**-0.055*****	**-0.345****	**-0.055*****	**-0.040*****	0.010
	(0.017)	**(0.172)**	**(0.017)**	**(0.014)**	(0.124)
Other	-0.035	0.168	-0.036	-0.056	-0.038
	(0.078)	(0.269)	(0.078)	(0.036)	(0.059)
Hispanic	-0.055*	-0.153	-0.050*	-0.016	0.054
	(0.029)	(0.153)	(0.027)	(0.023)	(0.156)
Not Hispanic	-0.005	-0.013	-0.004	-0.010	-0.012
	(0.007)	(0.012)	(0.007)	(0.008)	(0.011)
Weights	Pop	ATT x Pop	Pop	Pop	No
Covariates	Yes	Yes	Yes	Yes	Yes

Table A-9: Triple-Difference Estimates by Educational Attainment (ages 45-64) - Different Specifications

Table 4 in the text shows DD and triple-difference estimates for groups with different education levels, from county-level regressions, with county and year FE and population weights, of ln[(amenable mortality/100,000 persons)+1] over 2009-2016 on full-expansion dummy (=1 for Full-Expansion States in expansion years; 0 otherwise), and covariates. Third difference is ages 55-64 versus ages 65-74. This table provides triple difference results for principal coefficients of interest, from regressions in which we vary this specification as follows: using ATT*population weights; adding linear state trends; and running regressions at state- instead of county-level, with and without population weights. Standard errors use state clusters. *, **, *** indicates statistical significance at the 10%, 5%, and 1% levels, respectively; significant results at 5% level or better in **boldface**.

	Triple Difference Results				
Education Subsamples	**Main Specification (1)**	**ATT x Pop weights (2)**	**with Linear State Trends (3)**	**State-Level w. pop. weights (4)**	**State-Level unweighted (5)**
All Amenable	0.014	-0.001	0.014	0.007	0.007
	(0.009)	(0.011)	(0.009)	(0.009)	(0.010)
Elementary School	0.066	0.129*	0.045	0.045	0.062
	(0.048)	(0.068)	(0.046)	(0.031)	(0.040)
High School Incomplete	-0.011	-0.015	0.004	-0.023	-0.039
	(0.036)	(0.031)	(0.036)	(0.031)	(0.035)
High School Complete	0.010	-0.001	0.023	0.005	-0.017
	(0.014)	(0.021)	(0.019)	(0.013)	(0.023)
Some College	0.013	0.011	0.031*	0.011	0.005
	(0.011)	(0.018)	(0.017)	(0.013)	(0.023)
Weights	Pop	ATT x Pop	Pop	Pop	No
Covariates	Yes	Yes	Yes	Yes	Yes

Table A-10: Triple-Difference Estimates by Cause of Death (ages 55-64): Different Specifications

Table 5 in the text shows DD and triple-difference estimates for different causes of death, from county-level regressions, with county and year FE and population weights, of ln[(amenable mortality/100,000 persons)+1] over 2009-2016 on full-expansion dummy (=1 for Full-Expansion States in expansion years; 0 otherwise), and covariates. Third difference is ages 55-64 versus ages 65-74. This table provides triple difference results for principal coefficients of interest, from regressions in which we vary this specification as follows: using ATT*population weights; adding linear state trends; and running regressions at state- instead of county-level, with and without population weights. Standard errors use state clusters. *, **, *** indicates statistical significance at the 10%, 5%, and 1% levels, respectively; significant results at 5% level or better in **boldface**.

	Triple Difference Results				
Cause of Death	**Main Specification (1)**	**ATT x Pop weights (2)**	**with Linear State Trends (3)**	**State-Level w. pop. weights (4)**	**State-Level unweighted (5)**
Amenable	-0.004	-0.013	-0.003	-0.009	-0.015
	(0.008)	(0.012)	(0.008)	(0.010)	(0.011)
Non-Amenable	-0.006	-0.008	-0.006	-0.006	-0.005
	(0.012)	(0.017)	(0.012)	(0.012)	(0.012)
Cancer	-0.004	-0.017	-0.004	-0.006	-0.001
	(0.009)	(0.011)	(0.008)	(0.010)	(0.011)
Diabetes	-0.007	-0.034	-0.005	-0.016	0.018
	(0.020)	(0.025)	(0.020)	(0.016)	(0.030)
Cardiovascular	0.006	-0.005	0.007	-0.002	-0.022
	(0.010)	(0.016)	(0.010)	(0.011)	(0.016)
Respiratory	-0.010	0.003	-0.009	-0.013	-0.023
	(0.023)	(0.035)	(0.022)	(0.016)	(0.026)
HIV	-0.051	-0.022	-0.051	-0.030	0.112
	(0.060)	(0.078)	(0.060)	(0.058)	(0.112)
Weights	Pop	Att x Pop	Pop	Pop	No
Covariates	Yes	Yes	Yes	Yes	Yes

Table A-11: Triple Difference Estimates: Counties with high-vs-low Baseline Health Uninsurance and Poverty Levels: Different Specifications

Table 6 in the text shows DD and triple-difference estimates for high-vs-low pre-ACA uninsurance and high-vs-low poverty counties, from county-level regressions, with county and year FE and population weights, of ln[(amenable mortality/100,000 persons)+1] over 2009-2016 on full-expansion dummy (=1 for Full-Expansion States in expansion years; 0 otherwise), and covariates. Third difference is ages 55-64 versus ages 65-74. This table provides triple difference results for principal coefficients of interest, from regressions in which we vary this specification as follows: using ATT*population weights; and comparing all non-elderly adults (ages 18-64) to all elderly (age 65+). Standard errors use state clusters. *, **, *** indicates statistical significance at the 10%, 5%, and 1% levels, respectively; significant results at 5% level or better in **boldface**.

	Triple Difference Results					
	Main Specification		**ATT x Pop Weights**		**Age 18-64 vs. 65+**	
Subsamples	**Unins.**	**Poverty**	**Unins.**	**Poverty**	**Unins.**	**Poverty**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
All Amenable	0.003	0.000	-0.023	-0.018	0.004	0.012
	(0.020)	(0.013)	(0.025)	(0.016)	(0.014)	(0.012)
Male	-0.020	-0.024	-0.045	**-0.046****	-0.025	-0.004
	(0.028)	(0.018)	(0.038)	**(0.020)**	(0.017)	(0.016)
Female	0.050*	**0.037*****	0.024	0.025	**0.054*****	**0.034****
	(0.028)	**(0.014)**	(0.036)	(0.039)	**(0.020)**	**(0.013)**
White (Not Hispanic)	-0.017	-0.015	**-0.053****	**-0.030*****	-0.027*	0.002
	(0.018)	(0.010)	**(0.024)**	**(0.010)**	(0.014)	(0.011)
Black (Not Hispanic)	-0.001	-0.073*	0.393	-0.303	-0.004	**-0.083*****
	(0.059)	(0.040)	(0.365)	(0.385)	(0.038)	**(0.032)**
Other	-0.083	-0.005	-0.354	-0.614	-0.057	0.060
	(0.137)	(0.107)	(0.411)	(0.512)	(0.079)	(0.074)
Hispanic	0.279	0.082	0.369	-0.004	0.056	-0.002
	(0.267)	(0.103)	(0.286)	(0.175)	(0.068)	(0.044)
Not Hispanic	0.003	-0.005	-0.028	-0.019	0.005	0.010
	(0.021)	(0.015)	(0.030)	(0.017)	(0.018)	(0.013)
Weights	Pop	Pop	Att x Pop	Att x Pop	Pop	Pop
Covariates	Yes	Yes	Yes	Yes	Yes	Yes

Table A-12: Synthetic Control Method: Weights on Donor States

Table shows the weights assigned to the Non-Expansion States (donor states)by the regular synthetic control method, used in text Figure 3.

Non-Expansion States	Synthetic Control Weights
Alabama	0
Florida	0.123
Georgia	0
Idaho	0
Kansas	0
Louisiana	0
Maine	0.038
Mississippi	0
Missouri	0.411
Nebraska	0
North Carolina	0
Oklahoma	0
South Carolina	0
South Dakota	0
Tennessee	0
Texas	0.023
Utah	0.041
Virginia	0.272
Wyoming	0.091

Figure A-1. Time Trends in Amenable Mortality for Persons Aged 18-64

Figure shows amenable mortality rate for persons age 18-64 for Full-Expansion, Substantial Expansion, Mild Expansion, and Non-Expansion States, over 1999-2016, using county population weights. State groups are defined in Table 1. Dashed vertical line separate pre-expansion from expansion period.

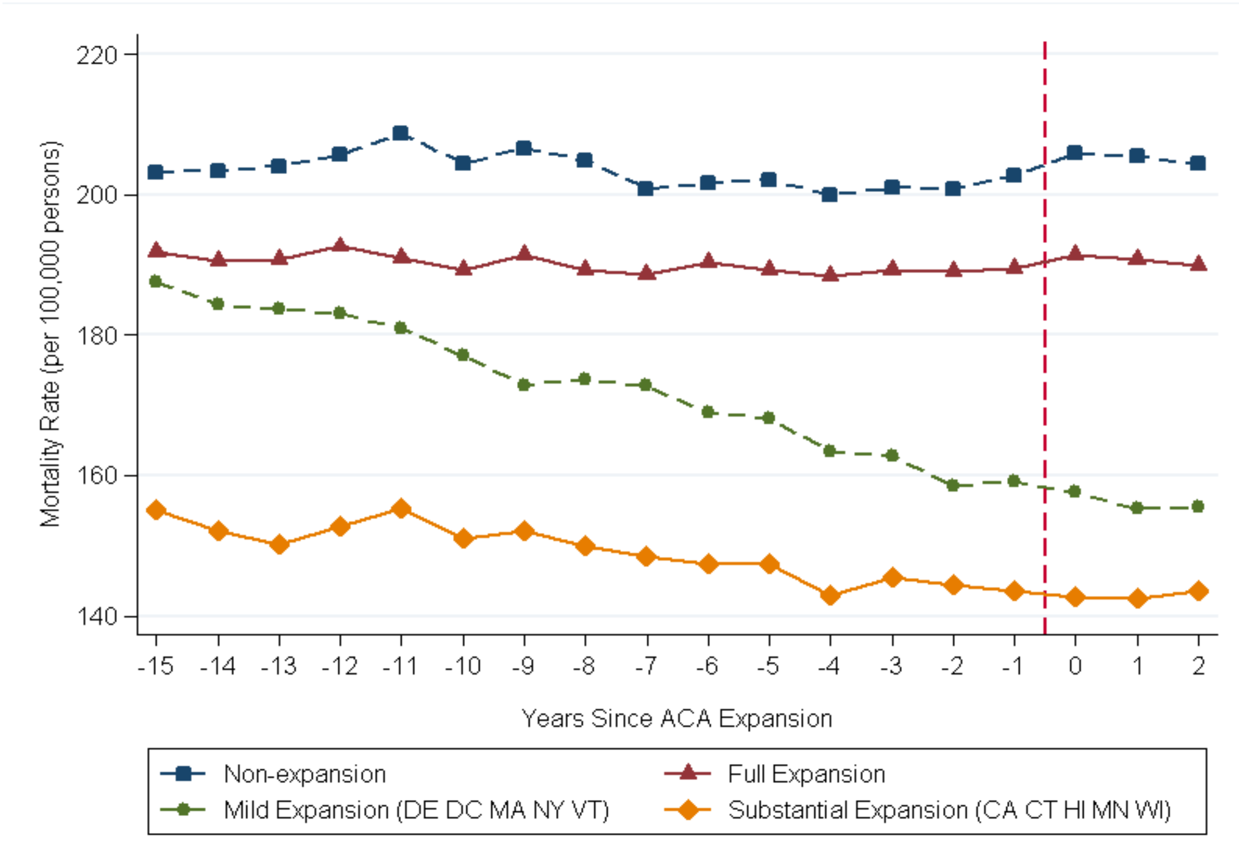


Figure A-2. Synthetic Control Results for Near-Elderly Amenable Mortality

Synthetic control results for *ln*((amenable mortality/100,000 persons)+1) for Full-Expansion States (treated as a single treated unit) versus synthetic control drawn from Non-Expansion States, over 1999-2016. Covariates for constructing donor pool are same as in Figure 2, plus uninsurance rate in 2013. The y-axis shows *ln*((amenable mortality/100,000 persons)+1) for Full-Expansion States, combined into single treated unit (using population weights), and their synthetic control. Vertical dotted line separates pre-expansion from expansion period.

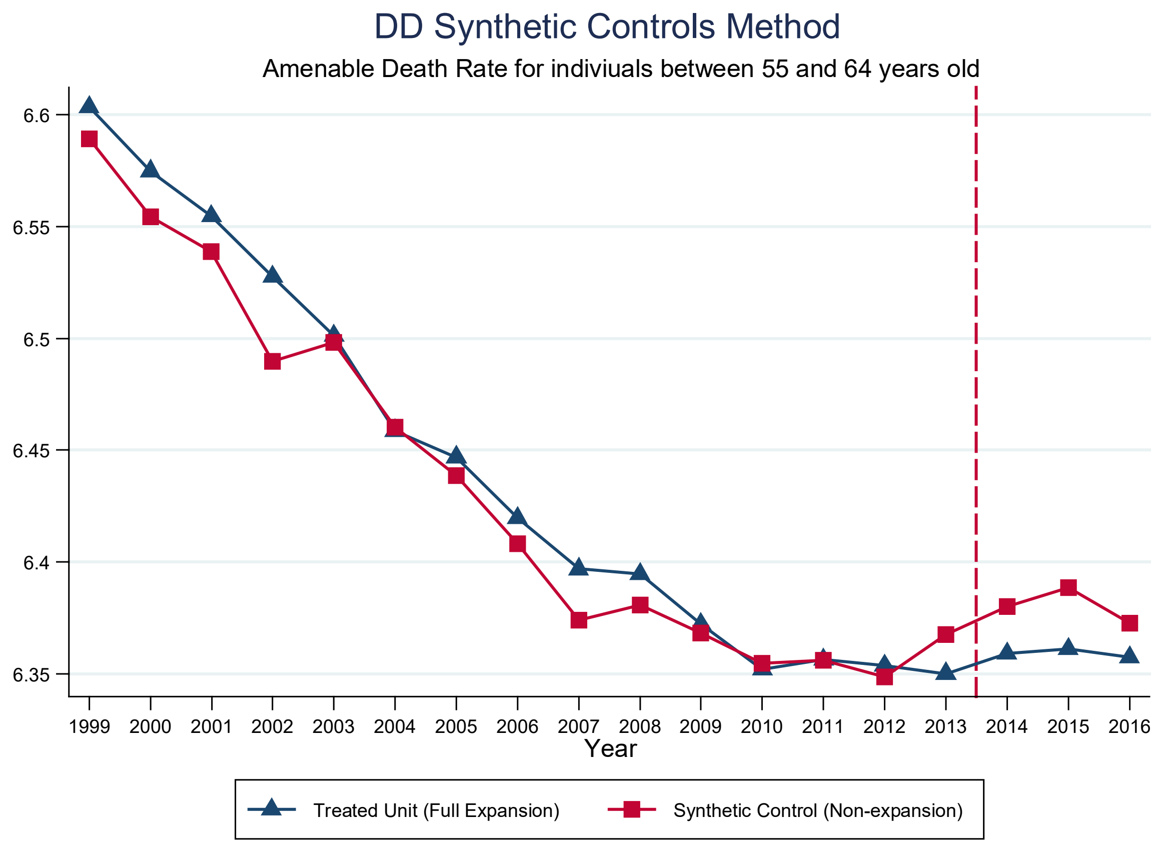


Figure A-3. Generalized Synthetic Control Method (gsynth)

Synthetic control results, using Xu's (2017) generalized synthetic control (gsynth) method, for *ln*(amenable mortality/100,000 + 1) for Full-Expansion States versus synthetic control for each state over 1999-2015. The donor pool consists of every non-expansion state's 55 to 64 year-old death rate as well as every state's untreated 65 to 74 year old population. This design is intended to crudely approximate triple-difference results. States are equally weighted. Covariates for constructing synthetic control are same as in the specifications with covariates in Table 2 of the text. The y-axis shows coefficient on Full-Expansion dummy. Vertical bars around point estimates show 95% CIs. Dashed vertical line separates pre-expansion from expansion period.

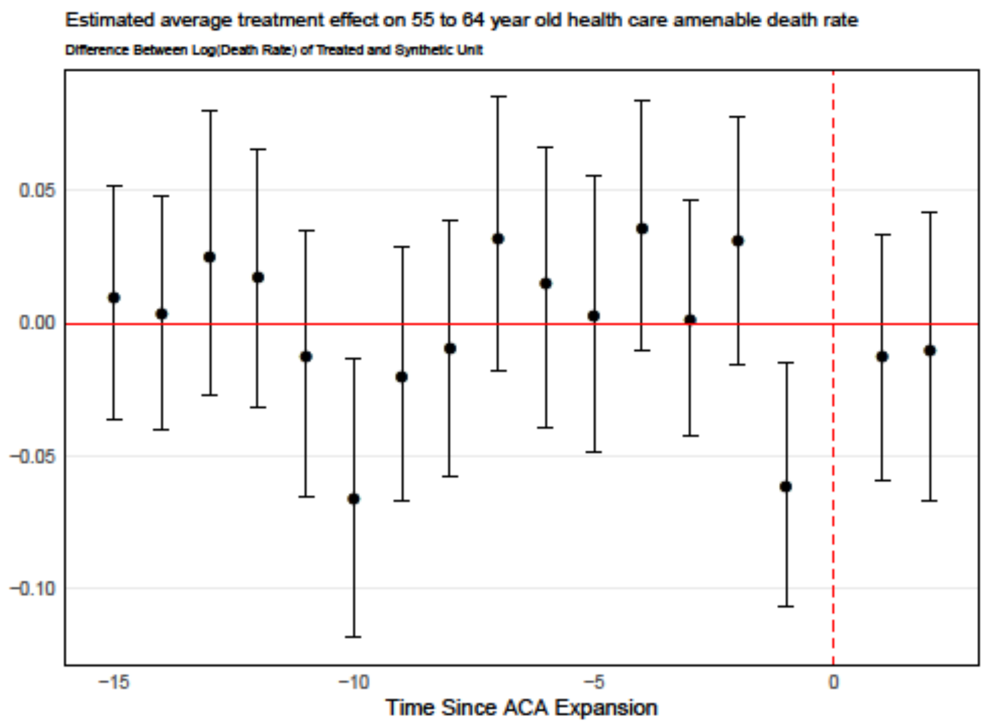
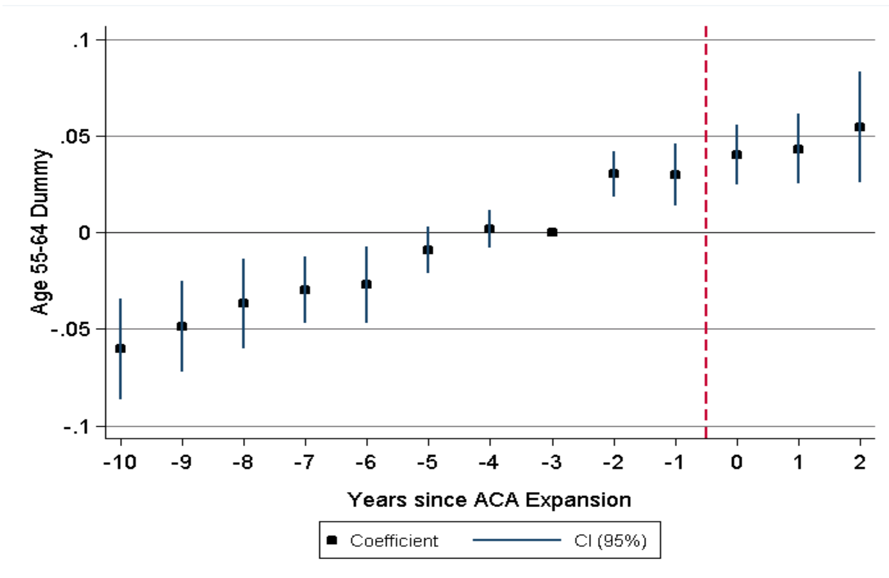


Figure A-4. Age Discontinuity Leads-and-Lags Results, Separately for Full-Expansion and No-Expansion States

Graphs from leads-and-lags regressions of *ln*((amenable mortality/100,000 persons)+1) for 55-64 versus 65-74 age groups in Full-Expansion (Panel A) and No-Expansion States (Panel B), over 2004-2016. Covariates are listed in paper. Regressions include county and year FE, and county-population weights. y-axis shows coefficients on lead and lag dummies; vertical bars show 95% confidence intervals (CIs) around coefficients, using standard errors clustered on state. Coefficient for year -3 is set to zero.

Panel B. Amenable Mortality in Full-Expansion-States



Panel B. Amenable Mortality in No Expansion-States

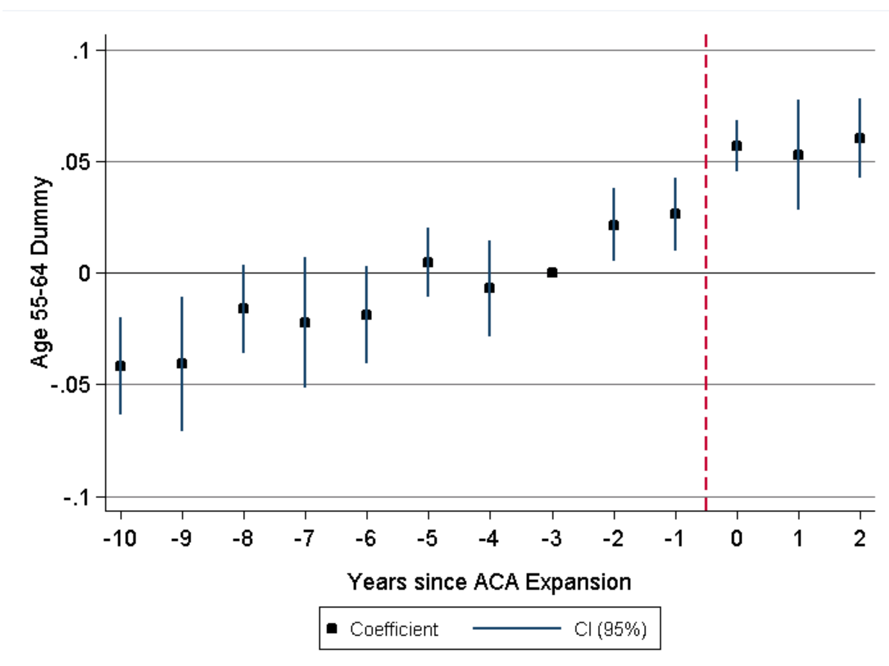


Figure A-5. Triple Difference Leads-and-Lags Graphs: Demographic Groups

Graphs from leads and lags regressions of triple differences for indicated subsamples, of *ln*((amenable mortality/100,000 persons)+1) for persons aged 55-74, in Full-Expansion States versus No-Expansion States, over 2004-2016; the third difference is age 55-64 versus age 65-74. Covariates are same as in Figure 2. Regressions include county and year FE, and county-population weights. y-axis shows coefficients on lead and lag dummies; vertical bars show 95% CIs around coefficients, using standard errors clustered on state. Coefficient for year -3 is set to zero.

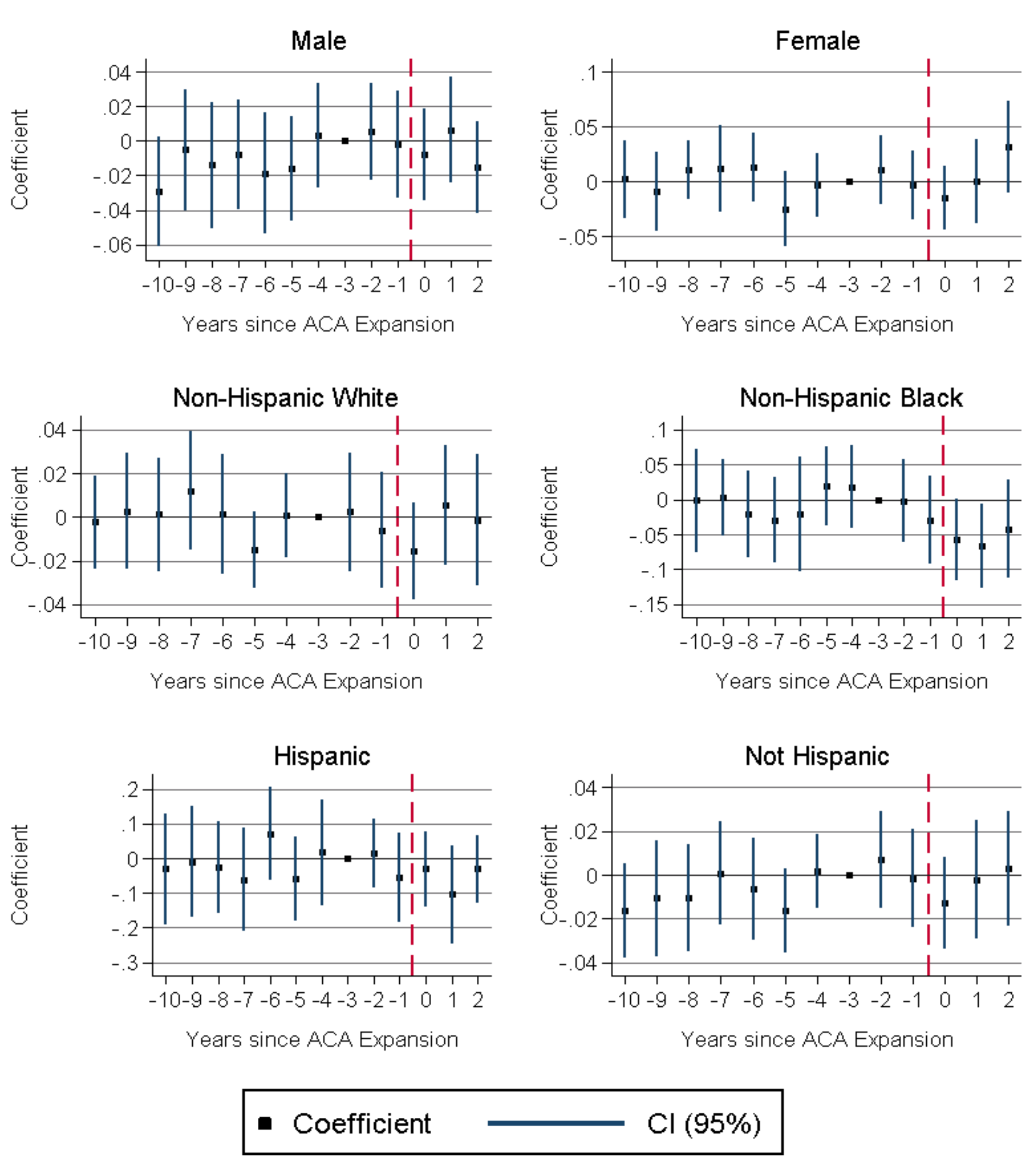


Figure A-6. Triple Difference Leads-and-Lags Graphs: By Education Level

Graphs show leads and lags regressions of triple differences for indicated subsamples, of *ln*((amenable mortality/100,000 persons)+1) for persons aged 45+, in Full-Expansion States versus No-Expansion States, over 2004-2016; the third difference is age 45-64 versus age 65+. Covariates are same as in Figure 2. Regressions include county and year FE, and county-population weights. y-axis shows coefficients on lead and lag dummies; vertical bars show 95% CIs around coefficients, using standard errors clustered on state. Coefficient for year -3 is set to zero.

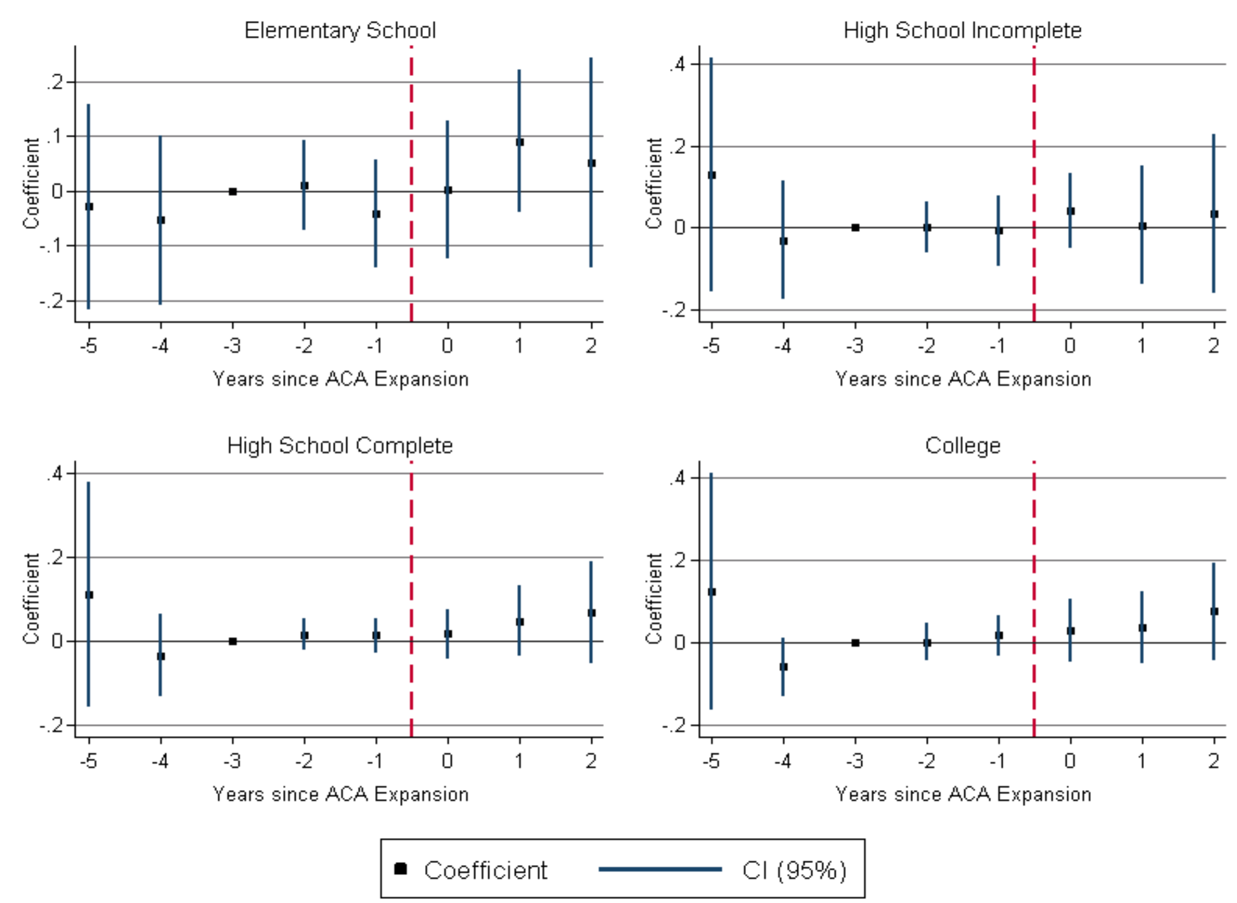


Figure A-7. Triple Difference Leads-and-Lags Graphs: By Causes of Death

Graphs show triple difference leads and lags regressions of *ln*[(mortality/100,000 persons)+1] among persons with indicated primary cause of death, aged 55-74, in Full-Expansion States versus No-Expansion States, over 2004-2016; the third difference is age 55-64 versus age 65-74. Covariates are listed in the paper. Regressions include county and year FE, and county population weights. Y-axis shows coefficients on leads and lags dummies; vertical bars show 95% CIs around coefficients, using standard errors clustered on state. Coefficient for year -3 is set to zero. Dashed vertical line separate pre-expansion from expansion period.

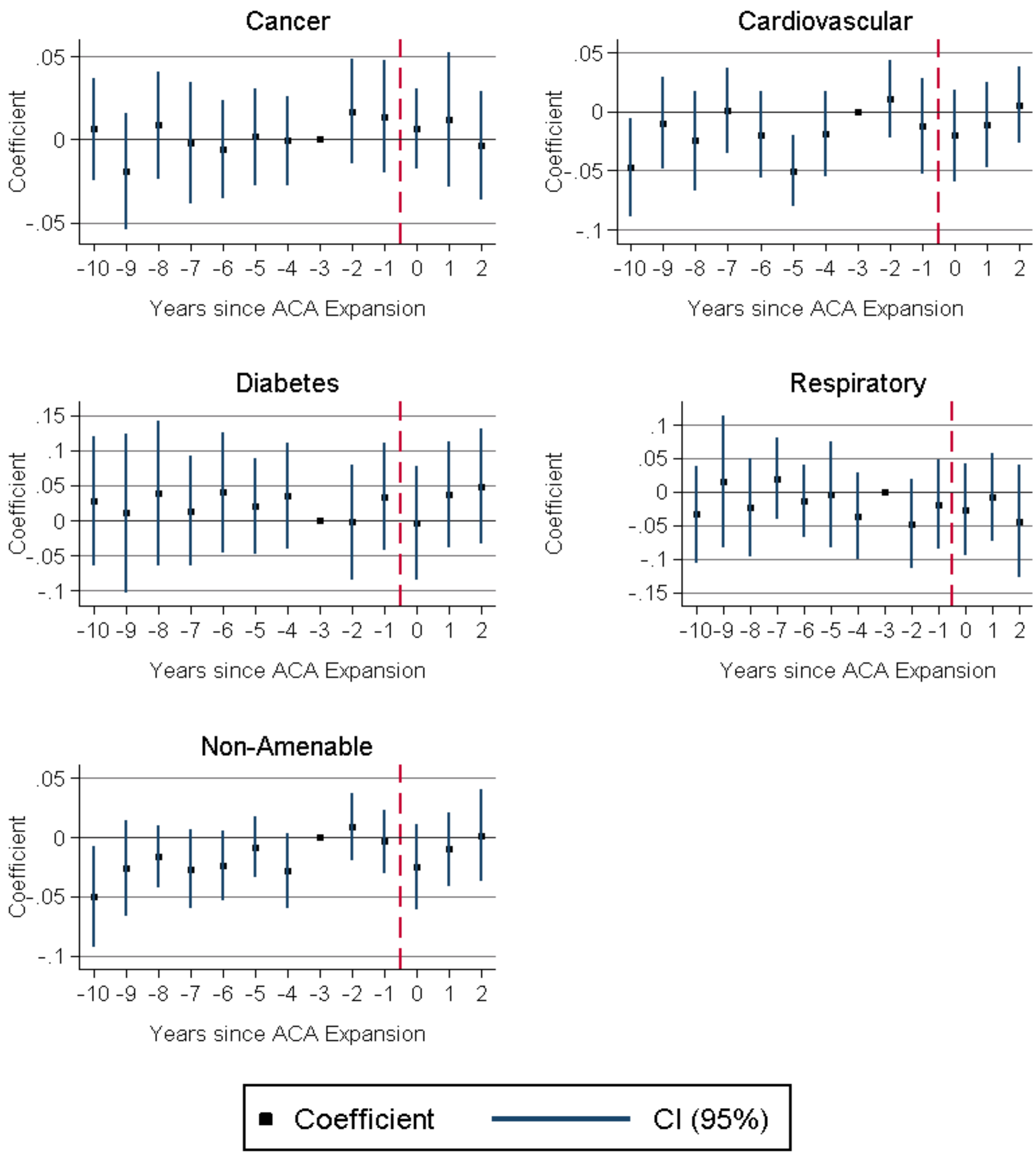
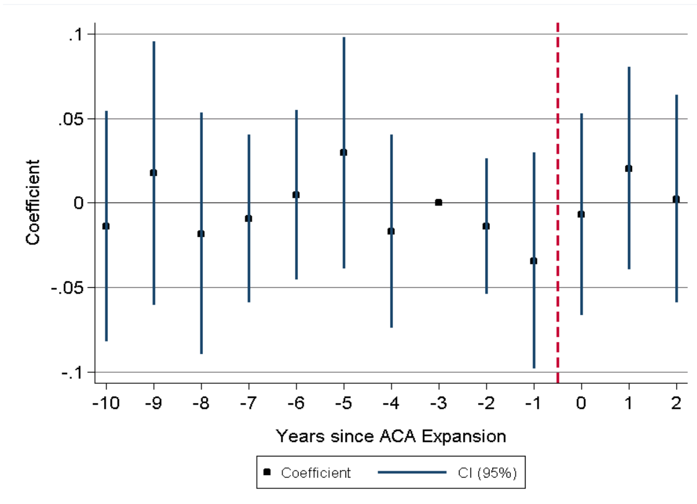


Figure A-8: Leads and Lags Graphs for High-vs-Low Uninsurance and Poverty

Graphs show leads and lags regressions of triple differences for high versus low uninsurance and high vs. low poverty counties, of *ln*((amenable mortality/100,000 persons)+1+ for persons aged 55-64, in Full-Expansion States versus No-Expansion States, over 2004-2016. High (low) uninsurance counties are those with highest (lowest) uninsurance rates in 2013 containing 20% of U.S. population, and similarly for high (low) poverty counties. Covariates are same as in Figure 2. Regressions include county and year FE, and county-population weights. y-axis shows coefficients on lead and lag dummies; vertical bars show 95% CIs around coefficients, using standard errors clustered on state. Coefficient for year -3 is set to zero.

Panel A. High-Uninsurance vs. Low-Uninsurance Counties



Panel B. High-Poverty vs. Low-Poverty Counties

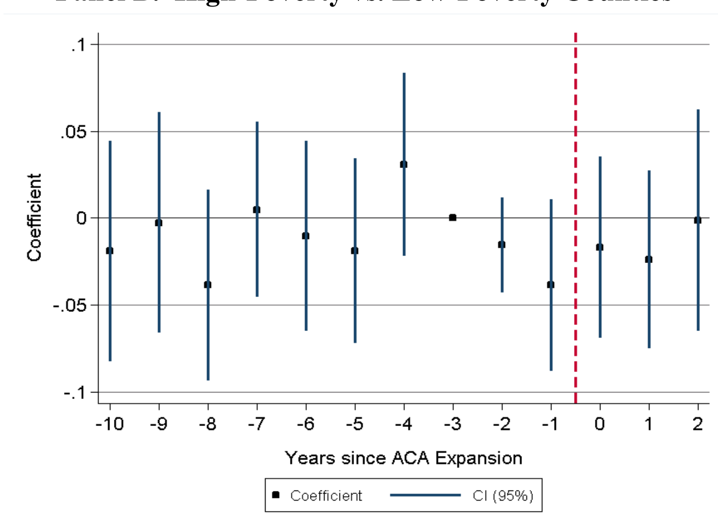
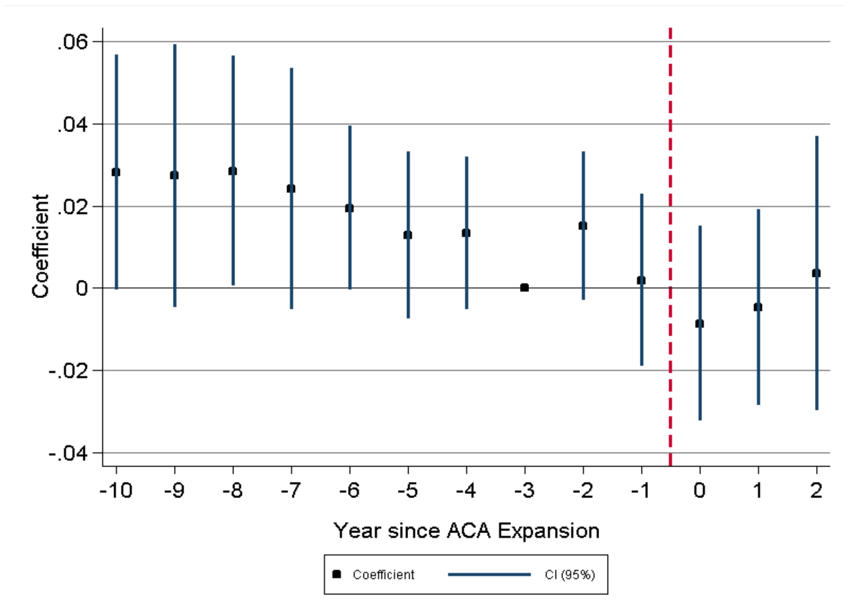


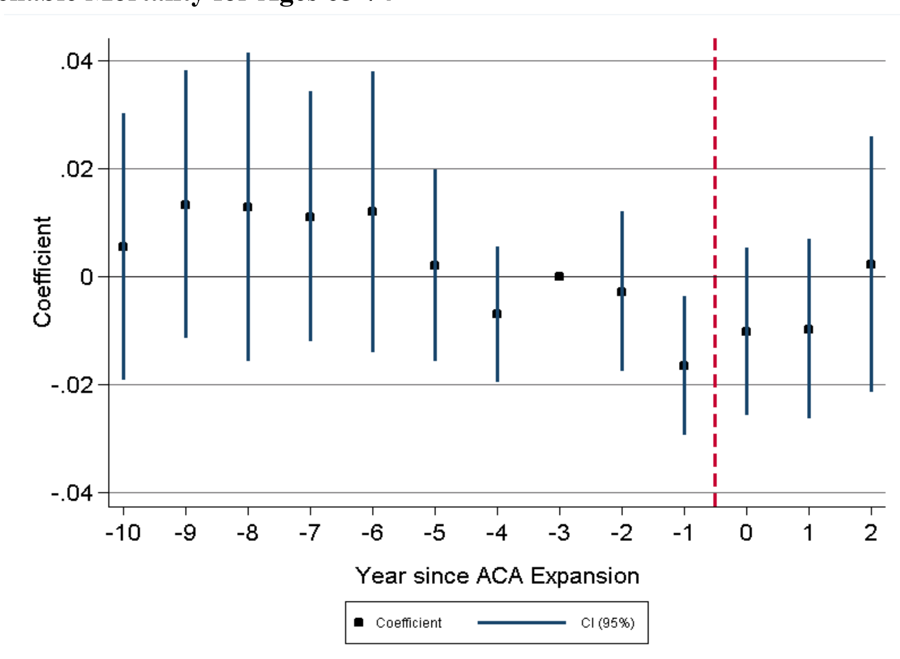
Figure A-9. DiD and Triple Difference Leads-and-Lags Results: Amenable Mortality, with ATT x Population Weights

Graphs from leads and lags regressions of *ln*[(amenable mortality/100,000 persons)+1] for Full-Expansion States versus control group of Non-Expansion States, over 2004-2016. Covariates are listed in paper. Regressions include county and year FE, and ATT x Population weights. Y-axis shows coefficients on lead and lag dummies; vertical bars show 95% confidence intervals (CIs) around coefficients, using standard errors clustered on state. Coefficient for year -3 is set to zero. Dashed vertical line separate pre-expansion from expansion period.

Panel A. Amenable Mortality for Ages 55-64



Panel B. Amenable Mortality for Ages 65-74



Panel C. Triple difference. Leads and lags graphs for amenable mortality for persons age 55-64 in Full-Expansion States, relative to (i) persons age 65-74 in Full-Expansion States, and (ii) persons age 55-64 in Non-Expansion States.

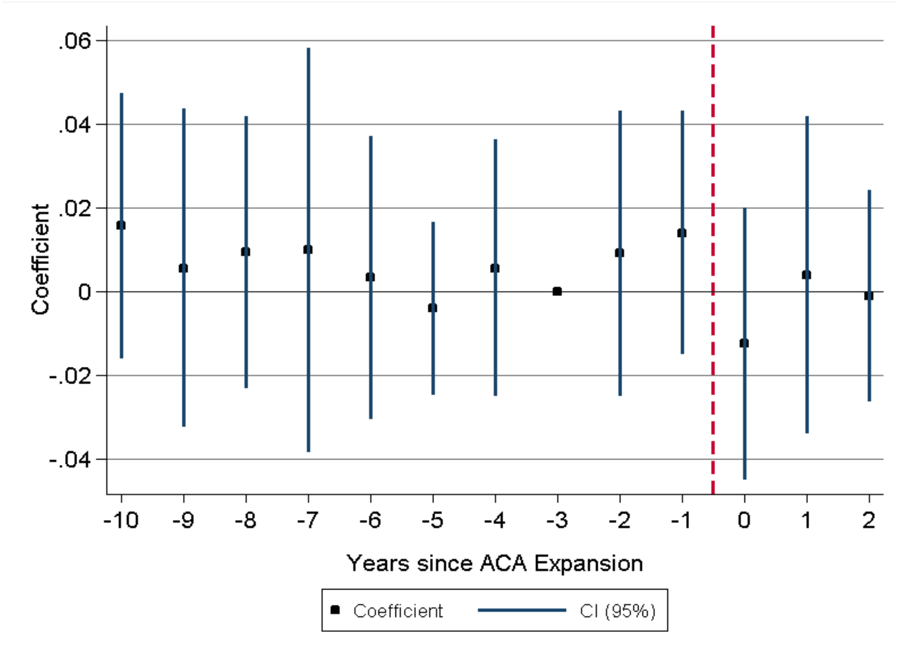
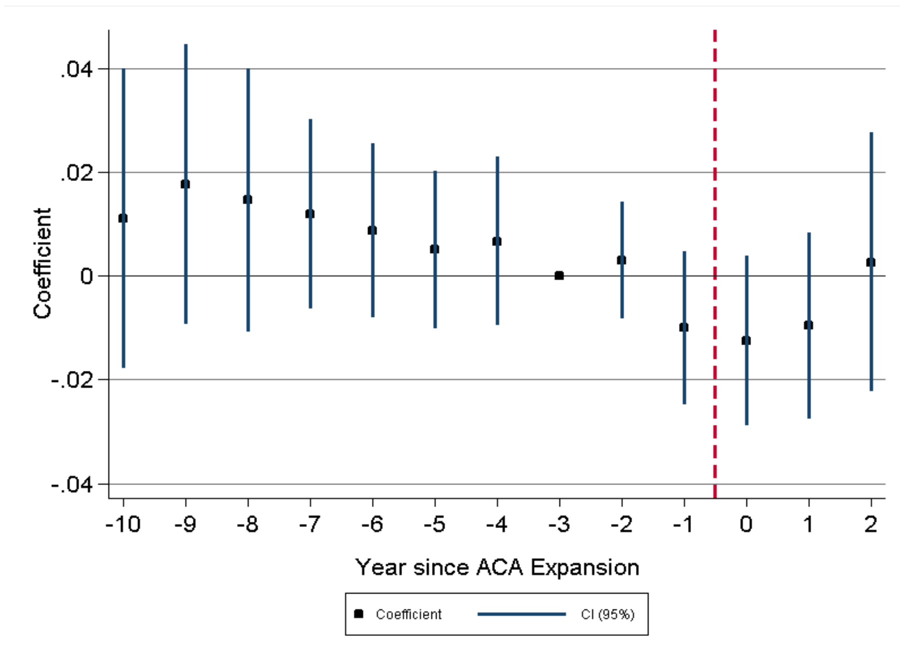


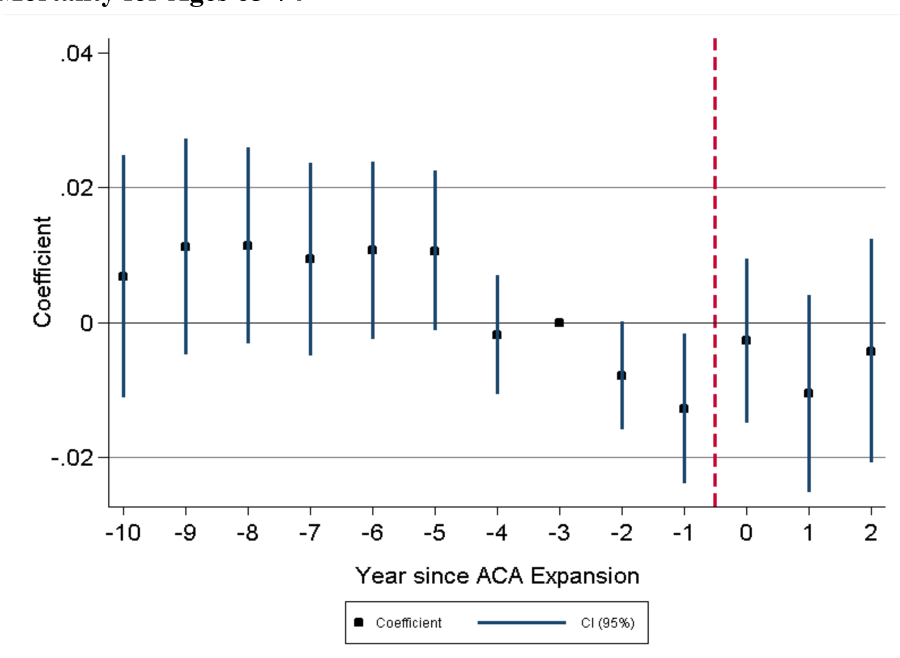
Figure A-10. DiD and Triple Difference Leads-and-Lags Results for Total Mortality

Graphs from leads and lags regressions of *ln*[(all mortality/100,000 persons)+1] for Full-Expansion States versus control group of Non-Expansion States, over 2004-2016. Covariates are listed in paper. Regressions include county and year FE, and county-population weights. Y-axis shows coefficients on lead and lag dummies; vertical bars show 95% confidence intervals (CIs) around coefficients, using standard errors clustered on state. Coefficient for year -3 is set to zero. Dashed vertical line separate pre-expansion from expansion period.

Panel A. All Mortality for Ages 55-64



Panel B. All Mortality for Ages 65-74



Panel C. Triple difference. Leads and lags graphs for all mortality for persons age 55-64 in Full-Expansion States, relative to (i) persons age 65-74 in Full-Expansion States, and (ii) persons age 55-64 in Non-Expansion States.

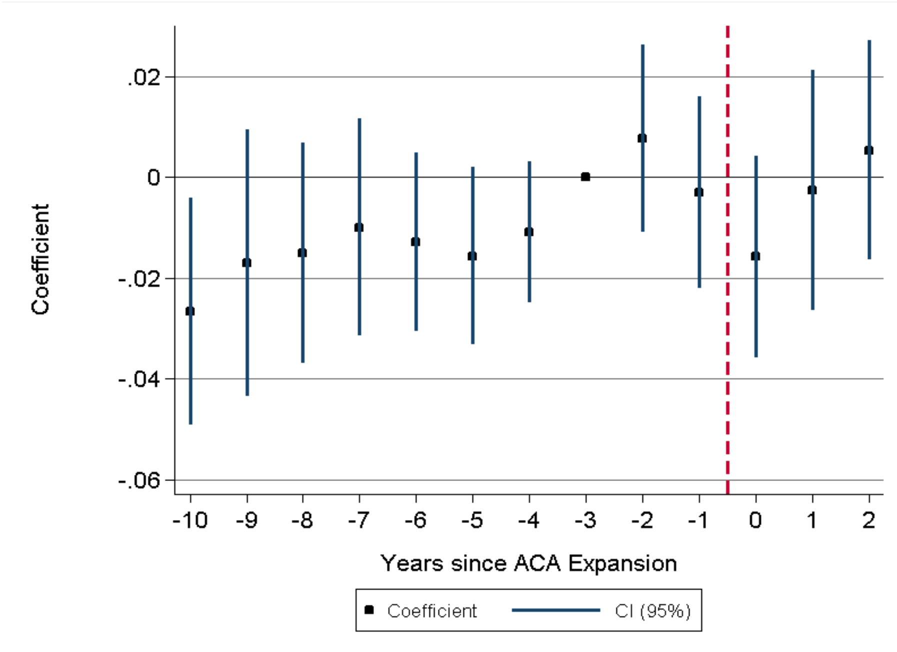
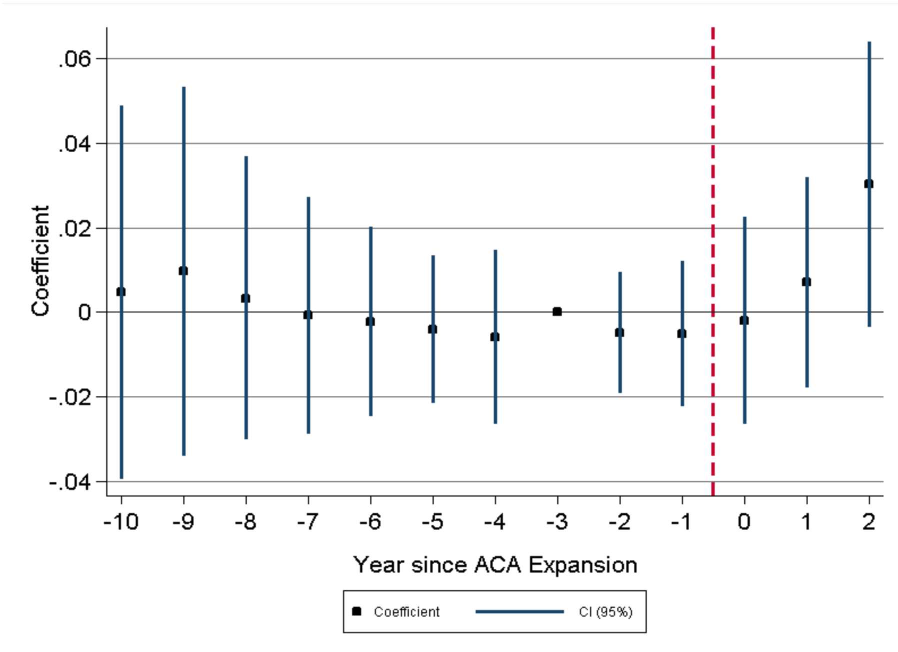


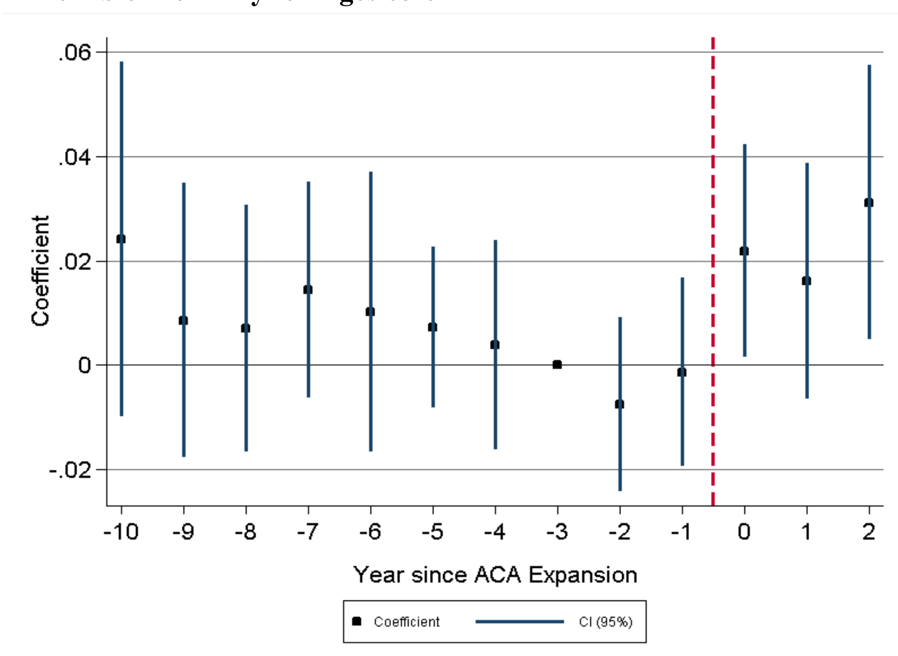
Figure A-11. DiD and Triple Difference Leads-and-Lags Results for Non-Amenable Mortality

Graphs from leads and lags regressions of *ln*[(non-amenable mortality/100,000 persons)+1] for Full-Expansion States versus control group of Non-Expansion States, over 2004-2016. Covariates are listed in paper. Regressions include county and year FE, and county-population weights. Y-axis shows coefficients on lead and lag dummies; vertical bars show 95% confidence intervals (CIs) around coefficients, using standard errors clustered on state. Coefficient for year -3 is set to zero. Dashed vertical line separate pre-expansion from expansion period.

Panel A. Non-Amenable Mortality for Ages 55-64



Panel B. Non-Amenable Mortality for Ages 65-74



Panel C. Triple difference. Leads and lags graphs for non-amenable mortality for persons age 55-64 in Full-Expansion States, relative to (i) persons age 65-74 in Full-Expansion States, and (ii) persons age 55-64 in Non-Expansion States.

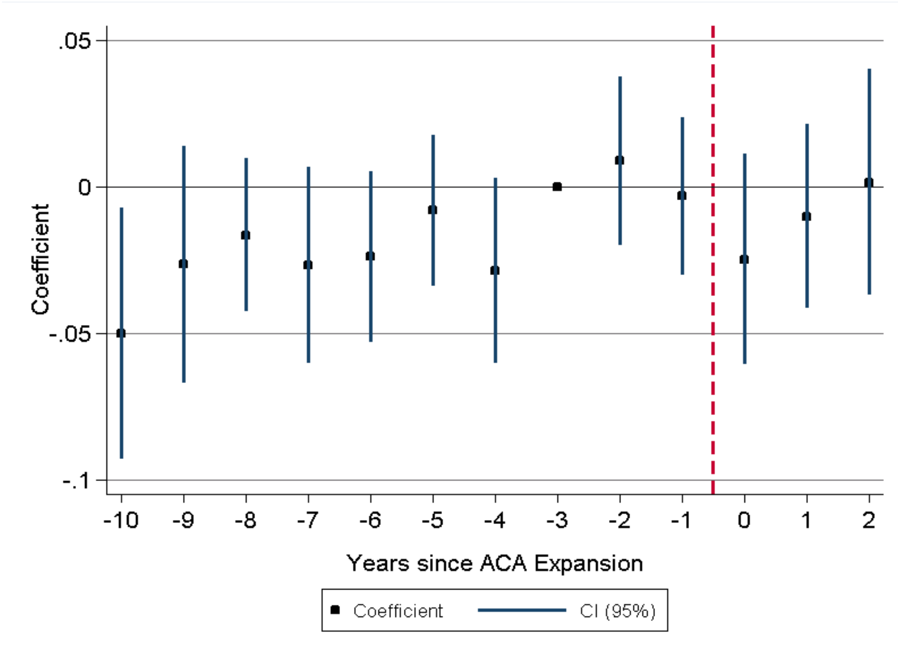


Figure A-12. Triple Difference Leads-and-Lags Graphs: Demographic Groups, with ATT x Population Weights

Graphs from leads and lags regressions of triple differences for indicated subsamples, of *ln*[(amenable mortality/100,000 persons)+1] for persons aged 55-74, in Full-Expansion States versus No-Expansion States, over 2004-2016; the third difference is age 55-64 versus age 65-74. Covariates are listed in the paper. Regressions include county and year FE, and Att x Pop weights. Y-axis shows coefficients on lead and lag dummies; vertical bars show 95% CIs around coefficients, using standard errors clustered on state. Coefficient for year -3 is set to zero. Dashed vertical line separate pre-expansion from expansion period.

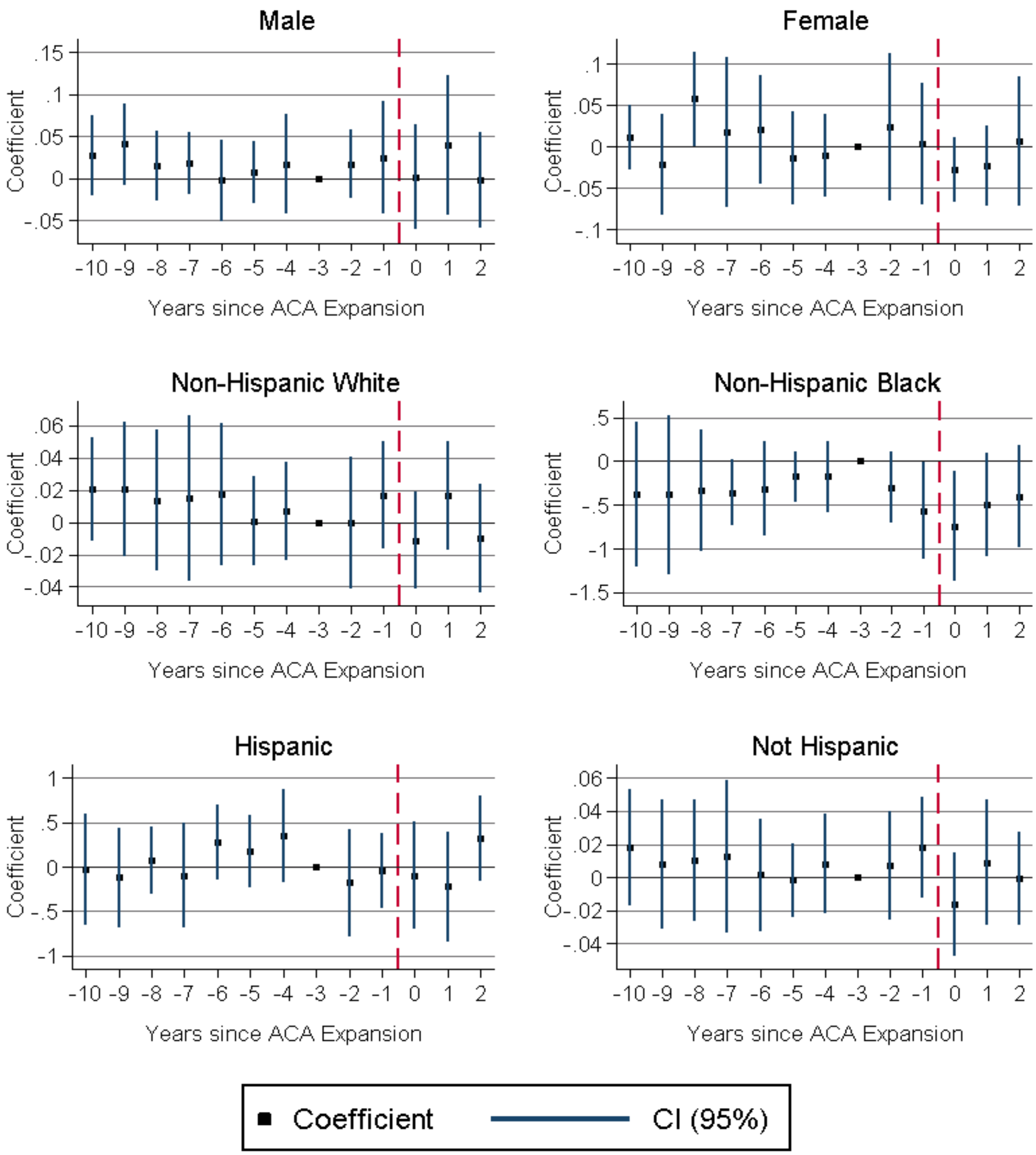


Figure A-13. Triple Difference Leads-and-Lags Graphs: By Education Level, with ATT x Population Weights

Graphs show leads and lags regressions of triple differences for indicated subsamples, of *ln*[(amenable mortality/100,000 persons)+1] for persons aged 45+, in Full-Expansion States versus No-Expansion States, over 2004-2016; the third difference is age 45-64 versus age 65+. Covariates are listed in the paper. Regressions include county and year FE, and ATT x Population weights. y-axis shows coefficients on lead and lag dummies; vertical bars show 95% CIs around coefficients, using standard errors clustered on state. Coefficient for year -3 is set to zero. Dashed vertical line separate pre-expansion from expansion period.

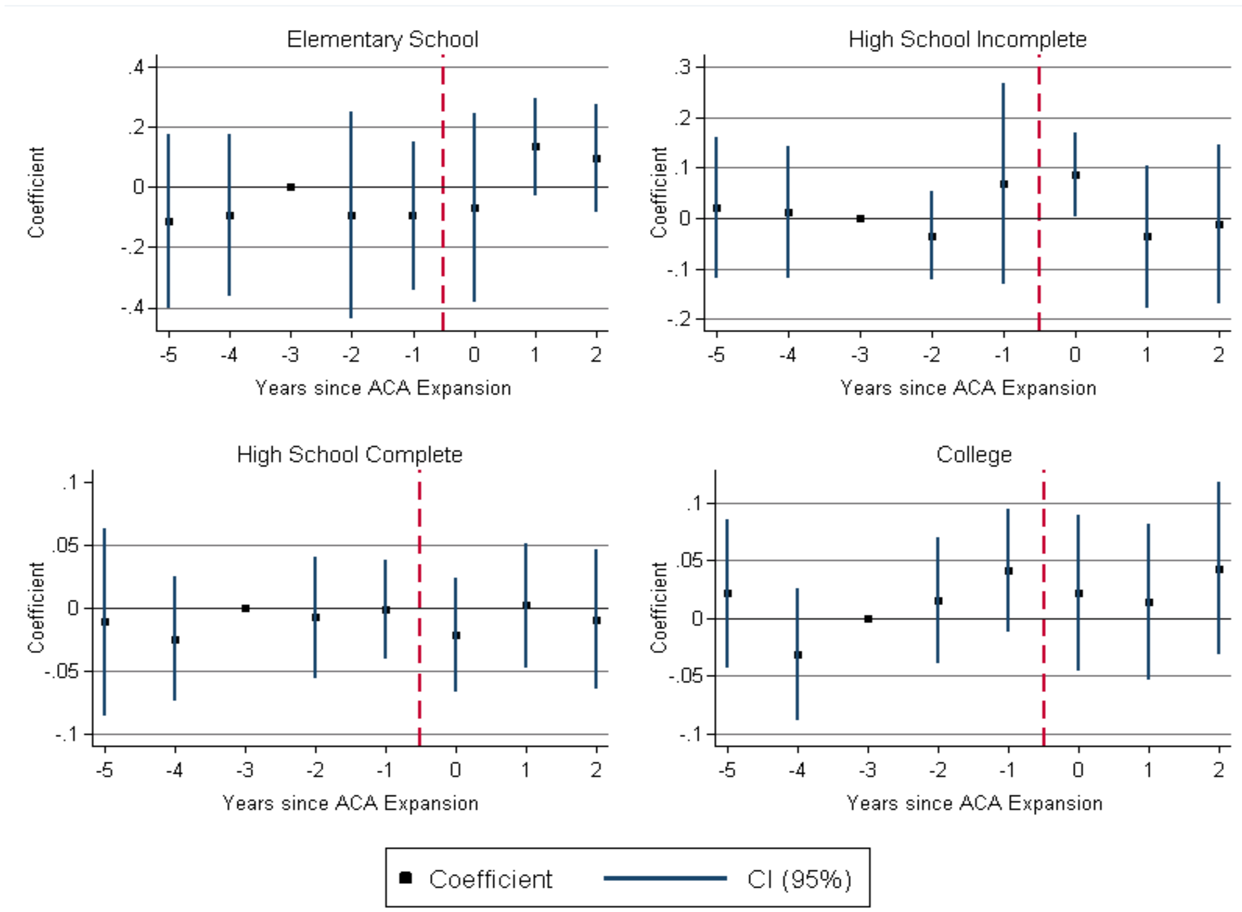


Figure A-14. Triple Difference Leads-and-Lags Graphs: By Causes of Death, ATT x Population Weights

Graphs show triple difference leads and lags regressions of *ln*[(mortality/100,000 persons)+1] among persons with indicated primary cause of death, aged 55-74, in Full-Expansion States versus No-Expansion States, over 2004-2016; the third difference is age 55-64 versus age 65-74. Covariates are listed in the paper. Regressions include county and year FE, and ATT x population weights. Y-axis shows coefficients on leads and lags dummies; vertical bars show 95% CIs around coefficients, using standard errors clustered on state. Coefficient for year -3 is set to zero. Dashed vertical line separate pre-expansion from expansion period.

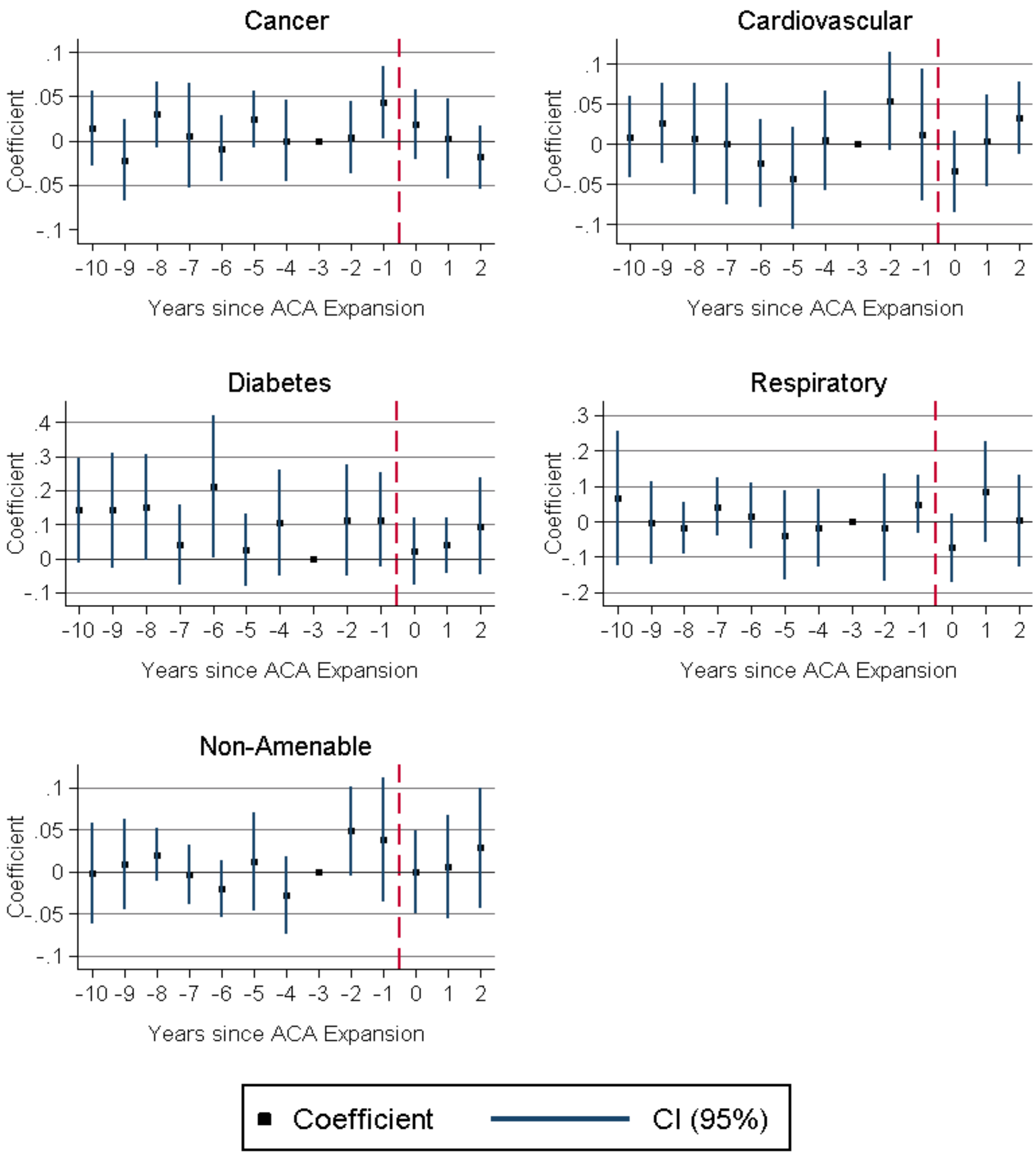
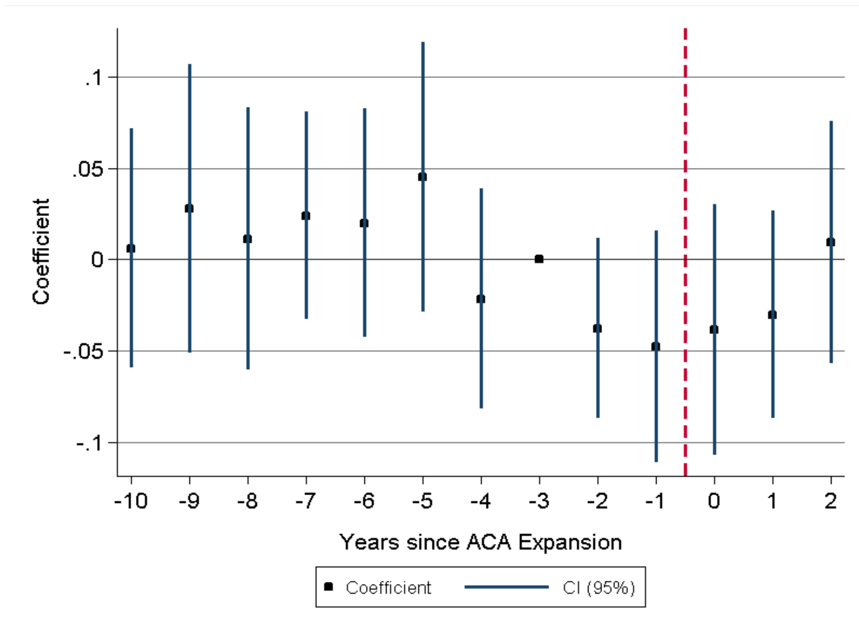


Figure A-15: Leads and Lags Graphs for High-vs-Low Uninsurance and Poverty, ATT x Pop weights

Graphs show leads and lags regressions of triple differences for high versus low uninsurance and high vs. low poverty counties, of *ln*[(amenable mortality/100,000 persons)+1] for persons aged 55-64, in Full-Expansion States versus No-Expansion States, over 2004-2016. High (low) uninsurance counties are those with highest (lowest) uninsurance rates in 2013 containing 20% of U.S. population, and similarly for high (low) poverty counties. Covariates are listed in the paper. Regressions include county and year FE, and ATT x Pop weights. Y-axis shows coefficients on lead and lag dummies; vertical bars show 95% CIs around coefficients, using standard errors clustered on state. Coefficient for year -3 is set to zero. Dashed vertical line separate pre-expansion from expansion period.

Panel A. High-Uninsurance vs. Low-Uninsurance Counties



Panel B. High-Poverty vs. Low-Poverty Counties

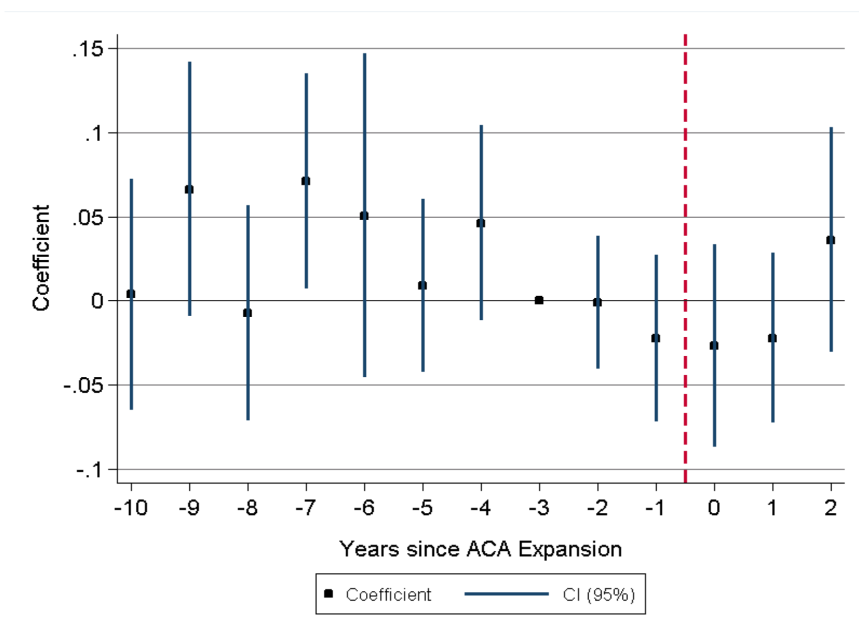
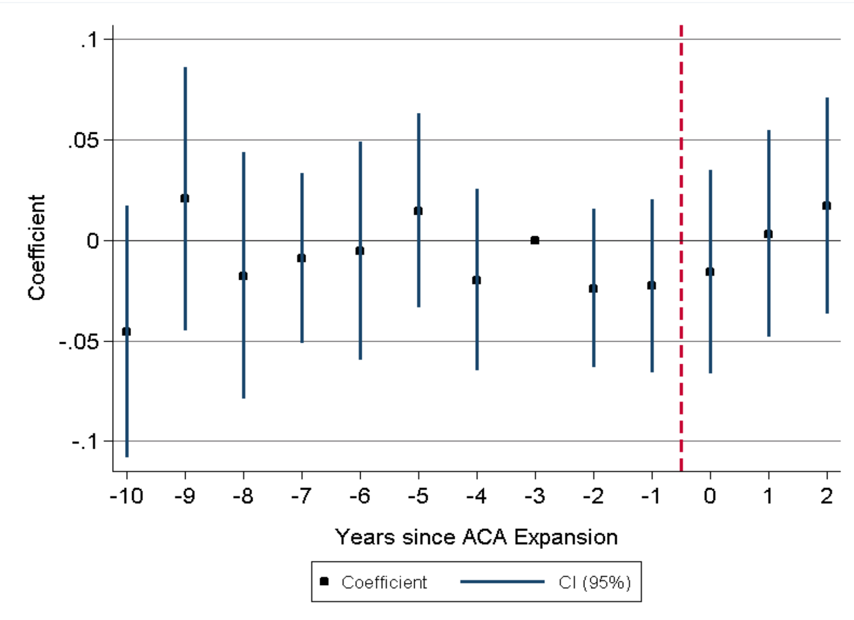


Figure A-16: Leads and Lags Graphs for High-vs-Low Uninsurance and Poverty, 18-64 years

Graphs show leads and lags regressions of triple differences for high versus low uninsurance and high vs. low poverty counties, of *ln*[(amenable mortality/100,000 persons)+1] for persons aged 18-64, in Full-Expansion States versus No-Expansion States, over 2004-2016. High (low) uninsurance counties are those with highest (lowest) uninsurance rates in 2013 containing 20% of U.S. population, and similarly for high (low) poverty counties. Covariates are listed in the paper. Regressions include county and year FE, and county population weights. Y-axis shows coefficients on lead and lag dummies; vertical bars show 95% CIs around coefficients, using standard errors clustered on state. Coefficient for year -3 is set to zero. Dashed vertical line separate pre-expansion from expansion period.

Panel A. High-Uninsurance vs. Low-Uninsurance Counties



Panel B. High-Poverty vs. Low-Poverty Counties

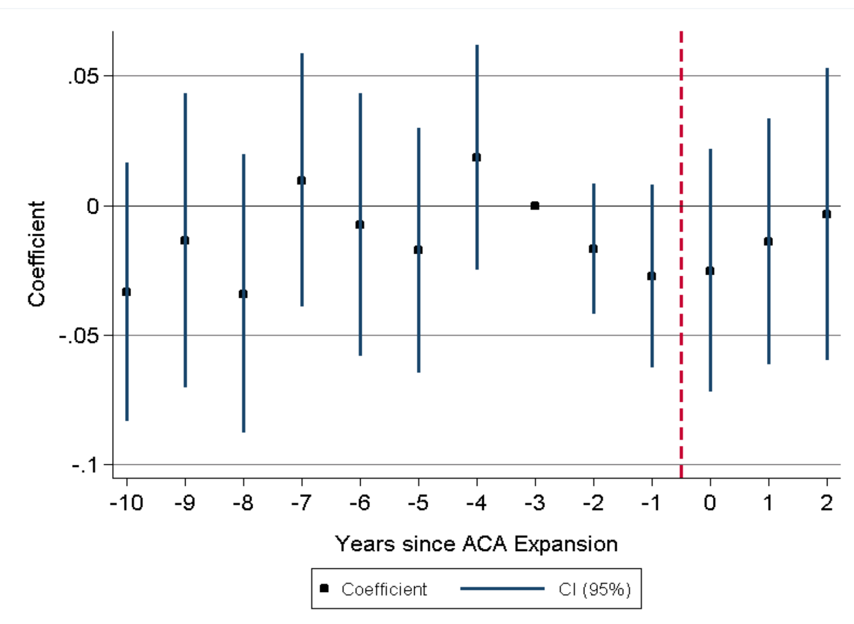


Figure A-17. DiD Leads-and-Lags Results for Ages 18-64, Amenable Mortality

Graphs from DiD leads and lags regressions of *ln*[(amenable mortality/100,000 persons)+1] for Full-Expansion States versus control group of Non-Expansion States, over 2004-2016. Covariates are listed in paper. Regressions include county and year FE, and county population weights. Y-axis shows coefficients on lead and lag dummies; vertical bars show 95% confidence intervals (CIs) around coefficients, using standard errors clustered on state. Coefficient for year -3 is set to zero. Dashed vertical line separate pre-expansion from expansion period.

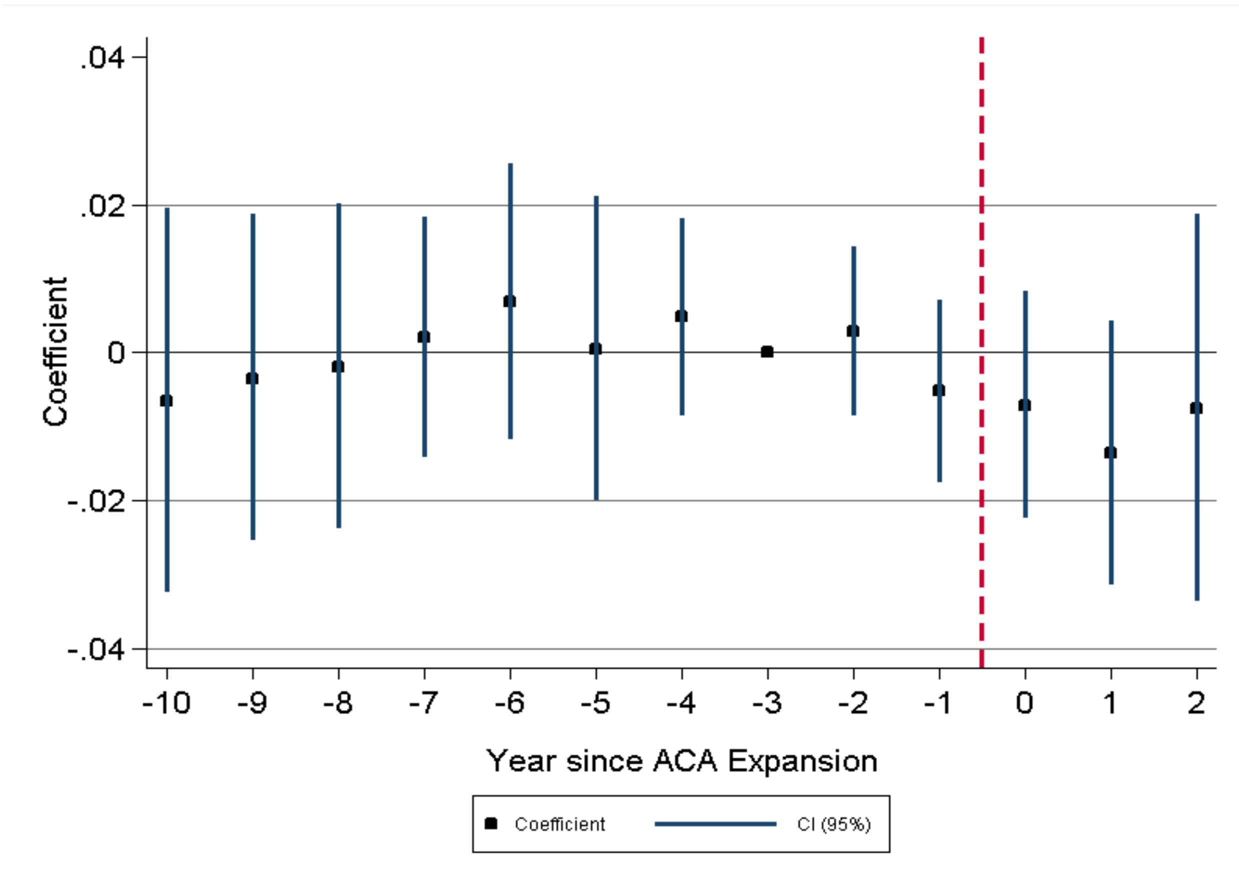
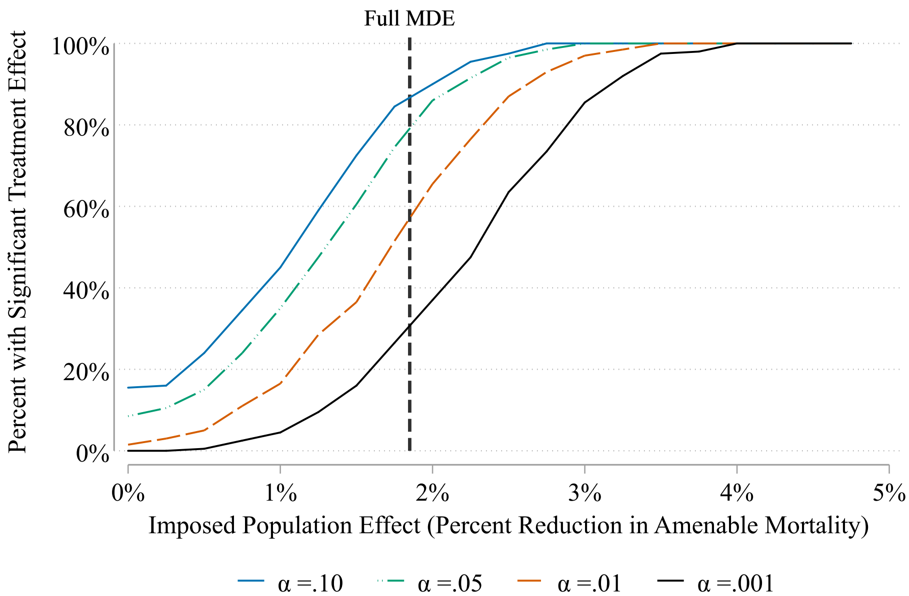


Figure A-18. Power Analyses for Full Sample: State Level DD and Triple Differences

Power curves for simulated Medicaid expansion as of January 1, 2012, applied to persons aged 55-64 during pre-treatment period (2007-2013). Graphs show power (likelihood of detecting a statistically significant effect on amenable mortality, at the indicated confidence levels, for two-tailed test), given imposed "true" population average effect. Curves are based on 1,000 replications of the DD (top graph) and triple difference (bottom graph) regression models used in Table 2, with covariates. In each draw, we select 20 pseudo-treated states at random from the combined set of 41 treated and control states, and remove a fraction of the observed deaths at random from the treated states, where the fraction reflects an imposed treatment effect (for the entire population), and we vary the imposed treatment effect from 0-5% in increments of 0.1%. Curves for α = .10/.05/.01/.001 correspond to 90%/95%/99%/99.9% confidence levels, respectively. Dashed vertical line indicates minimum detectable effect at 95% confidence level, with 80% power, for full sample (Full MDE).



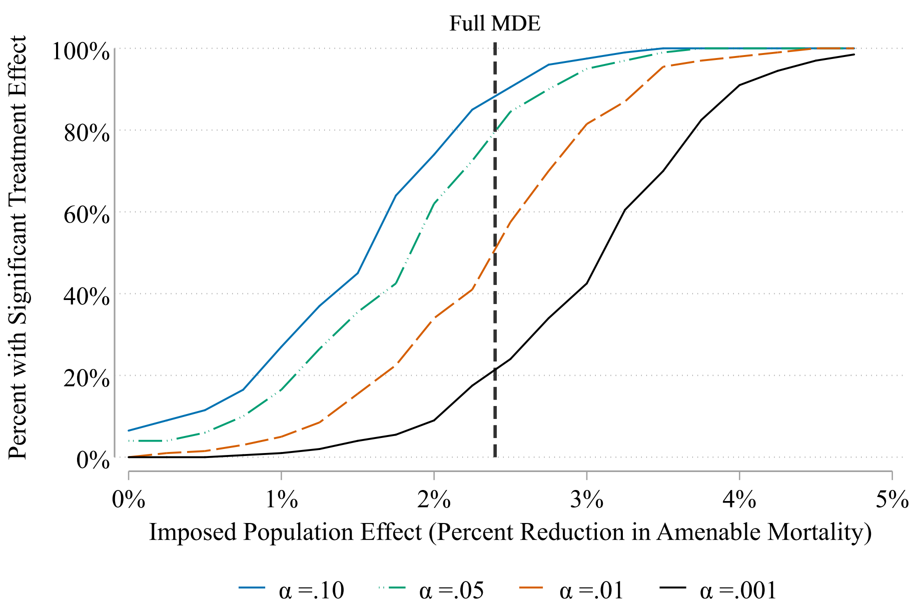
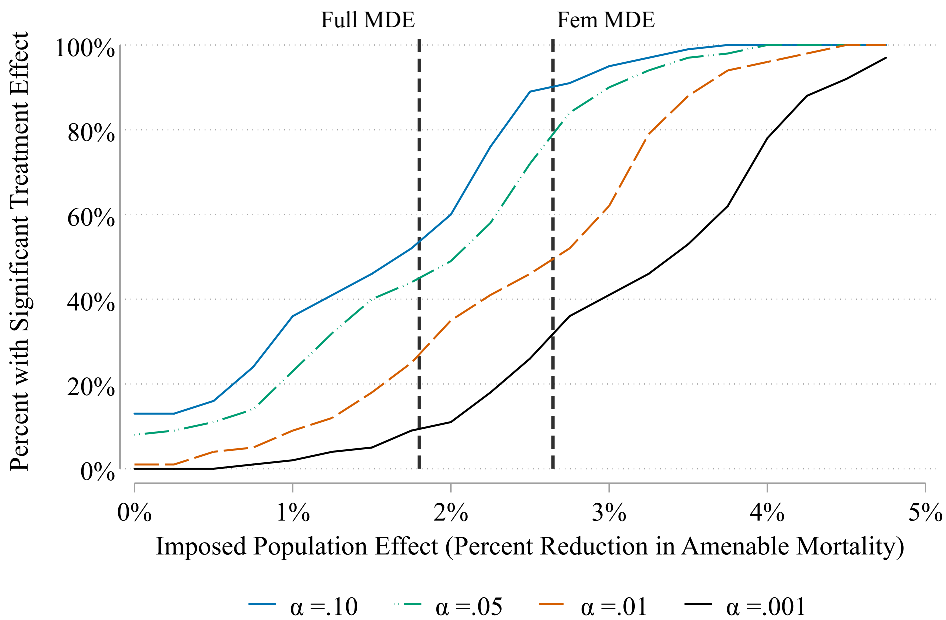


Figure A-19. Power Analysis for Women: DD and Triple Differences

Power curves for simulated Medicaid expansion as of January 1, 2012, applied to females aged 55-64 during pre-treatment period (2007-2013). Graphs show power (likelihood of detecting a statistically significant effect on amenable mortality, at the indicated confidence levels, for two-tailed test), given imposed "true" population average effect. Curves are based on 1,000 replications of the DD (top graph) and triple difference (bottom graph) regression models used in Table 2, with covariates. In each draw, we select 20 pseudo-treated states at random from the combined set of 41 treated and control states, and remove a fraction of the observed deaths at random from the treated states, where the fraction reflects an imposed treatment effect (for the entire population), and we vary the imposed treatment effect from 0-5% in increments of 0.1%. Curves for α = .10/.05/.01/.001 correspond to 90%/95%/99%/99.9% confidence levels, respectively. Dashed vertical lines indicate minimum detectable effects at 95% confidence level, with 80% power, for full sample (Full MDE) and for women (Fem MDE).



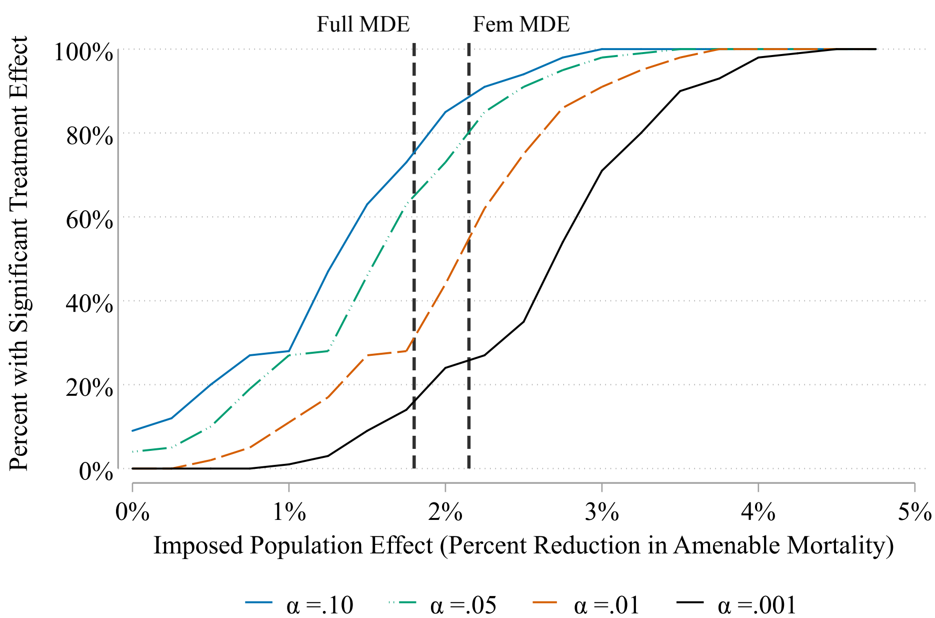
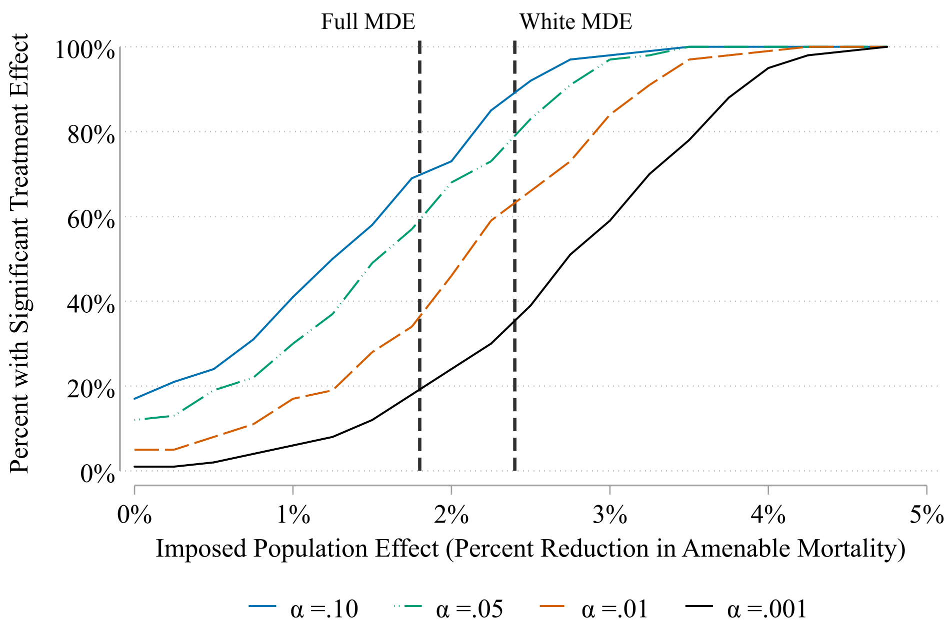


Figure A-20. Power Analysis for Non-Hispanic Whites: DD and Triple Differences

Power curves for simulated Medicaid expansion as of January 1, 2012, applied to non-Hispanic whites aged 55-64 during pre-treatment period (2007-2013). Graphs show power (likelihood of detecting a statistically significant effect on amenable mortality, at the indicated confidence levels, for two-tailed test), given imposed "true" population average effect. Curves are based on 1,000 replications of the DD (top graph) and triple difference (bottom graph) regression models used in Table 2, with covariates. In each draw, we select 20 pseudo-treated states at random from the combined set of 41 treated and control states, and remove a fraction of the observed deaths at random from the treated states, where the fraction reflects an imposed treatment effect (for the entire population), and we vary the imposed treatment effect from 0-5% in increments of 0.1%. Curves for α = .10/.05/.01/.001 correspond to 90%/95%/99%/99.9% confidence levels, respectively. Dashed vertical lines indicate minimum detectable effects at 95% confidence level, with 80% power, for full sample (Full MDE) and for non-Hispanic whites (White MDE).



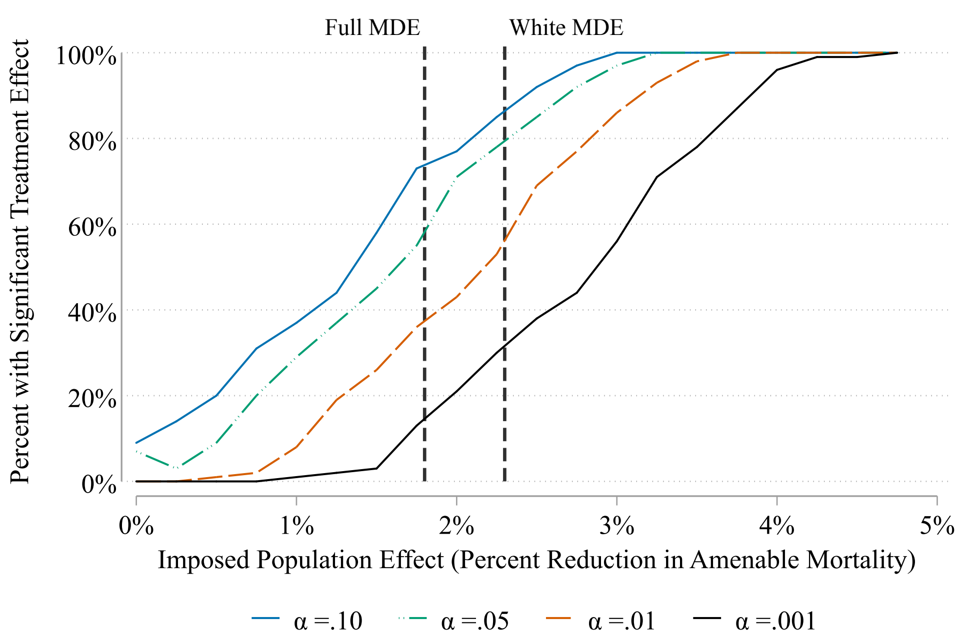
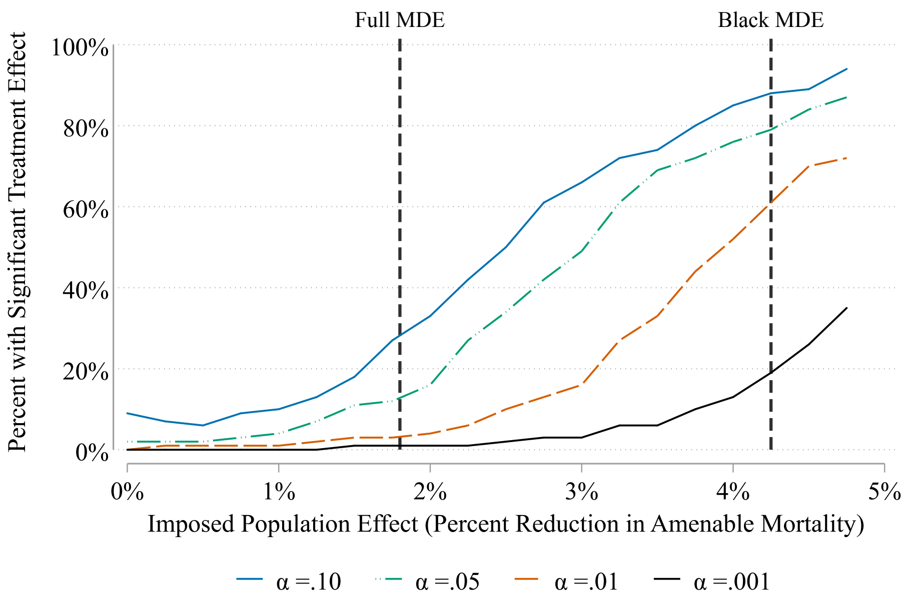


Figure A-21. Power Analysis for Non-Hispanic Blacks: DD and Triple Differences

Power curves for simulated Medicaid expansion as of January 1, 2012, applied to non-Hispanic blacks aged 55-64 during pre-treatment period (2007-2013). Graphs show power (likelihood of detecting a statistically significant effect on amenable mortality, at the indicated confidence levels, for two-tailed test), given imposed "true" population average effect. Curves are based on 1,000 replications of the DD (top graph) and triple difference (bottom graph) regression models used in Table 2, with covariates. In each draw, we select 20 pseudo-treated states at random from the combined set of 41 treated and control states, and remove a fraction of the observed deaths at random from the treated states, where the fraction reflects an imposed treatment effect (for the entire population), and we vary the imposed treatment effect from 0-5% in increments of 0.1%. Curves for α = .10/.05/.01/.001 correspond to 90%/95%/99%/99.9% confidence levels, respectively. Dashed vertical lines indicate minimum detectable effects at 95% confidence level, with 80% power, for full sample (Full MDE) and for non-Hispanic blacks (Black MDE).



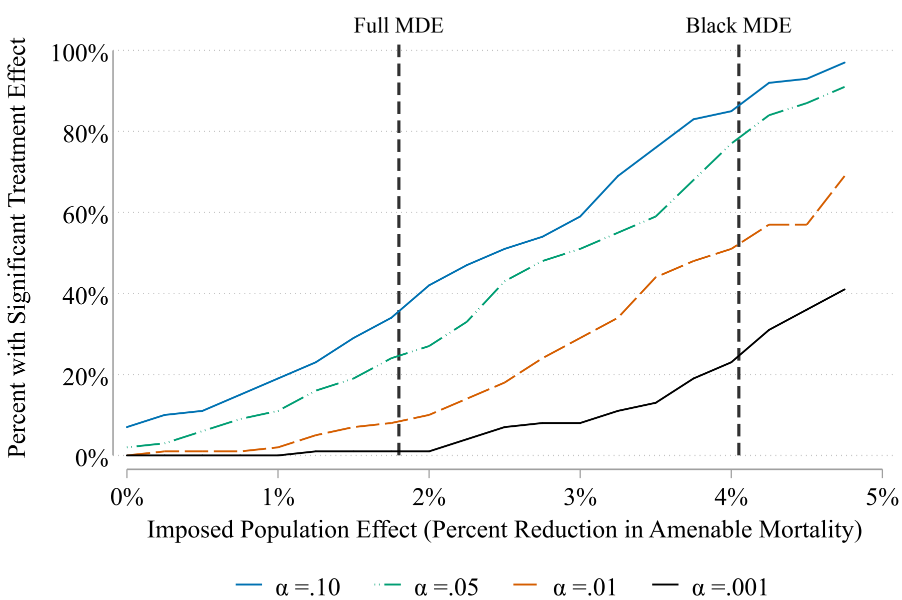
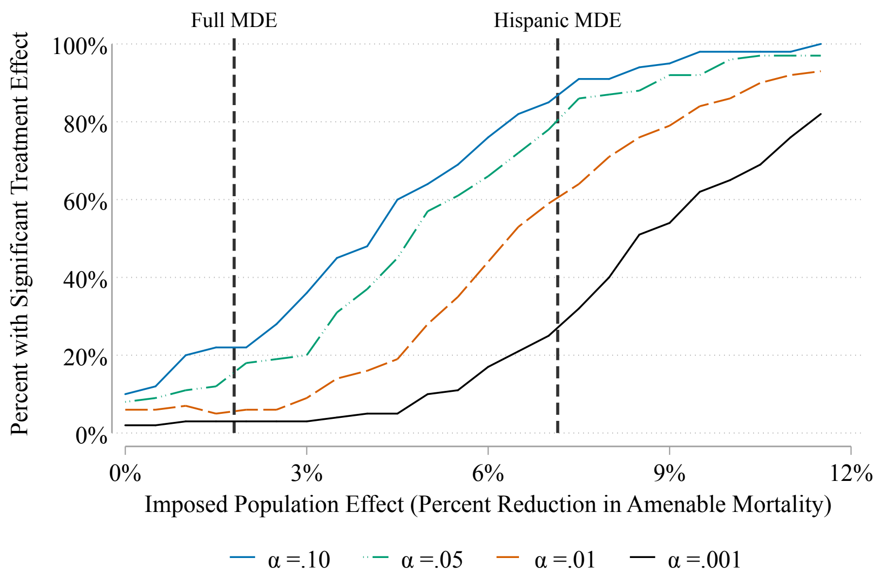


Figure A-22. Power Analysis for Hispanics: DD and Triple Differences

Power curves for simulated Medicaid expansion as of January 1, 2012, applied to non-white, non-black Hispanics aged 55-64 during pre-treatment period (2007-2013). Graphs show power (likelihood of detecting a statistically significant effect on amenable mortality, at the indicated confidence levels, for two-tailed test), given imposed "true" population average effect. Curves are based on 1,000 replications of the DD (top graph) and triple difference (bottom graph) regression models used in Table 2, with covariates. In each draw, we select 20 pseudo-treated states at random from the combined set of 41 treated and control states, and remove a fraction of the observed deaths at random from the treated states, where the fraction reflects an imposed treatment effect (for the entire population), and we vary the imposed treatment effect from 0-5% in increments of 0.1%. Curves for α = .10/.05/.01/.001 correspond to 90%/95%/99%/99.9% confidence levels, respectively. Dashed vertical line indicates minimum detectable effect at 95% confidence level, with 80% power for full sample (Full MDE) and for Hispanics (Hispanic MDE).



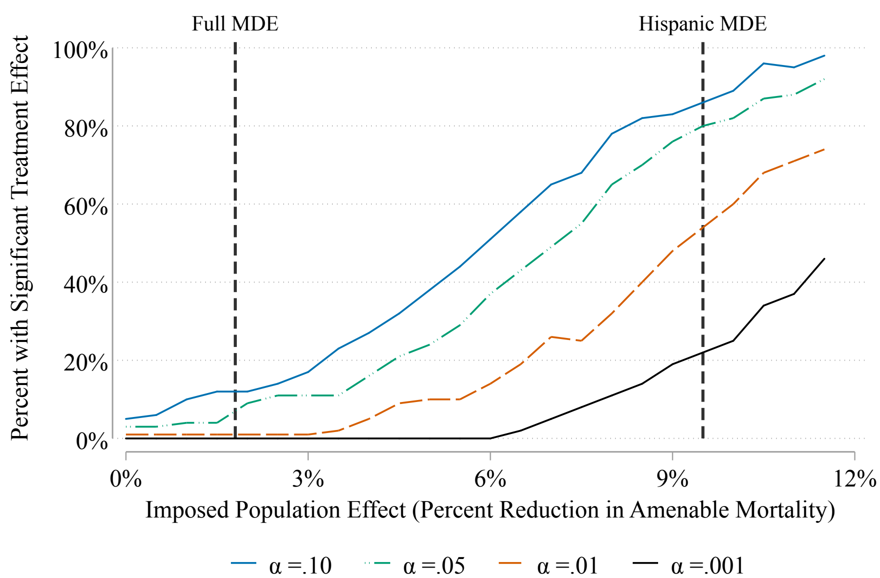


Figure A-23. Power Analysis for Low Education Subsample: DD Design

Power curves for simulated Medicaid expansion as of January 1, 2012, applied to those without a high school education aged 45-64 during pre-treatment period (2007-2013). Demographic data on education is available only for broad age groups (the best available was ages 45-64) so we present only DD and not triple difference results. Graphs show power (likelihood of detecting a statistically significant effect on amenable mortality, at the indicated confidence levels, for two-tailed test), given imposed "true" population average effect. Curves are based on 1,000 replications of the DD regression model used in Table 2, with covariates. In each draw, we select 20 pseudo-treated states at random from the combined set of 41 treated and control states, and remove a fraction of the observed deaths at random from the treated states, where the fraction reflects an imposed treatment effect (for the entire population), and we vary the imposed treatment effect from 0-5% in increments of 0.1%. Curves for α = .10/.05/.01/.001 correspond to 90%/95%/99%/99.9% confidence levels, respectively. Dashed vertical lines indicates minimum detectable effect at 95% confidence level, with 80% power for full sample (Full MDE) and for low-education subsample (Low Educ. MDE).

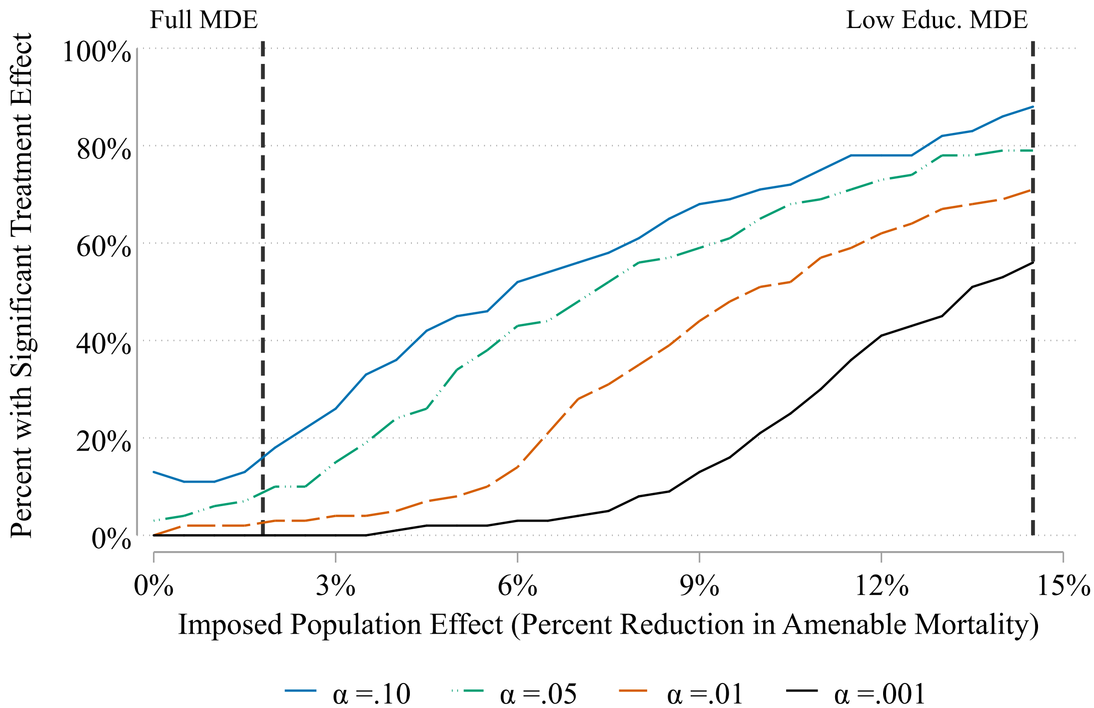


Figure A-24. Uninsurance Rate by Single Year of Age

Source: Authors’ calculations from American Community Survey 2009, 2013 and 2015

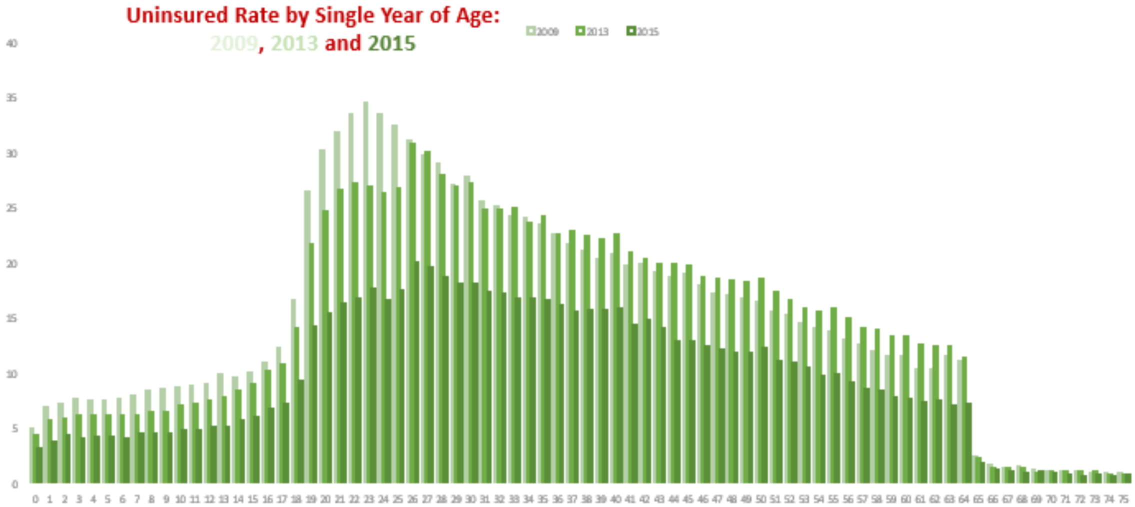


Figure A-25. Difference in Uninsurance Rate from 2012 to 2016 by Expansion Status

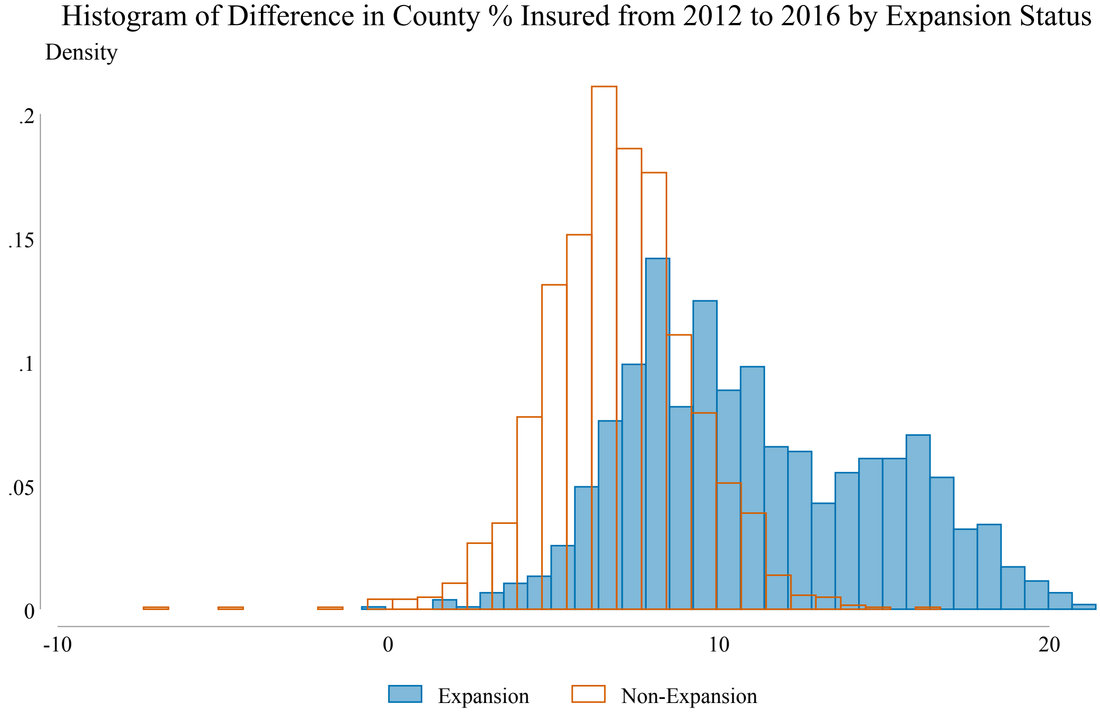
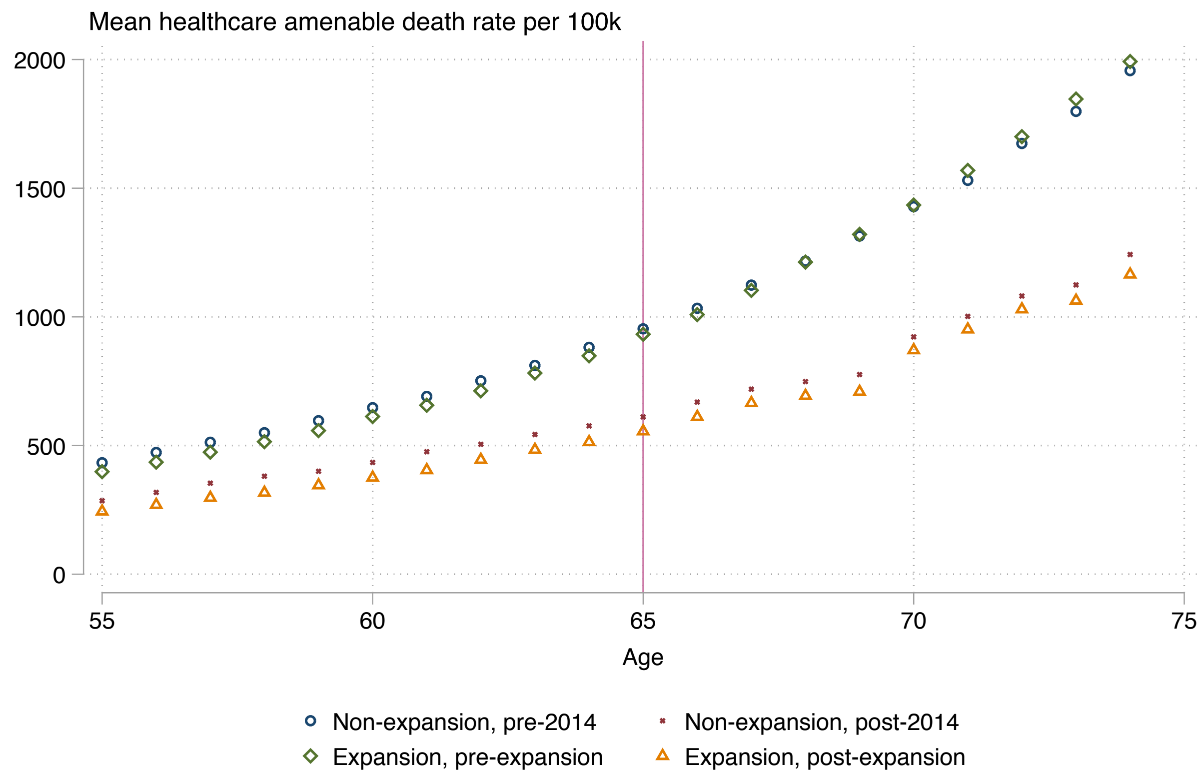


Figure A-26. Changes in mortality by single year of age

Mean health care amenable death rate per 100,000 by single year of age are reported for both expansion and non-expansion states before and after expansion. Difference across time (pre-2014 to post-2014 for non-expansion states; and pre-expansion to post-expansion in expansion states) illustrate that the death rate of each single year of age in expansion states have reduced relative to each analogous group in non-expansion states. The differences across age groups (55-64 v 65-74) illustrate that this improvement was not limited to those eligible for Medicaid. That is, the improvement occurred for Medicare enrollees as well. Thus even with disaggregated data by age, we do not find conclusive evidence of a Medicaid expansion impact on the mortality rate for the near elderly (55-64).

Source: Author calculations from restricted access mortality files.



Appendix References

Abadie, Alberto, Diamond, Alexis and Hainmueller, Jens. 2010. Synthetic control methods for comparative case studies: estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*, 105 (490): 493-505.

Baicker, Katherine, Sarah Taubman, Heidi Allen, Mira Bernstein, Jonathan Gruber, Joseph Newhouse, Eric Schneider, Bill Wright, Alan Zaslavsky, and Amy Finkelstein, 2013. The Oregon Experiment – Effects of Medicaid on Clinical Outcomes, 368 (18) *New England Journal of Medicine* 1713-1722.

Courtemanche, Charles, James Marton, Benjamin Ukert, Aaron Yelowitz, and Daniela Zapata. 2017. Impacts of the Affordable Care Act on health insurance coverage in Medicaid expansion and non-expansion states. J*ournal of Policy Analysis and Management*, Vol. 36, No. 1, 178–210

Goldman, Dana, Jayantha Bhattacharya, McCaffrey DF, Duan N, Arleen Leibowitz et al. 2001. The effect of insurance on mortality in an HIV+ population in care. *J. Am. Stat. Assoc.* 96(455):883–894

Finkelstein, Amy and Robin McKnight. 2008. What Did Medicare Do? The Initial Impact of Medicare on Mortality and Out of Pocket Medical Spending, 92 *Journal of Public Economics* 1644-1668.

Soni, Aparna, 2016. Synthetic Control Method with Multiple Treatment Units , unpublished research note.

Soni, Aparna, John Cawley, Lindsay Sabik, and Kosali Simon. 2018a. Effect of Medicaid Expansions of 2014 on Overall and Early-Stage Cancer Diagnoses. *American Journal of Public Health* 108, no.2. Feb 1) pp.216-218.

Soni, A., K. Simon, L. Sabik, and S. Sommers. 2018b Changes in Insurance Coverage Among Cancer Patients Under the Affordable Care Act . *JAMA Oncology,* Jan 2018 Vol 4 No 1 p.122.

Xu, Y., 2017. Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects, *Political Analysis* 25: 57-76.

Example Simulated Power Analysis from Black, Hollingsworth, Nunes, and Simon (2019)

Alex Hollingsworth

3 January 2019

This is an example of the type of simulated power analysis done in Black et al. (2019). This example is done with publicly available data. You can find the code, data, and output for this example hosted on Alex's GitHub page https://github.com/hollina/health_insurance_and_mortality.

This set-up is designed to mimic a typical DiD setting. Here we will compare 23 randomly chosen treated states to 18 randomly chosen control states. We will impose a series of treatment effects that gradually increase in magnitude and report whether or not these imposed treatment effects are detectable. We will vary the set of randomly chosen treated states. We will calculate the minimum detectable effect size at various power and significance levels. We will also explore a measure of believability, which is based upon Gelman and Carlin (2014) measures of sign and magnitude error.

In this simple design we used 5 years of pre-expansion data and 3 years of post-expansion data. Both state and year fixed-effects are included. Regressions are weighted by state-population and standard errors will be clustered at the state-level. The dependent variable will be the natural log of the all-cause non-elderly mortality rate per 100,000.

This code is simply an example of our simulated power analysis and is not an attempt to identify the impact of Medicaid expansion on mortality. Importantly, changing the research design (e.g. adding control variables, shifting to the county-level, changing the cause of death, using propensity score weights, or using a synthetic control estimator) will impact power. Our approach could be easily modified to accommodate any of these alternative research designs. Any improvements to the research design will very likely increase power and decrease the minimum detectable effect size.

Initial Set-up

Here we will set-up the power analysis and choose various required parameters/options.

First we clear the memory

```
. clear all
```

Choose the number of datasets we want to compose each estimate. For example, if we choose 2, then two sets of psuedo-treated states will be drawn and the power analysis will be conducted twice for each effect size; once for each set of pseudo-treated states and effect size pair.

```
. local max_dataset_number  = 1000
```

Pick the number of psuedo-post-expansion years

```
. local number_post_years = 3

. local last_year = 2013-`number_post_years'+1
```

Set number of psuedo-pre-expansion years

```
. local number_pre_years = 5

. local first_year = `last_year'-`number_pre_years'
```

Set effect size step and max value in percent terms (0-1)

```
. local step_size = .0025 // Quarter of a percent

. local end_value = .05 // End at 5%
```

Create a local macro from the choices above

```
. local step_macro

. forvalues x = 0(`step_size')`end_value' {
  2.     local step_macro `step_macro'  `x'
  3. }
```

Determine the length of the macro above, so percent complete can be displayed later

```
. local num :  word count `step_macro'

. local num = `num'
```

Calculate the max number of rows so percent complete can be displayed later

```
. local max_row = `max_dataset_number'*`num'
```

Create excel sheet to store results from simulation. Note: I have $dropbox set via my profile.do to point to my Dropbox folder.

```
. putexcel set
"$dropbox/health_insurance_and_mortality/state_level_public_dat
> a_example/output/power_simulation_results.xlsx", replace
```

Initialize cells names in excel sheet

```
. putexcel A1 = ("dependent_variable")
file
/Users/hollinal/Dropbox/health_insurance_and_mortality/state_level_public
```

```
> _data_example/output/power_simulation_results.xlsx saved

. putexcel B1 = ("controls")
file
/Users/hollinal/Dropbox/health_insurance_and_mortality/state_level_public
> _data_example/output/power_simulation_results.xlsx saved

. putexcel C1 = ("weight")
file
/Users/hollinal/Dropbox/health_insurance_and_mortality/state_level_public
> _data_example/output/power_simulation_results.xlsx saved

. putexcel D1 = ("treated_states")
file
/Users/hollinal/Dropbox/health_insurance_and_mortality/state_level_public
> _data_example/output/power_simulation_results.xlsx saved

. putexcel E1 = ("effect_size")
file
/Users/hollinal/Dropbox/health_insurance_and_mortality/state_level_public
> _data_example/output/power_simulation_results.xlsx saved
. putexcel F1 = ("deaths_reduced_per_year")
file
/Users/hollinal/Dropbox/health_insurance_and_mortality/state_level_public
> _data_example/output/power_simulation_results.xlsx saved

. putexcel G1 = ("total_deaths_reduced")
file
/Users/hollinal/Dropbox/health_insurance_and_mortality/state_level_public
> _data_example/output/power_simulation_results.xlsx saved
. putexcel H1 = ("coef")
file
/Users/hollinal/Dropbox/health_insurance_and_mortality/state_level_public
> _data_example/output/power_simulation_results.xlsx saved

. putexcel I1 = ("se")
file
/Users/hollinal/Dropbox/health_insurance_and_mortality/state_level_public
> _data_example/output/power_simulation_results.xlsx saved

. putexcel J1 = ("df")
file
/Users/hollinal/Dropbox/health_insurance_and_mortality/state_level_public
> _data_example/output/power_simulation_results.xlsx saved
```

Import and clean mortality data

Import data extracted from CDC wonder. All cause mortality 0-64 by state and year. The data were gathered on 1 January 2019.

```
. import delimited
"$dropbox/health_insurance_and_mortality/state_level_public
> _data_example/data/Multiple Cause of Death, 1999-2017.txt"
(8 vars, 1,077 obs)
```

Drop total variables

```
. drop if missing(year)
(108 observations deleted)
```

Drop unneeded variables from CDC Wonder

```
. drop notes
```

Drop years after expansion

```
. drop if year>=2014
(204 observations deleted)
```

Drop if year before first desired year

```
. drop if year<`first_year'
(357 observations deleted)
```

Change state name to be state postal code

```
. replace state ="AL" if state=="Alabama"
(8 real changes made)

. replace state ="AK" if state=="Alaska"
(8 real changes made)

. replace state ="AZ" if state=="Arizona"
(8 real changes made)

. replace state ="AR" if state=="Arkansas"
(8 real changes made)

. replace state ="CA" if state=="California"
(8 real changes made)

. replace state ="CO" if state=="Colorado"
(8 real changes made)

. replace state ="CT" if state=="Connecticu "
(0 real changes made)

. replace state ="DE" if state=="Delaware"
(8 real changes made)

. replace state ="DC" if state=="District of Columbia"
(8 real changes made)

. replace state ="FL" if state=="Florida"
(8 real changes made)

. replace state ="GA" if state=="Georgia"
(8 real changes made)

. replace state ="HI" if state=="Hawaii"
(8 real changes made)
```

```
. replace state ="ID" if state=="Idaho"
(8 real changes made)

. replace state ="IL" if state=="Illinois"
(8 real changes made)

. replace state ="IN" if state=="Indiana"
(8 real changes made)

. replace state ="IA" if state=="Iowa"
(8 real changes made)

. replace state ="KS" if state=="Kansas"
(8 real changes made)

. replace state ="KY" if state=="Kentucky"
(8 real changes made)

. replace state ="LA" if state=="Louisiana"
(8 real changes made)

. replace state ="ME" if state=="Maine"
(8 real changes made)

. replace state ="MD" if state=="Maryland"
(8 real changes made)

. replace state ="MA" if state=="Massachusetts"
(8 real changes made)

. replace state ="MI" if state=="Michigan"
(8 real changes made)

. replace state ="MN" if state=="Minnesota"
(8 real changes made)

. replace state ="MS" if state=="Mississippi"
(8 real changes made)

. replace state ="MO" if state=="Missouri"
(8 real changes made)

. replace state ="MT" if state=="Montana"
(8 real changes made)

. replace state ="NE" if state=="Nebraska"
(8 real changes made)

. replace state ="NV" if state=="Nevada"
(8 real changes made)

. replace state ="NH" if state=="New Hampshire"
(8 real changes made)

. replace state ="NJ" if state=="New Jersey"
(8 real changes made)

. replace state ="NM" if state=="New Mexico"
(8 real changes made)

. replace state ="NY" if state=="New York"
(8 real changes made)
```

```
. replace state ="NC" if state=="North Carolina"
(8 real changes made)

. replace state ="ND" if state=="North Dakota"
(8 real changes made)

. replace state ="OH" if state=="Ohio"
(8 real changes made)

. replace state ="OK" if state=="Oklahoma"
(8 real changes made)

. replace state ="OR" if state=="Oregon"
(8 real changes made)

. replace state ="PA" if state=="Pennsylvania"
(8 real changes made)

. replace state ="RI" if state=="Rhode Island"
(8 real changes made)

. replace state ="SC" if state=="South Carolina"
(8 real changes made)

. replace state ="SD" if state=="South Dakota"
(8 real changes made)

. replace state ="TN" if state=="Tennessee"
(8 real changes made)

. replace state ="TX" if state=="Texas"
(8 real changes made)

. replace state ="UT" if state=="Utah"
(8 real changes made)

. replace state ="VT" if state=="Vermont"
(8 real changes made)

. replace state ="VA" if state=="Virginia"
(8 real changes made)

. replace state ="WA" if state=="Washington"
(8 real changes made)

. replace state ="WV" if state=="West Virginia"
(8 real changes made)

. replace state ="WI" if state=="Wisconsin"
(8 real changes made)

. replace state ="WY" if state=="Wyoming"
(8 real changes made)
```

Add expansion status to each state

```
. gen expansion4=0

. label define expansion4 0 "0. Non-expansion" 1 "1. Full expansion" ///
>     2 "2. Mild expansion" 3 "3. Substantial expansion"

. label values expansion4 expansion4
```

```
. local full AZ AR CO IL IA KY MD NV NM NJ ND OH OR RI WV  WA

. foreach x in `full' {
  2.     replace expansion4=1 if state=="`x'"
  3. }
(8 real changes made)
(8 real changes made)
(8 real changes made)
(8 real changes made)
(8 real changes made)
(8 real changes made)
(8 real changes made)
(8 real changes made)
(8 real changes made)
(8 real changes made)
(8 real changes made)
(8 real changes made)
(8 real changes made)
(8 real changes made)
(8 real changes made)
(8 real changes made)

. local mild DE DC MA NY VT

. foreach x in `mild' {
  2.     replace expansion4=2 if state=="`x'"
  3. }
(8 real changes made)
(8 real changes made)
(8 real changes made)
(8 real changes made)
(8 real changes made)

. local medium CA CT HI MN WI

. foreach x in `medium' {
  2.     replace expansion4=3 if state=="`x'"
  3. }
(8 real changes made)
(0 real changes made)
(8 real changes made)
(8 real changes made)
(8 real changes made)
```

Account for mid-year expansions

```
. replace expansion4=1 if state=="MI"  //MI expanded in April 2014
(8 real changes made)

. replace expansion4=1 if state=="NH"  //NH expanded in August 2014
(8 real changes made)

. replace expansion4=1 if state=="PA"  //PA expanded in Jan 2015
(8 real changes made)

. replace expansion4=1 if state=="IN"  //IN expanded in Feb 2015
(8 real changes made)

. replace expansion4=1 if state=="AK"  //AK expanded in Sept 2015
(8 real changes made)

. replace expansion4=1 if state=="MT"  //MT expanded in Jan 2016
```

```
(8 real changes made)

. replace expansion4=1 if state=="LA"  //LA expanded in July 2016
(8 real changes made)
```

Keep only full or non-expansion states

```
. drop if expansion4==2 | expansion4==3
(72 observations deleted)
```

Store number of expansion states

```
. distinct statecode  if expansion4==1

             |        Observations
             |      total   distinct
-------------+----------------------
   statecode |        184         23

. scalar number_expand = r(ndistinct)
```

Save data to be called in power analysis

Save temporary dataset to be called

```
. compress
  variable expansion4 was float now byte
  variable population was double now long
  variable state was str20 now str11
  (5,376 bytes saved)

. save
"$dropbox/health_insurance_and_mortality/state_level_public_data_exampl
> e/temp/temp_data.dta", replace
(note: file
/Users/hollinal/Dropbox/health_insurance_and_mortality/state_level
> _public_data_example/temp/temp_data.dta not found)
file
/Users/hollinal/Dropbox/health_insurance_and_mortality/state_level_public
> _data_example/temp/temp_data.dta saved
```

Run simulated power analysis

Start a timer to show how long this takes

```
. timer on 1
```

Set row number for excel sheet

```
. local row =2
```

Run a loop. Performing the power analysis once for each of the desired number of datasets. The following output is supressed for the html document even though it runs. This is to ensure the document is not too long.

```
. forvalues dataset_number = 1(1)`max_dataset_number'    {
  2.     // Display the dataset number
.     qui di "`dataset_number'"
.     // Open main dataset for analysis
.     qui use
"$dropbox/health_insurance_and_mortality/state_level_public_data
> _example/temp/temp_data.dta", clear
.     // Set seed for reproducibility. We want the seed to be the same within
> a dataset.
.     qui local rand_seed = 1234 + `dataset_number'
  5.     qui set seed   `rand_seed'
.
///////////////////////////////////////////////////////////////////////
> ///////
>     // Generate a random variable for each state, then the first N in rank
w
> ill be
.     // considered expansion states. Where N is # of expansion states
.     qui bysort statecode: gen random_variable = runiform() if _n==1
  7.     qui bysort statecode: carryforward random_variable, replace
.     // Rank the states
.     qui egen rank = group(random_variable)
.     // Given this random ordering of states, assign expansion status to the
> # set above
.     qui gen expansion = 0
 10.     qui replace expansion=1 if rank <=number_expand
.     // Do this same thing for the treatment variable
.     qui gen treatment = 0
 12.     qui replace treatment = 1 if expansion==1 & year>=`last_year'
.     // Create Post variable
.     qui gen post = 0
 14.     qui replace post =1 if year>=`last_year'
.     // Store basic data from regression in excel sheet
.     qui putexcel A`row' = ("all_deaths")
 16.     qui putexcel B`row' = ("no controls")
 17.     qui putexcel C`row' = ("population")
.     // Add list of states to excel sheet
.     qui capture drop test
 19.     qui gen test = ""
.     qui levelsof state if treatment ==1, local(treated_states)
 21.     foreach x in `treated_states' {
 22.         qui replace test = test + ", " + "`x'"
 23.     }
.     qui local state_list `=test[1]'
 25.     qui putexcel D`row' = ("`state_list'")
.     // Generate a death rate with no effect
.     qui gen death_rate = (deaths/population)*100000
.     // Gen order variable
.     qui gen order = _n
.
///////////////////////////////////////////////////////////////////////
> /////
>     // Create a reduced deaths variable by a given percentage using the
bino
> mial for each effect size
```

```
.     qui local counter = 1
.     foreach x in `step_macro' {
 30.         qui gen reduced_deaths_`counter' = 0
 31.         qui replace reduced_deaths_`counter' = rbinomial(deaths,`x') if t
> reatment==1
 32.         qui replace reduced_deaths_`counter'=0 if missing(reduced_deaths_
> `counter')
.         qui gen deaths_`counter' = deaths - reduced_deaths_`counter'
 34.         qui replace deaths_`counter'=0 if missing(deaths_`counter')
.         qui gen death_rate_`counter'= ln((deaths_`counter'/population)*10000
> 0+1)
.         // Store the effect size in excel sheet
.         qui putexcel E`row' = (`x')
.         // Store the number of reduced deaths in excel sheet
.         qui sum reduced_deaths_`counter' if year>=`last_year'
 38.         qui putexcel F`row' = (`r(sum)'/`number_post_years')
 39.         qui putexcel G`row' = (`r(sum)')
.         // Move the row and counter one forward
.         qui local counter = `counter' + 1
 41.         qui local row = `row' + 1
 42.     }
.     // Move the row counter back to the top
.     qui local row = `row' - `num'
. ////////////////////////////////////////////////////////////////////////
> /////
>     // Run regression of treatment on reduced deaths variable for each effec
> t size
.     // Reset the counter
.     qui local counter = 1
.     forvalues counter = 1(1)`num' {
.         qui reghdfe death_rate_`counter' ///
>             treatment ///
>             i.post i.expansion ///
>             [aweight=population] ///
>             ,  absorb(statecode year)  vce(cluster statecode)
.         // Store results
.         qui putexcel H`row' =(_b[treatment])
 48.         qui putexcel I`row' = (_se[treatment])
 49.         qui putexcel J`row' =(`e(df_r)')
.         // Display Percent Complete
.         qui di "/////////////////////////////////////////////////////////////
> //////////"
 51.         qui di "/////////////////////////Percent Complete/////////////
> ////////////"
 52.         qui di ((`row'-1)/`max_row')*100
 53.         qui di "//////////////////////////////////////////////////////////
> /////////////"
.         qui local row = `row' + 1
 55.         qui local counter = `counter' + 1
 56.     }
 57. }
```

Stop timer

```
. timer off 1

. timer list
   1:  79905.50 /        1 =   79905.5020
```

Erase temporary dataset used for analysis

```
. erase
"$dropbox/health_insurance_and_mortality/state_level_public_data_examp
> le/temp/temp_data.dta"
```

Import and clean results from simulated power analysis

Import simulation results

```
. import excel
"$dropbox/health_insurance_and_mortality/state_level_public_dat
> a_example/output/power_simulation_results.xlsx", sheet("Sheet1") firstrow
cl
> ear
```

Calculate z-scores and p-values

```
. gen z_score = abs(((coef - 0)/se))

. gen p_value =  2*ttail(df,z_score)
```

Calculate indicator for power threshold for each observation

```
. gen power_10 = 0

. gen power_05 = 0

. gen power_01 = 0

. gen power_001 = 0
. replace power_10 =  1 if p_value<= .1
(12,536 real changes made)

. replace power_05 =  1 if p_value<= .05
(11,065 real changes made)

. replace power_01 =  1 if p_value<= .01
(8,209 real changes made)

. replace power_001 = 1 if p_value<= .001
(4,872 real changes made)
```

Calculate a count variable

```
. gen count = 1
```

Make sign error

```
. gen s_error_10 = 0

. replace s_error_10 =1 if power_10==1 & coef>=0
(174 real changes made)

. gen s_error_05 = 0

. replace s_error_05 =1 if power_05==1 & coef>=0
(85 real changes made)

. gen s_error_01 = 0

. replace s_error_01 =1 if power_01==1 & coef>=0
(17 real changes made)

. gen s_error_001 = 0

. replace s_error_001 =1 if power_001==1 & coef>=0
(0 real changes made)
. replace s_error_10 =. if effect_size==0
(1,000 real changes made, 1,000 to missing)

. replace s_error_05 =. if effect_size==0
(1,000 real changes made, 1,000 to missing)

. replace s_error_01 =. if effect_size==0
(1,000 real changes made, 1,000 to missing)

. replace s_error_001 =. if effect_size==0
(1,000 real changes made, 1,000 to missing)
```

Make magnitude error

```
. gen m_error = abs(coef/effect_size)
(1,000 missing values generated)

. gen m_error_10 = m_error
(1,000 missing values generated)

. replace m_error_10 = . if power_10==0
(6,628 real changes made, 6,628 to missing)

. gen m_error_05 = m_error
(1,000 missing values generated)

. replace m_error_05 = . if power_05==0
(8,030 real changes made, 8,030 to missing)

. gen m_error_01 = m_error
(1,000 missing values generated)

. replace m_error_01 = . if power_01==0
```

```
(10,820 real changes made, 10,820 to missing)

. gen m_error_001 = m_error
(1,000 missing values generated)

. replace m_error_001 = . if power_001==0
(14,130 real changes made, 14,130 to missing)
```

Generate Beliveabilitiy

```
. gen believe_10 = 0

. replace  believe_10 = 1 if power_10 ==1 & s_error_10==0 & m_error_10<=2
(11,081 real changes made)
. gen believe_05 = 0

. replace  believe_05 = 1 if power_05 ==1 & s_error_05==0 & m_error_05<=2
(9,934 real changes made)
. gen believe_01 = 0

. replace  believe_01 = 1 if power_01 ==1 & s_error_01==0 & m_error_01<=2
(7,502 real changes made)
. gen believe_001 = 0

. replace  believe_001 = 1 if power_001 ==1 & s_error_001==0 & m_error_001<=2
(4,519 real changes made)
```

Collapse by effect size to calculate power, % sign error, average magnitude error and % believable

```
. collapse (sum) count *power_* *s_error_* *believe_* (mean) *m_error_*,
by(ef
> fect_size)
```

Generate sign error ratio, rather than raw count

```
. replace s_error_10 = (s_error_10/power_10)*100
(5 real changes made)

. replace s_error_05 = (s_error_05/power_05)*100
(4 real changes made)

. replace s_error_01 = (s_error_01/power_01)*100
(2 real changes made)

. replace s_error_001 = (s_error_001/power_001)*100
(0 real changes made)
. replace s_error_10 = . if effect_size==0
(1 real change made, 1 to missing)

. replace s_error_05 = . if effect_size==0
(1 real change made, 1 to missing)

. replace s_error_01 = . if effect_size==0
(1 real change made, 1 to missing)

. replace s_error_001 = . if effect_size==0
```

```
(1 real change made, 1 to missing)
```

Make power and believability out of 100

```
. ds *power* *believe_*
power_10      power_01      believe_10    believe_01
power_05      power_001     believe_05    believe_001

. foreach x in `r(varlist)' {
  2.      replace `x' = (`x'/count)*100
  3. }
(20 real changes made)
(20 real changes made)
(20 real changes made)
(20 real changes made)
(16 real changes made)
(15 real changes made)
(14 real changes made)
(13 real changes made)
```

Make effect size 0-100

```
. replace effect_size=effect_size*100
(19 real changes made)
```

Plot power curves

First determine closest point where the power_05 hits 80%

```
. gen distance_from_80 = (power_05-80)^2

. sort distance_from_80

. sum effect_size in 1

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 effect_size |          1           3           .          3          3

. local mde=`r(mean)'
```

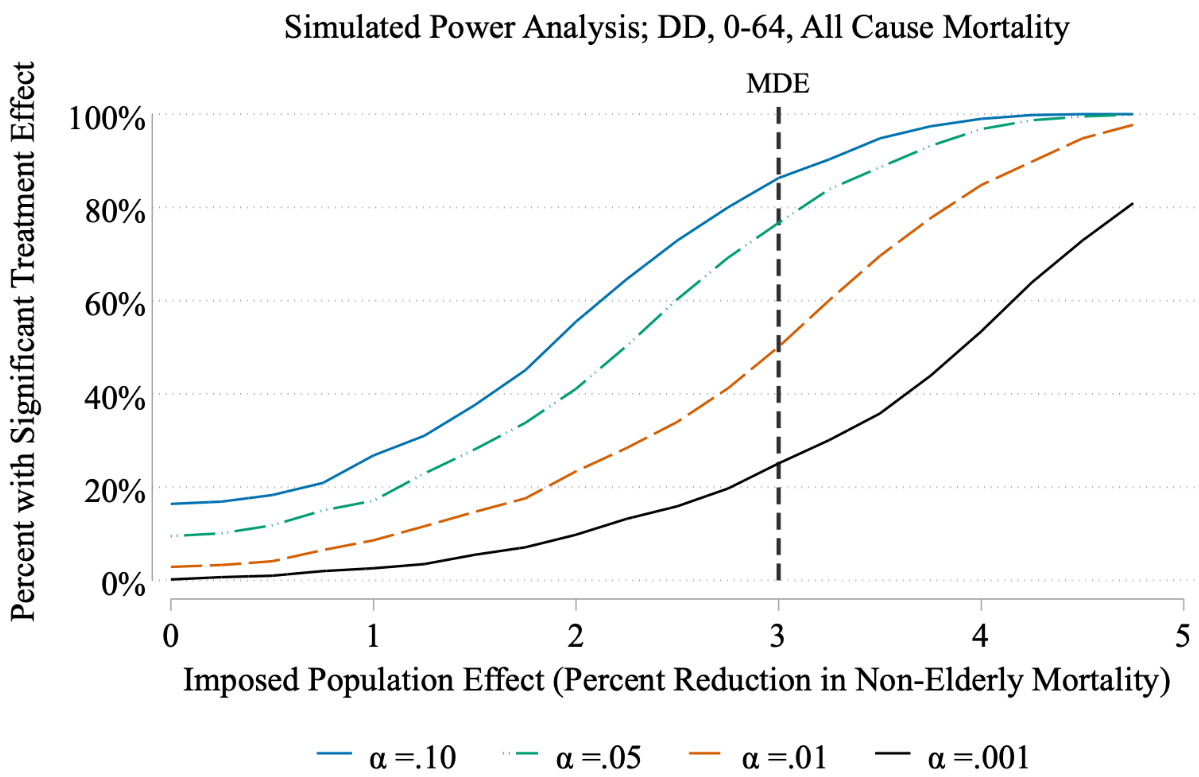
Add label to graph with this MDE

```
. capture drop mde_label

. gen mde_label = ""
(20 missing values generated)

. set obs `=_N+1'
number of observations (_N) was 20, now 21

. replace mde_label = "MDE" in `=_N'
variable mde_label was str1 now str3
```

```
(1 real change made)

. replace effect_size = `mde' in `=_N'
(1 real change made)
. capture drop full_power

. gen full_power = 102.5
```

Plot power curve

```
. sort effect_size

. twoway connected power_10 effect_size ,  lpattern("l") color(sea)
msymbol(no
> ne) mlabcolor(sea) mlabel("") mlabsize(3) mlabpos(11) ///
>     || connected  power_05 effect_size ,  lpattern(".._") color(turquoise)
m
> symbol(none) mlabcolor(turquoise) mlabel("") mlabsize(3) mlabpos(3) ///
>     || connected  power_01 effect_size , lpattern("_") color(vermillion)
msy
> mbol(none) mlabcolor(vermillion) mlabel("") mlabsize(3) mlabpos(3) ///
>     || connected  power_001  effect_size ,  lpattern("l") color(black)
msymb
> ol(none) mlabcolor(black) mlabel("") mlabsize(3) mlabpos(3) ///
>     || scatter full_power effect_size , mlabel(mde_label) msymbol(none)
mlab
> pos(12) mlabsize(3.5) ///
>         xline(`mde', lpattern(dash) lcolor(gs3) lwidth(.5) noextend) ///
>         ytitle("Percent with Significant Treatment Effect", size(4)) ///
>         xtitle("Imposed Population Effect (Percent Reduction in Non-Elderly
> Mortality)", size(4) ) ///
>         xscale(r(0 5)) ///
>         xlabel(, nogrid labsize(4)) ///
>         ylabel(0 "0%" 20 "20%"  40 "40%" 60 "60%" 80 "80%" 100 "100%",gmax
n
> oticks labsize(4)) ///
>         legend(order( 1 2 3 4) pos(6) col(4) ///
>             label(1 "{&alpha} =.10") label(2 "{&alpha} =.05") ///
>             label(3 "{&alpha} =.01") label(4 "{&alpha} =.001") size(4)) ///
>             title("Simulated Power Analysis; DD, 0-64, All Cause Mortality"
> " ", size(4))
.     graph export
"$dropbox/health_insurance_and_mortality/state_level_public
> _data_example/scripts/markdown/simulated_power_analysis.png",  replace
width
> (800)
(file
/Users/hollinal/Dropbox/health_insurance_and_mortality/state_level_publi
> c_data_example/scripts/markdown/simulated_power_analysis.png written in PNG
> format)
```

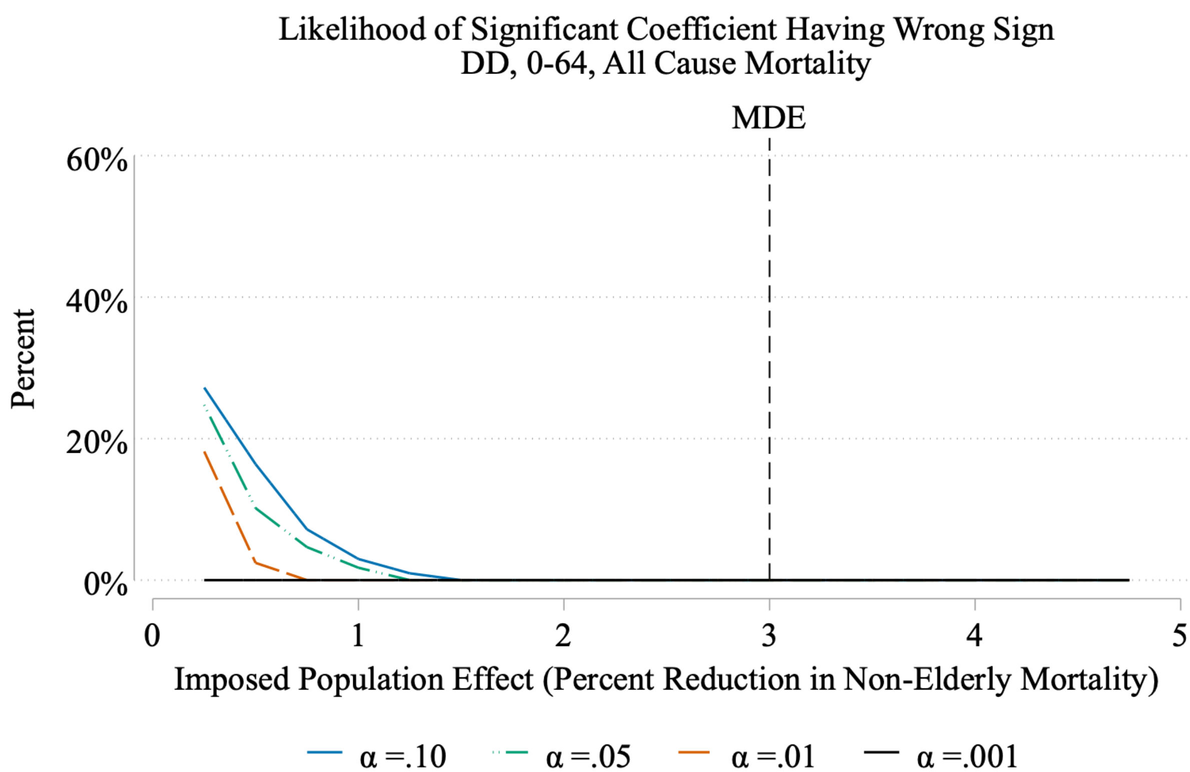


Simulated Power Analysis; DD, 0-64, All Cause Mortality

Plot sign error

```
. sum s_error_10

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
  s_error_10 |         19     2.88117    7.122931          0   27.21893

. gen  s_error_label= 62.5

. twoway connected s_error_10 effect_size ,  lpattern("l") color(sea)
msymbol(
> none) mlabcolor(sea) mlabel("") mlabsize(3) mlabpos(11) ///
>     || connected s_error_05 effect_size  ,  lpattern("..._")
color(turquoise)
>  msymbol(none) mlabcolor(turquoise) mlabel("") mlabsize(3) mlabpos(3) ///
>     || connected s_error_01 effect_size  , lpattern("_") color(vermillion)
m
> symbol(none) mlabcolor(vermillion) mlabel("") mlabsize(3) mlabpos(3) ///
>     || connected s_error_001  effect_size  ,  lpattern("l") color(black)
msy
> mbol(none) mlabcolor(black) mlabel("") mlabsize(3) mlabpos(3) ///
>     || scatter s_error_label effect_size  , mlabel(mde_label) msymbol(none)
>  mlabpos(12) mlabsize(4)  ///
>     ytitle("Percent", size(4)) ///
>         xtitle("Imposed Population Effect (Percent Reduction in Non-Elderly
> Mortality)", size(4)) ///
>         legend(size(4) order(1 2 3 4) pos(6) col(4) label(1 "{&alpha}
=.10")
```

```
>  label(2 "{&alpha} =.05") label(3 "{&alpha} =.01") label(4 "{&alpha}
=.001")
> ) ///
>          xscale(r(0 5)) ///
>          xline(`mde', lpattern(dash) lcolor(grey) noextend) ///
>          xlabel( , nogrid labsize(4)) ///
>          ylabel(0 "0%" 20 "20%"  40 "40%" 60 "60%",gmax noticks labsize(4))
/
> //
>          title("Likelihood of Significant Coefficient Having Wrong Sign"
"DD,
>  0-64, All Cause Mortality" " ", size(4))
(note:  named style grey not found in class color, default attributes used)
.      graph export
"$dropbox/health_insurance_and_mortality/state_level_public
> _data_example/scripts/markdown/s_error.png", replace width(800)
(file
/Users/hollinal/Dropbox/health_insurance_and_mortality/state_level_publi
> c_data_example/scripts/markdown/s_error.png written in PNG format)
```

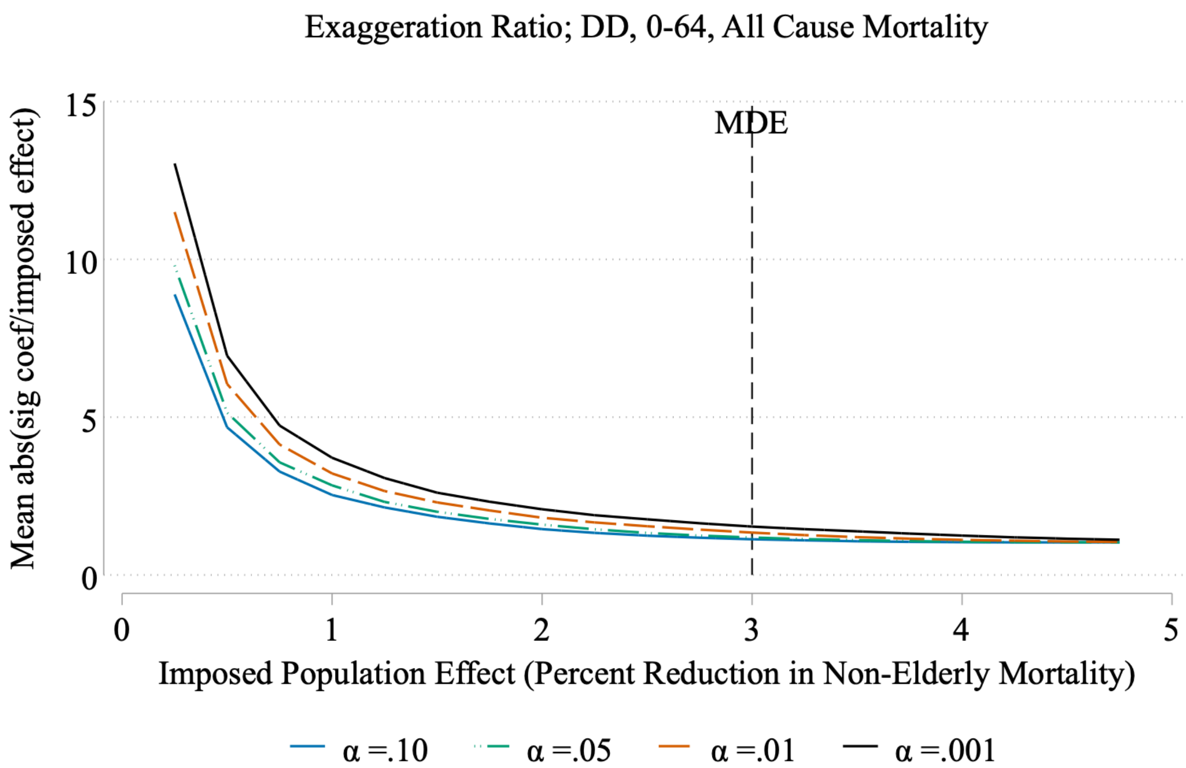


Likelihood of Significant Coefficient Having Wrong Sign DD, 0-64, All Cause Mortality

Plot magnitude error

```
. sum m_error_001

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 m_error_001 |         19    2.851967    2.875496   1.114921   13.03762

. gen  height= `r(max)'*1.05
```

```
. twoway connected m_error_10 effect_size ,  lpattern("l") color(sea)
msymbol(
> none) mlabcolor(sea) mlabel("") mlabsize(3) mlabpos(11) ///
>     || connected m_error_05 effect_size ,  lpattern(".._") color(turquoise)
> msymbol(none) mlabcolor(turquoise) mlabel("") mlabsize(3) mlabpos(3) ///
>     || connected m_error_01 effect_size , lpattern("_") color(vermillion)
ms
> ymbol(none) mlabcolor(vermillion) mlabel("") mlabsize(3) mlabpos(3) ///
>     || connected m_error_001  effect_size ,  lpattern("l") color(black)
msym
> bol(none) mlabcolor(black) mlabel("") mlabsize(3) mlabpos(3) ///
>     || scatter height effect_size , mlabel(mde_label) msymbol(none)
mlabpos
> (12) mlabsize(4)  ///
>     ytitle("Mean abs(sig coef/imposed effect)", size(4)) ///
>         xtitle("Imposed Population Effect (Percent Reduction in Non-Elderly
> Mortality)", size(4)) ///
>         legend(size(4) order(1 2 3 4) pos(6) col(4) label(1 "{&alpha}
=.10")
>  label(2 "{&alpha} =.05") label(3 "{&alpha} =.01") label(4 "{&alpha}
=.001")
> ) ///
>         xscale(r(0 5)) ///
>         xline(`mde', lpattern(dash) lcolor(grey) noextend) ///
>         xlabel(, nogrid labsize(4)) ///
>         ylabel(, gmax noticks labsize(4)) ///
>         title("Exaggeration Ratio; DD, 0-64, All Cause Mortality"  " ",
size
> (4))
(note:  named style grey not found in class color, default attributes used)
.         graph export
"$dropbox/health_insurance_and_mortality/state_level_pu
> blic_data_example/scripts/markdown/m_error.png", replace width(800)
(file
/Users/hollinal/Dropbox/health_insurance_and_mortality/state_level_publi
> c_data_example/scripts/markdown/m_error.png written in PNG format)
```

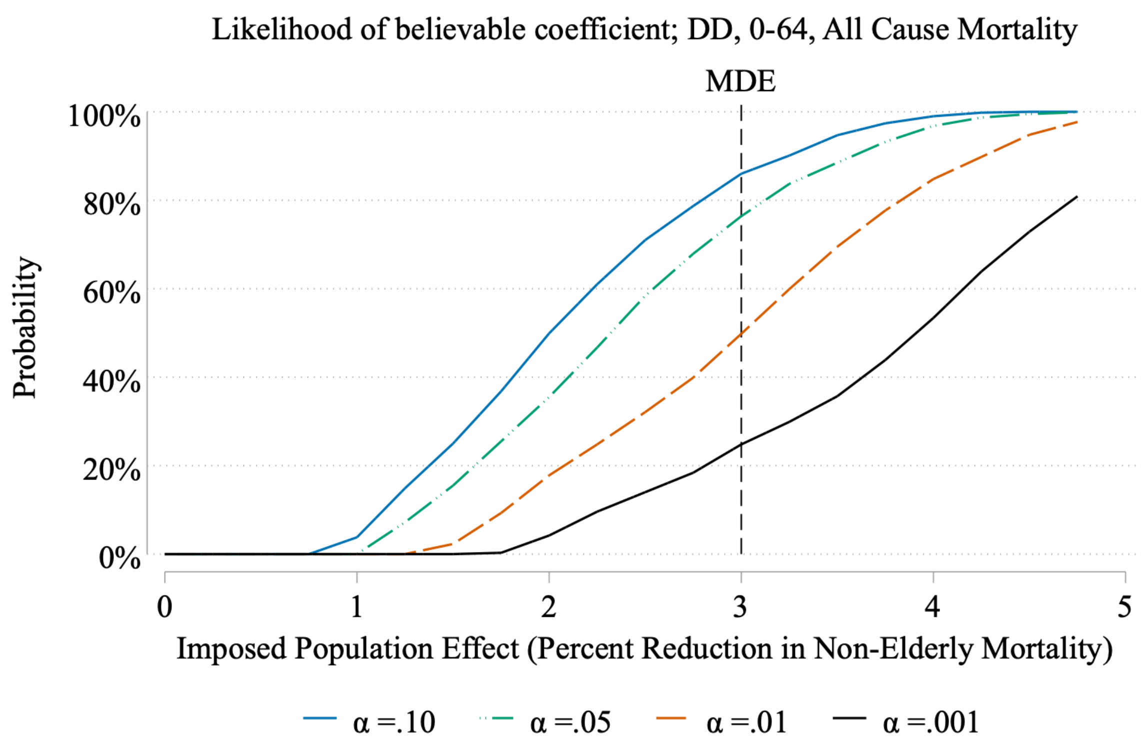


Exaggeration Ratio; DD, 0-64, All Cause Mortality

Plot believability

```
. twoway connected believe_10 effect_size ,  lpattern("l") color(sea)
msymbol(
> none) mlabcolor(sea) mlabel("") mlabsize(3) mlabpos(11) ///
>     || connected believe_05 effect_size ,  lpattern(".._") color(turquoise)
> msymbol(none) mlabcolor(turquoise) mlabel("") mlabsize(3) mlabpos(3) ///
>     || connected believe_01 effect_size , lpattern("_") color(vermillion)
ms
> ymbol(none) mlabcolor(vermillion) mlabel("") mlabsize(3) mlabpos(3) ///
>     || connected believe_001  effect_size ,  lpattern("l") color(black)
msym
> bol(none) mlabcolor(black) mlabel("") mlabsize(3) mlabpos(3) ///
>     || scatter full_power effect_size , mlabel(mde_label) msymbol(none)
mlab
> pos(12) mlabsize(4)  ///
>     xtitle("Imposed Population Effect (Percent Reduction in Non-Elderly
Mort
> ality)", size(4)) ///
>          legend(size(4) order(1 2 3 4) pos(6) col(4) label(1 "{&alpha}
=.10")
>  label(2 "{&alpha} =.05") label(3 "{&alpha} =.01") label(4 "{&alpha}
=.001")
> ) ///
>                ytitle("Probability", size(4)) ///
>          xscale(r(0 5)) ///
>          xline(`mde', lpattern(dash) lcolor(grey) noextend) ///
>          xlabel(, nogrid labsize(4)) ///
```

```
>         ylabel(0 "0%" 20 "20%"  40 "40%" 60 "60%" 80 "80%" 100 "100%",gmax
n
> oticks labsize(4)) ///
>         title("Likelihood of believable coefficient; DD, 0-64, All Cause
Mor
> tality" " ", size(4))
(note:  named style grey not found in class color, default attributes used)
.       graph export
"$dropbox/health_insurance_and_mortality/state_level_publi
> c_data_example/scripts/markdown/believable.png", replace width(800)
(file
/Users/hollinal/Dropbox/health_insurance_and_mortality/state_level_publi
> c_data_example/scripts/markdown/believable.png written in PNG format)
```



Likelihood of believable coefficient; DD, 0-64, All Cause Mortality

Conclusion

Using this simple example, we can see that for this simple research design the minimum mortality reduction that is believable, well-powered, and significant at the 5% level is around 3%. Changing the research design (e.g. adding control variables, shifting to the county-level, changing the cause of death) would certainly impact power.

This simple research design is a DiD comparing 23 random treated states to 18 random control states. In this simple design we used 5 years of pre-expansion data and 3 years of post-expansion data. Both state and year fixed-effects were included. Regressions were weighted by state-population and standard errors were clustered at the state-level. The dependent variable was the natural log of the all-cause non-elderly mortality rate per 100,000.