

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

August 2018

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

[NBER working paper citation to come]

Northwestern University, Pritzker Law School

Law and Economics Research Paper No. 18-xx

Northwestern University, Institute for Policy Research

Working Paper 2018-xx

This paper can be downloaded from the SSRN electronic library at:

<http://ssrn.com/abstract=3xxxxxx>

The Online Appendix can be downloaded without charge from SSRN at:

<http://ssrn.com/abstract=3yyyyyy>

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

Bernard Black, Alex Hollingsworth, Leticia Nunes, and Kosali Simon*

Thursday, August 16, 2018

Abstract. A large literature examines the effect of health insurance on mortality. We add to this literature by exploiting quasi-experimental variation provided by the Affordable Care Act (ACA) expansions in health insurance coverage. We make two main contributions. First, using various differences in differences (DiD) and triple difference designs to compare mortality trends in Medicaid expansion and non-expansion states, as well as other forms of identification using all ACA insurance expansions, we find no convincing evidence that ACA expansions have changed mortality for non-elderly adults. However, confidence intervals are large, thus our results should not be interpreted as evidence that health insurance has no effect on overall mortality for this age group. Second, we provide a simulation-based power analysis, showing that even the 50-state natural experiment provided by the ACA is underpowered to detect plausibly sized mortality effects in available datasets, and discuss data element and sample size needs for the literature to advance. Our power analysis, which applies pseudo-shocks in the pre-treatment period, can serve as a guide for other natural-experiment studies where assessing plausible effect sizes and exploring statistical power can inform research design and increase the validity of reported results.

* Black: Nicholas J. Chabreja Professor at Northwestern University, Pritzker Law School and Kellogg School of Management, Northwestern University; Hollingsworth: School of Public and Environmental Affairs, Indiana University; Nunes: Department of Economics, Fundação Getulio Vargas,. Simon: School of Public and Environmental Affairs, Indiana University and NBER. We thank Ted Joyce, Robert Kaestner, [***others to come**] and participants in workshops at Bar-Ilan University Conference on Law and Big Data (May 2018); Interdisciplinary Center Herzliya (May 2018); ASSA 2018; Stanford University Health Economics Seminar (May 2018); ASHEcon (2018); APPAM (2017); Texas A&M (May 2017); [***and to come**].

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 careful 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

Some 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). Other studies examine the effect of insurance on non-physical health outcomes, such as mental health stress levels and financial health (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 natural experiment studies of low-frequency outcomes. Our study examines this relationship using mortality microdata from 1999-2015. We use both difference-in-differences (DiD) and triple-difference/age discontinuity approaches to study the effect of state Medicaid expansions and the ACA in total on mortality. 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,

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.”

diabetes, respiratory disease); and sociodemographic factors at the individual (gender, race/ethnicity, and education) and the county (baseline percent uninsured, 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 limiting the age bands makes the above and below-65 groups more comparable. This follows the approach taken in 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. One reason for this “null result” is that our “first stage” is weak: the identifying variation (the relative change in uninsurance rates for Medicaid expansion versus non-expansion states) is a small fraction of the population. The increase in health insurance coverage attributable to Medicaid expansion depends on the population under study, but is only around 5% even when we hone in on low-educated populations; for the full sample of non-elderly adults is only around 1% of the population. 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 any effects of health insurance on mortality are likely to emerge only over time.

Our second contribution is to use simulation-based power analysis, to assess whether ACA expansion effects of plausible size can be reliably detected with the available datasets. We conclude that even the 50-state natural experiment provided by the ACA is severely underpowered to detect plausible-sized effects on mortality. It will be extremely challenging for a study such as ours to reliably detect effects of insurance coverage on mortality unless these data can be linked at the individual level to large-sample data on health, income, and perhaps longitudinal insurance status, thus allowing us to identify sufficiently large subsamples with a larger first stage and/ or a higher sensitivity of health and mortality to healthcare use. Even with such hypothetical data, only fairly large effects of health insurance on mortality could be reliably detected.

We estimate power in our pre-treatment period data (pre 2014) by first applying a pseudo-shock to health insurance rates at the beginning of 2012 as if the ACA expansion had occurred

then. We choose pseudo-treated states at random, apply pseudo mortality shocks of different sizes to those states (by randomly removing deaths from our mortality data) and repeat this process 1,000 times. The repeated re-randomization across states effectively converts the non-parallel pre-treatment trends, for which we find evidence in the data, into additional noise, which reduces power but does not lead to bias. We then assess the likelihood that these pseudo shocks we introduce in 2012-2013 would be detected, using methods similar to our actual specifications. 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 power) is about 0.023. Together with a 0.012 first-stage this implies a roughly 200% drop in mortality among the newly insured. The DD and triple difference models have similar power. Power does not improve when we examine vulnerable subgroups: non-parallel trends remain common and the gain in power from a higher first-stage and a higher base mortality rate is offset by smaller sample sizes.

To put into perspective the implied although obviously impossible 200% minimum detectable effect of gaining insurance on the mortality of the newly insured, the introduction of sulfa drugs reduced maternal mortality by 24-36% (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%]).² 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

² 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.

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.

A further implication from the power analysis is that, if significant effects of expanding health insurance eligibility on general adult mortality are found as in some prior studies these are likely to be false positives, reflecting the “luck” of a draw from the right tail of a probability distribution or other ways that shock-based studies can produce false positives, including failure to adequately balance treated and control units or address non-parallel trends, specification searches, and “file-drawer bias.” McCrary, Christensen and Fanelli (2016) propose a minimum t -statistic of around 3 to correct for file-drawer bias alone.

Power analyses are common in the design stage of a randomized trial; researchers use them to ensure that the trial does not “fail” due to inadequate sample size. They are rare, however, for DiD and other observational studies. We propose that many shock-based studies with panel data would benefit from assessing plausible effect sizes and including power analyses in an explicit “design stage”, thus reducing the chance of inadvertently publishing false positive results (Rubin, 2008).

For example, we find non-parallel 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 using persons aged 65-74 years as a within-state control group reduces this problem, but non-parallel trends remain for blacks and Hispanics (Figure 4). DD and triple difference regressions ignore these non-parallel trends. As a result, we find implausibly large, statistically significant effects of ACA expansion on mortality for blacks and Hispanics, in both DD and triple difference specifications. The power analysis reduces the temptation to interpret these significant coefficients as robust results.

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) may experience lower marginal gains from insurance than the already insured; (ii) emergency care and substantial healthcare access for vulnerable populations

were already provided through prior policy interventions; and (iii) access to health insurance does not equate to access to healthcare, as even the uninsured consume substantial healthcare, and 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 emerging in 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. Part III provides an overview of the conceptual concerns that inform our analysis. The past 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 IV summarizes the ACA insurance expansions. Part V describes our data and presents summary statistics. Part VI summarizes our empirical approach. Part VII presents full-sample results.

In Part VIII, we search for evidence of heterogeneous effects for different subpopulations. Confidence intervals are large and no statistically convincing evidence points to detectable effects even in subsamples where we would expect effects to be more likely. Part IX 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 X concludes.

II. Prior Research

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 both morbidity and mortality, physical and mental health, elderly adults, non-elderly adults, pregnant women, and children/infants, short and long run effects, as well as specific diseases and demographic subpopulations.

Our second contribution focuses on the statistical power of insurance-to-mortality studies to detect effects of plausible size. The use of a formal power analysis in an observational study is

generalizable to other settings. Power analyses are often used in randomized trial design to determine sample size, and are also used in grant applications for observational studies, but are rarely if ever reported in published research. Exceptions include [*cites to come].

For the first aim, we focus on the effect of health insurance on health for the general adult population.³ Historically, the first rigorous evidence of effect of health insurance on health 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; their point estimate and 95% CI are [-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 change in health insurance policy 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, and also find no significant effect of turning 65 on population

³ In early research using a natural experiment, Currie and Gruber (1996a,b) find that Medicaid expansions of the late 1980s and early 1990s reduced infant mortality by 8% and all-cause child mortality by 5%. Currie and Gruber (1997) find that neonatal mortality was improved if the mother resided close to NICU unit. 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. This finding is puzzling since emergency care is widely available regardless of insurance, and has been mandatory since 1996 under the Emergency Medical Treatment and Active Labor Act (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.

mortality (point estimate +0.5%, 95% CI [-3.3%, +4.3%]) .⁴ 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. This study focuses on hospital care for persons with underlying vulnerable health, and finds that having Medicare insurance 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.

Doyle (2005) studies a subpopulation with strong need for emergency medical care (victims of auto accidents 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 increased in-hospital mortality by 39%, relative to the average rate following an auto accident (1.5 more deaths per 100, relative to the mean of 3.8 deaths per 100) (point estimate 0.015, 95% CI [0.003, 0.027], which he attributes to difference in treatment intensity, rather than pre-accident differences in health, so this paper also speaks to a specific channel involving in-hospital treatment intensity, for emergency care for severe traumatic injury.⁵

Levy and Meltzer (2004, 2008) review the literature and conclude that, consistent with the Finkelstein and McKnight (2008) and Card, Dobkin and Maestas (2004) studies, 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 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 from receiving health insurance immediately, versus 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

⁴ The overall mortality results are included in the 2004 NBER working paper but not later published papers (Card, Dobkin and Maestas, 2008, 2009).

⁵ 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, finding that 6-month mortality fell by 71% as a result of gaining insurance.

and thus confidence bounds are wide. They do find that those receiving insurance have higher self-reported health. The second recent experiment 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 no statistically significant improvement in adult mortality or measures of physical health after 2 years. They do find increased health care use, increased diabetes detection and care, 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, but this difference shrinks rapidly 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 of [-26%, +13%]. Thus, both experiments find statistically insignificant effects for relatively vulnerable populations (the disabled for Weathers and Stegman, 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 (Sommers Baicker and Epstein {SBE} 2012; Sommers Long and Baicker (SLB), 2014; Powell (2018); McClellan 2017 find large effects of health insurance on mortality rates. SBE 2012 consider 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; SLB 2014 and Powell 2018 considers the Massachusetts insurance expansion in 2006. McClellan (2017) considers the ACA mandate for young adults to be covered under their parents' employment-based insurance policies. Dunn and Shapiro (forthcoming) considers the effect of Medicare part D prescription drug coverage for elderly adults.

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. One must keep in perspective the chain of causation between the policy change and health or mortality; to assess both the plausible magnitude of any treatment effect and the challenges in measuring that effect using available datasets. 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 information on pre-ACA insurance status (which

one could use to exclude the always insured from the sample, thus increasing the first stage),⁶ or income (which determines eligibility for Medicaid or subsidized private insurance).

Several concepts inform our analysis and interpretation of our results. One is the existence of prior policies that provide vulnerable populations with health insurance, or with health care regardless of health insurance status. These include health insurance or health care for the elderly and disabled through Medicare or Medicaid; pregnant women through Medicaid, many low-to middle income children through the Children’s Health Insurance Program, persons needing emergency care through the 1996 Emergency Medical Treatment and Active Labor Act (EMTALA), persons with specific high-cost health conditions (AIDS through the 1990 Ryan White Act; end-stage renal disease under Medicare since 1972), 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’ practical or political ability to target groups likely to be uninsured and promote a high takeup rate, the less likely studies are to find detectable effects on health. 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 “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 effect for never-takers or always takers (the already insured).

Third, there could be substantial treatment heterogeneity even within the compliers, with health insurance improving health for some, but neutral or even detrimental for others, 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.

⁶ 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.

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).

These concerns, taken together, highlight the complex relationship between health insurance and health outcomes, and anticipates 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 contain records on approximately 2.6 million deaths a year.⁷ 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.⁸ Other data in the mortality files include (1) race, ethnicity, and gender; (2) year of death; (3) age at death (which we collapse to 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 2009-2015 for our main DD analysis (thus, a 5-year pre-treatment period), 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).⁹

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) at or soon after January 1, 2014, and in three more states around or soon after January 1, 2015. "Standard" expansion included coverage for all non-elderly adults with family income less than 138% of the federal

⁷ The public use version of this data is at <http://wonder.cdc.gov/mortSQL.html>, but that version suppresses death counts in county-years with 10 or fewer deaths in any query.

⁸ 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.

⁹ 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.

poverty level (FPL). Of these 30 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 20 “Full Expansion States”; the control group is the 21 “Non-Expansion States” – several of whom expanded Medicaid toward the end of or after our sample period. 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, and the change in percent uninsured in each state from 2013-2015 for persons aged 50-64 years; Appendix Table A-1 provide additional details on each state’s expansion status.

The second major way the ACA expanded coverage was nationwide establishment of “marketplaces” with private insurance subsidies for those with income between 138% and 400% of FPL in expansion states, and from 100% in Non-Expansion states and WI, which expanded Medicaid only to 100% of FPL). Our study design exploits mainly variation in 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 persons aged 50-64 averaged 6.1% in Full Expansion States, versus 5.2% in No-Expansion States; the difference between the two groups is 0.9%.¹⁰

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. Because the “first stage” of the encouragement design is only around 1% of the population, we consider particular subgroups who were more likely to be affected by Medicaid expansion, for whom we can also measure mortality. Even then, we face first stages of 5% or less.

¹⁰ Here, we use uninsurance rates for persons age 50-64 as the closest available match in the Small Area Health Insurance Estimates (SAHIE) data on uninsurance rates to our principal treatment group, which is aged 55-64. The drop in uninsurance rates was somewhat larger for the entire adult population. See Appendix. If one weights states equally, rather than by population, the drops in uninsurance rates are 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, it answers a different question, which is ‘how is the average US state affected?’

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 found a 40% increase in emergency department (ED) visits among the newly Medicaid eligible (Taubman et al., 2014), and Gosh 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.

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, which also lets us assess whether pre-treatment trends are parallel; and (iii) a “distributed lag” model, which allows the treatment effect to cumulate over the post-treatment period. Treatment is recorded in event time, relative to the year in which each expansion state expanded Medicaid. For states which 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; and pre-expansion otherwise (see Table 1 for details). All models use county-level data, county and year fixed effects (FE), county population weights, and standard errors clustered at the state level.¹¹ The simple DD model is:

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

Here i indexes individuals; j indexes county, s indexes state, t indexes time in years, the dependent variable, Y_{jt} is $\ln((\text{deaths})/100,000 \text{ persons})+1$; we add 1 to the mortality rate to avoid dropping county-years with zero deaths. We limit the sample to full-expansion and non-expansion states to form a stronger comparison. The predictor variable of interest is $Post = 1$ for full-expansion states

¹¹ 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 Appendix for details).

in post-expansion years (2014 and 2015 for the 17 states that fully expand 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-federal physicians/1,000 persons.¹² We convert all amounts to 2010 dollars.¹³ 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 (τ_t and ϑ_{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 2010-2013. As expected, there are differences in a number of covariates. Expansion states are **[*one sentence summary to be added]**. To address covariate imbalance, we also implemented an inverse propensity score weighting approach in which we compute ATT weights, and use ATT*population weights;¹⁴ we also implemented an approach in which we roll up our county-level results to the state level, but weight states equally instead of by population. Results from both approaches, 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

¹² 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 (we interpolate from adjacent years for 2009 because data is missing) unemployment rate, median household income, percent in managed care (we interpolate from adjacent years for 2006-2007, because data is missing), 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/dhds/maps/atlas/index.htm>.

¹³ Source: www.bls.gov/cpi/. We use the annual average consumer price index for all urban consumers.

¹⁴ 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-Expansion or Non-Expansion state, using all variables in Table A1 to generate the fitted propensity p . ATT weights are calculated as $(p/(1-p))$.

amenable mortality seeks to capture deaths from conditions that are potentially preventable with timely care; examples include heart disease, stroke, cancer, diabetes, and infections.¹⁵

To study the time pattern of any apparent treatment effect, and to assess whether pre-treatment trends differ between Full Expansion 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).

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

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, 0 otherwise. For states that expanded Medicaid on January 1, 2014, $D_{st}^1 = 1$ for 2014 and $D_{st}^2 = 1$ 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 $\beta_{-3} \equiv 0$.

In the appendix, we also report results from a “distributed lag” model, which allows the treatment effect to evolve during the post reform period:

$$Y_{jt} = \alpha + \sum_{k=1}^2 (\beta_k * D_{jt}^{k-lag}) + \partial X_{jt} + \tau_t + \vartheta_j + \varepsilon_{jt} \quad [3]$$

Here the first treatment lag D_{jt}^{1-lag} equals 1 in for Full-Expansion States beginning in the first expansion year; while D_{jt}^{2-lag} turns on in the second expansion year. Thus, the coefficient on D_{jt}^{1-lag} estimates the impact of expansion in the first expansion year; while the coefficient on D_{jt}^{2-lag} estimated the *additional* impact in the second expansion year after reform. One can then combine the lagged effects to obtain an overall treatment effect ($\sum_{k=0}^2 \beta^k$), and accompanying t -statistic (using the `lincom` command in Stata). The principal difference between the leads-and-lags and distributed lag models is that the leads-and-lags model provides a coefficient and standard error for each year by itself, relative to a base year. In contrast, the distributed lag model provides estimates for annual incremental changes, starting from a pre-reform average; we then compute a “sum of coefficients” for the post-reform period.

¹⁵ 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.

We find evidence from the event study model (described below) 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 age 65-74, who are eligible for Medicare and should not be affected by Medicaid expansion, and limit the sample to persons aged 55-74, thus comparing mortality trends for persons age 55-64 to those for persons age 65-74. 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 which 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 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:

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

We similarly compare counties with high poverty rates in 2013, containing 20% of the sample population, to counties with low poverty rates, 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 the response to health insurance, for which we also have mortality data: education, gender, and race/ethnicity; for example, we expect that lower educated subgroups will have larger first stages and higher mortality rates, and thus be more likely to produce detectable improvements in mortality.

B. Power Analysis

We conduct a power analysis by artificially introducing treatment effects of different sizes into the data in pre-treatment period, and then assessing how often our DD and triple difference regression models can detect these effects at the 90%, 95%, 99%, and 99.9% confidence levels (two-tailed test). 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. We use a simulated (instead of a closed-form) power analyses where we average results over a thousand iterations, as a closed form requires parameterizing the error term for both variance and covariance terms (Burlig et al., 2017) and is especially hard to construct with panel 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. A simulated power analysis avoids these challenges, and allows us to use the same experimental 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, instead of 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 20 states are considered to be pseudo-treated and 21 to be 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 an average treatment effect of a reduction in amenable mortality (from 0% to 6%, in 0.25% increments) on treated persons, all persons age 55-64, living in a pseudo-expansion 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 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, any particular pseudo-treated county will not have its mortality rate decrease by exactly 0.5%, but the average pseudo-treated county will have the imposed treatment effect.

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), which 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 *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 (a believable coefficient).

VII. Full-Sample Results

We present full-sample results in this section, principally for adults age 55-64, and some limited results for adults age 45-64. See Appendix for similar results for all non-elderly adults. In Part VIII, we conduct alternative analyses to assess the effects of ACA-induced insurance variation on mortality, focusing on vulnerable subgroups or particular causes of death, which might be more conducive to producing statistically detectable effects.

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. 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 (and in our case, we show they are not), or by our third difference (they partly are), any DD analysis is suspect. More specifically, over 2010-

2015, 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. 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 over 2010-2013. There is little additional divergence during 2014-2015. The simple DD coefficient would be 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.¹⁶

[FIGURE 1 about here]

B. Covariate Balance

Appendix Table A1 provides a covariate balance table showing means, and the normalized difference in means, between Full-Expansion 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 (our principal outcome; see Figure 1) and uninsurance rates. In light of these differences, we also reran the analyses reported below, which use population weights, with $ATT * \text{population weights}$ instead; results were similar to those we present; see Appendix. We use population weighted results as our main specification as they are more transparent.

C. Leads-and-Lags Regression 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

¹⁶ 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.”

change in relative mortality in the first two expansion years. In the Appendix we provide leads-and-lags graphs for non-amenable and total mortality, which are similar to Figure 2.

[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 age 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) will help, but 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 65-74 to provide background for our triple difference regression estimates. There is again evidence of non-parallel trends, with mortality dropping in Full Expansion versus No 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 Expansion versus No Expansion states and for age 55-64 versus 65-74. Non-parallel trends are muted, but standard errors are larger than in Panel A, and 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 a dip 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 yet too small to be statistically convincing, given the year-to-year variation we observe in the pre-treatment period.

We considered an alternate DD specification which compares persons age 55-64 to those age 65-74 in the same state, but concluded that inference would be unreliable due to strongly non-parallel pre-treatment trends (rising relative mortality for those age 55-64). See Appendix. The triple difference specification appears to be the best available in limiting the extent of non-parallel pre-treatment trends, but remains suspect because it depends on non-parallel trends in the double differences tending to offset each other out in the pre-treatment period, with no basis for confidence that they would continue to do so in the treatment period.

D. Synthetic Control Results

We also 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),¹⁷ to construct a synthetic match using the Non-Expansion States as donor states. We report results in Figure 3.

[FIGURE 3 around here]

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. In any synthetic control analysis, that gap will tend to expand during the treatment period, once it is no longer minimized by construction; the question is whether the gap becomes statistically significant. The 95% CIs show that it does not. Moreover, visually, much of that gap arises in 2013. Put differently, the synthetic control 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. In the end, we are not persuaded that, for our data, the synthetic control approach is an improvement over the triple-difference design.¹⁸

¹⁷ We used code for this approach from Soni (2016).

¹⁸ 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, which do not otherwise seem good matches for the Full-Expansion states (Alaska, Maine, Wyoming). The Appendix includes a table showing the weights on each donor state.

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, and lets one conduct DD analyses on the resulting treated and control units, including generating analytical standard errors (which the original method does not provide). We performed versions of this procedure which weight all counties equally and versions where we weight by population by running the procedure on bootstrapped datasets with draws weighted by population. The results (see Appendix), were unsatisfactory. With gsynth, as with the simpler synthetic control method presented above, there is a relative drop in amenable mortality in Full-Expansion States in 2013, if we use county level data. We believe that this poor pre-treatment fit is due to small counties, which have highly varying death rates and are very hard to fit even if the donor pool is quite large. In addition to the poor pre-treatment fit, the implied post-period treatment effect with county population weights was around .05, and [*0.xx] without these weights. These are far too large to be true effects.¹⁹

E. DD and Triple-Difference Regression Results

We next turn to regression analysis. Table 2 shows results from DD regressions, with county and year FE and county population weights, following equation [1], separately for our principal treatment group (ages 55-64) and the placebo group (ages 65-74), and triple difference results. While both DD and triple difference specifications are suspect because of parallel trends problems, this concern is less severe for the triple difference, thus we focus on those results and show the DD results principally for comparison. 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 age 55-64 population, because we expect the effects of health insurance to be higher for this group than for younger persons, and because doing so lets us pursue the triple difference approach in a more convincing manner. 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

¹⁹ We did further work using gsynth; see Appendix for details. We implemented gsynth both with and without covariates. Without covariates, coefficient were similar to Figure 3B in the pre-treatment period, but the drop in relative mortality in 2014 was much smaller, and not statistically significant. We also used county-level data; results were again implausible, and were driven by large year-to-year swings in mortality in the smallest counties.

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 age 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 (age 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.²⁰

[TABLE 2 around here]

VIII. Is There Evidence of Heterogeneous Effects?

We next investigate whether the discouraging conclusion for the general adult population – no evidence of a statistically significant effect, and far too little power to detect an effect of plausible magnitude – would change if we focused on subgroups that are more likely to be affected by the ACA health insurance expansion. 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, specific cause of death, and county poverty and baseline uninsurance rates.

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

²⁰ 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 the logic behind using mortality among persons age 65-64 as a third difference, yet we need the third difference to address the non-parallel pre-treatment trends observed in Figure 2A.

first-stage, but it is central to assessing what coefficient magnitudes are reasonable. For education, population data is only 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 the 138% of FPL threshold for full ACA expansion), at the county*year level, but cannot directly study these subsamples because we lack information on income and insurance in the mortality data.

A. Variation Between Demographic Groups

We begin in Figure 4 with leads and lags graphs of the triple differences in amenable mortality for samples subdivided on gender and on race/ethnicity (white non-Hispanic, black non-Hispanic, and Hispanic). 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 3). Thus, the point estimates in Figure 4 (around -0.05) are not possible as true effects of Medicaid expansion. There is milder evidence of a drop in mortality in event year +1 (second expansion year) for Hispanics, but here too there is a drop; the point estimate for 2015 (around -0.10) is too large to be possible, let alone plausible.

[Figure 4 around here]

We turn next to DD and triple difference regression results for amenable mortality for these subsamples, starting with demographic subsamples in Table 3. The “all” row in Table 3 are the same as in Table 2. The first column of Table 3 shows the first stage change in uninsurance rates for Full-Expansion vs. No-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 2.4% (not significant). The hope that we might attain greater power for, say Blacks or Hispanics, who were more likely to be affected by Medicaid expansion, fails. The loss in power due to reduced sample size outweighs the gains due to modestly higher first stages.

In Table 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 and

for Hispanics. However, as noted above, both of these estimates are suspect due to non-parallel pre-treatment trends and implausibly large point estimates.

[TABLE 3 around here]

B. Variation Based on Education

In Figure 5, 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 age 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 in Figure 5. There is no evidence of a post-expansion decline in mortality for any subgroup, including the less-educated groups.

[Figure 5 around here]

In Table 4, we show regression results by education level. The first row shows full sample results. These differ from Table 2 due to the broader age range that we are forced to use by data limitations. Note that in preferred triple difference specification, the point estimate 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 around 4% for persons without a high school degree, but drops for 1.6% for high school graduates with no college, and to only 0.5% for persons with some college. However, the non-high-school graduates are only 12% of the age 45-64 population, so the power gained from a stronger first-stage is largely offset by loss of sample size.

The first row shows full sample results. The second through fifth row shows effects for the four education groups, starting with the lowest group, those with only elementary school completion, while the other rows shows 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)

[Table 4 around here]

C. Variation by Primary Cause of Death

In Table 5, we present effects specific to the top 4 causes of death: cancer, diabetes, cardiovascular causes and respiratory illnesses plus HIV. The Appendix includes 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 are a group that could plausibly benefit from Medicaid expansion given the negative correlation between income and diabetes prevalence, and evidence that gaining Medicaid insurance predicted increased diagnosis of diabetes in the Oregon Medicaid Experiment (Baicker et al., 2013). And HIV is a specific condition where health insurance has predicted lower mortality in previous studies (Goldman et al 2001). However, for all of these causes of death, both DD and triple difference coefficients are insignificant.

[Table 5 around here]

D. Variation by Pre-ACA Uninsurance and Poverty Rates

We turn next to an effort to exploit pre-ACA 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 6. Neither graph shows evidence of a treatment effect. Both graphs show signs of a pre-treatment trend toward lower mortality in high-uninsurance counties (over 2009-2013) and high-poverty counties (over 2010-2013), which does *not* continue in the post-expansion period and indeed reverses for the high-uninsurance counties.

[Figure 6 around here]

We present regression estimates in Table 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 rises to 2.4% for the by-uninsurance subsample but to only 1.6% for the by-income subsample, but the sample size is, by construction, only 40% of the full sample. There is no evidence of significant effects of Medicaid expansion on mortality. For the full sample, the coefficients for both subsamples are insignificant, and the coefficient for high-vs-low uninsurance rates is positive (opposite from predicted). For demographic subsamples, six of the 14 coefficients are positive; only two are significant, one of those is positive, and the only significant negative coefficient is barely significant and yet of implausible magnitude.²¹

[Table 6 around here]

E. Heterogeneous Effects: Summary

In sum, our search for evidence of a significant effect of Medicaid expansion on mortality for particular subgroups comes up empty. Most regression coefficients are insignificant. When significance is found (for Non-Hispanic Blacks and for Hispanics, in Table 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. 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.

²¹ In the Appendix, 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.

IX. 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 100,000 per year (Appendix Table App-1), 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 ($0.5 \times 0.25 \times 600$) to 525. In expectation, a DD regression should show a 25% reduction in mortality rate.²² 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 24,²³ 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.²⁴ 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 ($0.05 \times 0.1 \times 600$) to 597. The standard deviation in the number of expected deaths remains the same, so the expected

²² The expected coefficient in a regression, such as those we run, with $\ln(\text{mortality rate} + 1)$ as the dependent variable should be around -0.22

²³ This uses the standard formula for the variance of a binomial distribution with probability $\text{Var} = n \times p \times (1-p)$. For $n = 100,000$ and $p = .006$, $\text{Var} = 596$ and $\sigma = \text{Var}^{0.5} = 24.42$.

²⁴ Variances due to independent sources add so $\text{Var}_{\text{tot}} = 596 + 144 = 740$, and $\sigma_{\text{tot}} = \text{Var}_{\text{tot}}^{0.5} = 27.56$.

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 $594/20 \approx 30$. 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 ($\sigma = 13.2$; $t = 0.23$).

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 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 ($\sigma = 6.64$, $t = 0.45$). 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 ($\sigma = 5.62$; $t = 0.53$). 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 ACA Medicaid expansion

studies.²⁵ From SAHIE data, the first stage for low-income, Medicaid-eligible adults (income < 138% of FPL) is around 5%. We also saw above that the first stage for low-educated adults is around 4%.²⁶ 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.²⁷ ACA-derived insurance gains were somewhat smaller 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.²⁸

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 1.2% (Table 3). 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 people gaining health insurance in Full Expansion States ($0.0012 * 14.1M$). 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

²⁵ 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.

²⁶ Kaestner et al. (2015) estimate a similar 3% first-stage for low-educated adults, age 19-64.

²⁷ 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 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.

²⁸ Appendix Figure [A-X], reproduced from the American Community Survey (ACS) shows the ACA-related change in uninsurance rates by age.

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.²⁹

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 personal characteristics (income, family status, pre-ACA insurance), 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 7 illustrates the results from our power simulation, using the amenable mortality rate for all persons age 55-64 year 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 around 2.2% for the DD simulation, and 2.3% 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. An overall mortality reduction of 2.3%, given the roughly 1.2% first stage, implies that Medicaid expansion would have to reduce the average amenable mortality rate of all newly insured persons by $(.023)/(.012) = 190\%$. If we apply a stricter significance standard, to account for specification error, specification searches, and file-drawer bias, the minimum detectable effect will

²⁹ 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).

be substantially higher – Figure 7 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 2.3% reduction in mortality needed for 80% power and 95% confidence translates into about $.023 * 14.1M * .006 = 170,000 = 1,950$ annual deaths – almost 20 times the maximum plausible effect.

[Figure 7 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 mortality for the controls of 0.008. 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. We illustrate these problems in Figure 8.

In Figure 8, 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 8 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 9. 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 $\ln(\text{state population in 2012})$. We also superimpose a regression line showing the best linear fit between the point estimates and $\ln(\text{population})$.

It is apparent from Figure 9 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 9 around here]

D. Power for Vulnerable Subgroups

We also conducted power analyses for the subsamples considered in Tables 3 and 4, and report results in the Appendix. Power is generally similar to, or lower than, that shown in Figure 7. 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 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.

1. Power with a DD Design and a Focused Subsample

We assume hypothetical first stages varying from 1% to [***xx**] % and hypothetical second stages varying from 1% to [***xx%**]. 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.

In focusing on a subsample with a higher first stage, we also lose sample size. We assume that with the hypothetical data, (i) researchers can perfectly 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 $170,000/.05 = 3.4\text{M}$ treated persons, and a similar number of controls.

In Figure 10, Panel A, we show power curves only for the 95% significance level. We vary (i) the assumed first stage (we show curves for 5%, 7.55, 10%, 12.5%, and 15% first stages) and (ii) the imposed mortality reduction for the newly insured (from 2.5% 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 [***xx%**] or more in this subsample among, or around [***y00**] annual deaths. Yet the gains are limited. As we concentrate new insurance on a smaller and smaller subsample, we gain precision within this subsample. But we have made no progress in addressing the largest source of noise in the simulations, which is the inter-state variation in underlying population mortality.

2. Power with a Triple Difference Design and a Focused Subsample

One could potentially do better with a triple difference design, where we use mortality trends among other similar persons in the same county as the third difference, which can address the non-parallel inter-state trends. We consider two plausible third differences. The first is, as above,

persons aged 65-74. The second is to imagine that we have can identify a within-county control subsample of other persons 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 In the best, albeit unrealistic scenario, the third-difference subsample would completely absorb the non-parallel trends.

In Figure 10, Panel B, we use this triple difference design and impose that best case. We pick at random both the first-stage subsample and a corresponding third-difference subsample of the same size. Since both are drawn at random from the same county, they have by construction the same expected mortality rate. We then show power curves similar to Panel A. **[*discussion to come]** These numbers imply that an analysis with this rich data and, say, a 10% first stage would be powered enough to detect a second-stage of around 6% in this “10%-first-stage subsample” , or around 600 avoided deaths per year. Better, but still not nearly good enough. Recall that with a 10% second stage, we expect around 100 fewer annual deaths among those who actually gain insurance.

[Figure 10 around here]

We can perhaps do still better, with our hypothetical data. Suppose that we abandon the second difference from mortality of similar persons in Non-Expansion States, and use the within-county control subsample in a DD design, maintaining the best-case assumption that the within-county control group perfectly satisfies the parallel trends assumption. We present results in Figure 10, Panel C. **[*discussion to come]**

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 shock-based research.

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 30% first stage among people who took the trouble to sign up for the Oregon Medicaid lottery, who were randomly offered Medicaid or assigned to control, they found 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 their study was undersized, even for that large point estimate, by a factor of around 16 – they would need a sample of 1.2M people to reliably find a 13% effect – and a sample of 8M people to find a 5% effect.

Second, our analysis of power to detect the effect of health insurance on non-elderly adult mortality has direct implications for prior 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 extremely large 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.³⁰ We used Kaestner’s approach to compute a minimum detectable effect (95% confidence). We ranked the absolute values of the placebo treatment effects generated by randomly assigning treated and control groups, and determined the $p < 5\%$ threshold coefficient size. The minimum detectable effect size is 6.9%. Thus, the minimum detectable mortality decline for the newly insured implied by this exercise is $6.9\%/5.6\% = 123\%$.

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 uninsurance 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? Meanwhile, a back of the envelope calculation using their reported 95% CI suggests a standard error ~2% which implies a minimum detectable effect ~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),

³⁰ We thank Robert Kaestner for providing his Stata code, which we used in our analysis.

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 two years after expansion by type of mortality (health care 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 no convincing evidence of an ACA-induced decline in mortality in Medicaid expansion states. Instead, results are mixed; there are often 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 effects could materialize with more time, 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 year or three 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, and health status (aside from mortality) are essential. Income and prior insurance information would permit a substantially larger first stage. Health data would provide a

more sensitive second stage, and might also permit analysis limited to health-vulnerable subpopulations, provided that these subpopulations still provide reasonable sample sizes. 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. These steps should improve researchers' ability to assess the plausibility of reported results in two senses –they should prevent apparently significant results from arising by chance (the usual meaning of a “false positive”) and make it less likely that researchers will report estimates many times larger than true effects.

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.
- 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).
- 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.
- 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 & Daniela Zapata. 2017. Impacts of the Affordable Care Act on health insurance coverage in Medicaid expansion and non-expansion states. *Journal of Policy Analysis and Management*, Vol. 36, No. 1, 178–210
- 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.
- Currie, Janet and Jonathan Gruber. 1997. The technology of birth: health insurance, medical interventions and infant health. Natl. Bur. Econ. Res. Work. Pap. 5985

- 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. *Journal of Health Economics* vol 53, p.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.
- 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
- 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
- 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.
- 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.
- 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.
- 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.
- 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
- Newhouse, Joseph, Insur. Exp. Group. 1993. *Free for All? Lessons from the RAND Health Insurance Experiment*. Cambridge, MA: Harvard Univ. Press.
- 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.
- 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, 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.
- 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
- Xu, Y., 2017. Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects, *Political Analysis* 25: 57-76.
- 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*, vol 51 no 3 p.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.

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.). For Full Expansion states and No-Expansion 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)			change in % unins. (2013-2015)
		2013	2014	2015	
Full Expansion		13.4	9.4	7.0	-6.4
Pop. weighted		12.9	9.3	6.8	-6.1
Arizona ¹		17.6	13.1	9.8	-7.8
Arkansas ²		16.5	11.3	8.7	-7.8
Colorado ³		13.6	9.2	7.2	-6.4
Illinois		14	9.8	6.6	-7.4
Indiana	Feb 2015	12.9	11	8.9	-4.0
Iowa ⁴		7.7	5.9	4.3	-3.4
Kentucky		14.4	7.5	5.3	-9.1
Maryland		10	7.2	5.1	-4.9
Michigan	Apr 2014	11.4	8.2	5.7	-5.7
Nevada		19.8	14.3	10.9	-8.9
New Hampshire	Aug 2014	11.6	9.7	6.8	-4.8
New Jersey ⁵		13.1	10.8	8	-5.1
New Mexico		19	15	11.3	-7.7
North Dakota		9.6	6.9	5.9	-3.7
Ohio		12.3	8.3	6.4	-5.9
Oregon ⁶		15.3	9.6	6.7	-8.6
Pennsylvania	Jan 2015	9.5	7.7	5.7	-3.8
Rhode Island		11.2	6.2	4.5	-6.7
Washington ⁵		13.1	8.2	6.2	-6.9
West Virginia		14.5	8.9	5.9	-8.6
Substantial Expansion		10.3	7.1	5.2	-5.1
Pop. weighted		15.0	10.3	7.0	-8.0
California ⁵	2010	18.1	12.3	8.2	-9.9
Connecticut ⁵	2010	9.7	6.2	5	-4.7
Hawaii ⁷	1994	7.3	5.5	4.4	-2.9
Minnesota ⁵	2010	7.3	5.1	3.6	-3.7
Wisconsin ⁸	2009	9.1	6.6	5	-4.1
Mild Expansion		7.6	5.7	4.4	-3.2
Pop. weighted		8.6	6.8	5.3	-3.2
Delaware ⁹	1996	9.5	7.2	5.5	-4.0
Dist. of Columbia ⁵	2010	6.7	5.2	3.5	-3.2
Massachusetts ¹⁰	2006	3.5	3	2.6	-0.9
New York ¹¹	2001	10.4	8.2	6.4	-4.0
Vermont ¹²	1996	7.9	5	4.1	-3.8
No Expansion		15.7	13.1	11.3	-4.4
Pop. weighted		17.6	14.6	12.4	-5.2
Alabama		13.4	11.6	10.2	-3.2
Alaska	Sep 2015	19.1	17.2	15.2	-3.9

State	Expansion Date	% uninsured (age 50-64)			change in % unins. (2013-2015)
		2013	2014	2015	
Florida		22	17.9	14.1	-7.9
Georgia		18.2	15.2	13.1	-5.1
Idaho		17.1	12.9	11.3	-5.8
Kansas		11.8	9.7	8.2	-3.6
Louisiana	Jul 2016	17.8	15.6	12.1	-5.7
Maine		12.5	11	8.5	-4.0
Mississippi		18.8	15.4	13.1	-5.7
Missouri		13.3	10.3	9.1	-4.2
Montana	Jan 2016	18	14.4	13.2	-4.8
Nebraska		10.7	8.8	6.9	-3.8
North Carolina		15.8	12.6	10.8	-5.0
Oklahoma		18.1	15.6	14.1	-4.0
South Carolina		17.1	13.7	11.5	-5.6
South Dakota		11.3	9.3	9.3	-2.0
Tennessee		15	12.7	10.8	-4.2
Texas		21	17.4	15.5	-5.5
Utah		13	11.3	9.8	-3.2
Virginia		12.3	10.8	8.8	-3.5
Wyoming		13.3	12.7	10.9	-2.4
National		13.5	10.4	8.3	-5.1
Pop. weighted		14.6	11.2	8.8	-5.8

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 $\ln((\text{mortality}/100,000 \text{ persons})+1)$ over 2009-2015 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**.

	DiD		DiD		Triple diff.	
	55-64 yrs	65-74 yrs	55-64 yrs	65-74 yrs	(5)	(6)
	(1)	(2)	(3)	(4)	(5)	(6)
Healthcare Amenable Mortality						
Full Expansion Dummy	-0.021** (0.010)	-0.021*** (0.007)	-0.010 (0.007)	-0.004 (0.005)	-0.012** (0.006)	-0.003 (0.006)
Full Expansion Dummy x Age 55-64 Dummy					-0.007 (0.010)	-0.007 (0.009)
Non-amenable Mortality						
Full Expansion Dummy	0.007 (0.011)	0.002 (0.010)	0.022** (0.011)	0.021** (0.010)	0.019* (0.010)	0.026*** (0.009)
Full Expansion Dummy x Age 55-64 Dummy					-0.011 (0.013)	-0.013 (0.012)
All Mortality						
Full Expansion Dummy	-0.012 (0.009)	-0.014** (0.006)	-0.002 (0.005)	0.002 (0.004)	-0.004 (0.005)	0.004 (0.005)
Full Expansion Dummy x Age 55-64 Dummy					-0.006 (0.009)	-0.007 (0.008)
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	19,656	19,656	19,656	19,656	39,312	39,312

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

First column shows annual averages over 2009-2015 for number of deaths and population in millions. Of the full sample (28.5M people), 14.1M 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 2015 for indicated demographic subsamples, for persons aged 50-64, from regression of percent uninsurance on Full Expansion dummy, with state and year FE and county 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 $\ln((\text{amenable mortality}/100,000 \text{ persons})+1)$ 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	172,598 (28.5)	604.69	1.174** (0.487)	-0.021*** (0.007)	-0.004 (0.005)	-0.007 (0.009)
Male	104,534 (13.7)	759.98	0.738 (0.498)	-0.017* (0.009)	-0.007 (0.007)	0.003 (0.011)
Female	68,063 (14.8)	460.26	1.078** (0.519)	-0.028*** (0.009)	-0.004 (0.009)	-0.012 (0.012)
White (Not Hispanic)	128,511 (21.7)	592.40	0.938** (0.440)	-0.015** (0.007)	-0.007 (0.006)	-0.007 (0.009)
Black (Not Hispanic)	31,793 (3.4)	931.06	1.424 (0.817)	-0.038** (0.017)	0.021 (0.014)	-0.056*** (0.019)
Other	3,527 (1.3)	269.61	- -	-0.099 (0.065)	-0.010 (0.048)	-0.081 (0.087)
Hispanic	8,848 (2.2)	397.78	2.444 (1.344)	-0.162*** (0.056)	-0.068 (0.049)	-0.072** (0.034)
Not Hispanic	163,750 (26.3)	622.18	- -	-0.021*** (0.007)	-0.003 (0.005)	-0.008 (0.008)
Pop. Weights			Yes	Yes	Yes	Yes
Covariates			Yes	Yes	Yes	Yes

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

First column shows annual averages over 2009-2015 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 2015 for indicated education-levels, for persons aged 45-64, from regression of percent uninsurance on Full Expansion dummy, with county and year FE and county 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 $\ln((\text{amenable mortality}/100,000 \text{ persons})+1)$ 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	251,302 (59.5)	422.29	1.109** (0.546)	-0.012 (0.008)	-0.016*** (0.006)	0.011 (0.010)
Elementary School	14,921 (2.6)	571.81	4.269** (2.146)	0.057 (0.051)	0.008 (0.058)	0.058 (0.046)
High School Incomplete	33,490 (4.4)	761.29	3.792** (1.614)	-0.022 (0.062)	-0.020 (0.063)	-0.008 (0.037)
High School Complete	109,260 (18.1)	604.11	1.617** (0.749)	-0.025 (0.042)	-0.034 (0.039)	0.010 (0.016)
Some College	86,379 (34.4)	251.00	0.493 (0.468)	-0.016 (0.035)	-0.021 (0.030)	0.014 (0.013)
Pop. Weights			Yes	Yes	Yes	Yes
Covariates			Yes	Yes	Yes	Yes

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

First column shows annual averages over 2009-2015 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 $\ln((\text{amenable mortality}/100,000 \text{ persons})+1)$ among persons with indicated primary cause of death, 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**.

By Cause of Death	deaths (pop. In M) (1)	DiD 55-64 yrs (2)	DiD 65-74 yrs (3)	Triple diff. (4)
All Amenable	172,598 (28.5)	-0.021*** (0.007)	-0.004 (0.005)	-0.007 (0.009)
Cancer	86,733 (28.5)	-0.002 (0.007)	0.006 (0.006)	-0.004 (0.009)
Diabetes	14,186 (28.5)	-0.036 (0.025)	0.014 (0.030)	-0.022 (0.024)
Cardiovascular	69,718 (28.5)	-0.019 (0.011)	-0.004 (0.010)	-0.001 (0.012)
Respiratory	16,129 (28.5)	-0.029 (0.020)	-0.020 (0.014)	-0.011 (0.021)
HIV	1,279 (28.5)	-0.038 (0.037)	0.015 (0.038)	-0.050 (0.060)
Pop. Weights		Yes	Yes	Yes
Covariates		Yes	Yes	Yes

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

First column shows annual averages over 2009-2015 for number of deaths and population 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 2009-2015, for full sample and indicated demographic 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**.

	(1)	(2)	(3)	(4)	(5)
Demographic Subsamples	Deaths (pop. in M)	First Stage	Triple diff. Uninsurance	First Stage	Triple diff. Poverty
All Amenable	65,642 (11.8)	2.338*** (0.586)	0.025 (0.024)	1.623** (0.736)	-0.010 (0.019)
Male	40,387 (5.7)	1.511*** (0.586)	-0.018 (0.036)	1.380** (0.660)	-0.044** (0.022)
Female	25,770 (6.1)	2.769*** (0.626)	0.070** (0.030)	1.809** (0.803)	0.030 (0.027)
White (Not Hispanic)	50,852 (9.0)		-0.009 (0.032)		-0.028 (0.020)
Black (Not Hispanic)	11,783 (1.4)		-0.048 (0.063)		0.017 (0.050)
Other	1,504 (0.5)		-0.167 (0.165)		-0.043 (0.152)
Hispanic	3,331 (0.9)		0.689 (0.595)		0.028 (0.112)
Not Hispanic	60,321 (10.3)		0.020 (0.024)		-0.018 (0.019)

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 1999-2014, using county population weights. State groups are defined in Table 1. Vertical line separate pre-expansion from expansion period.

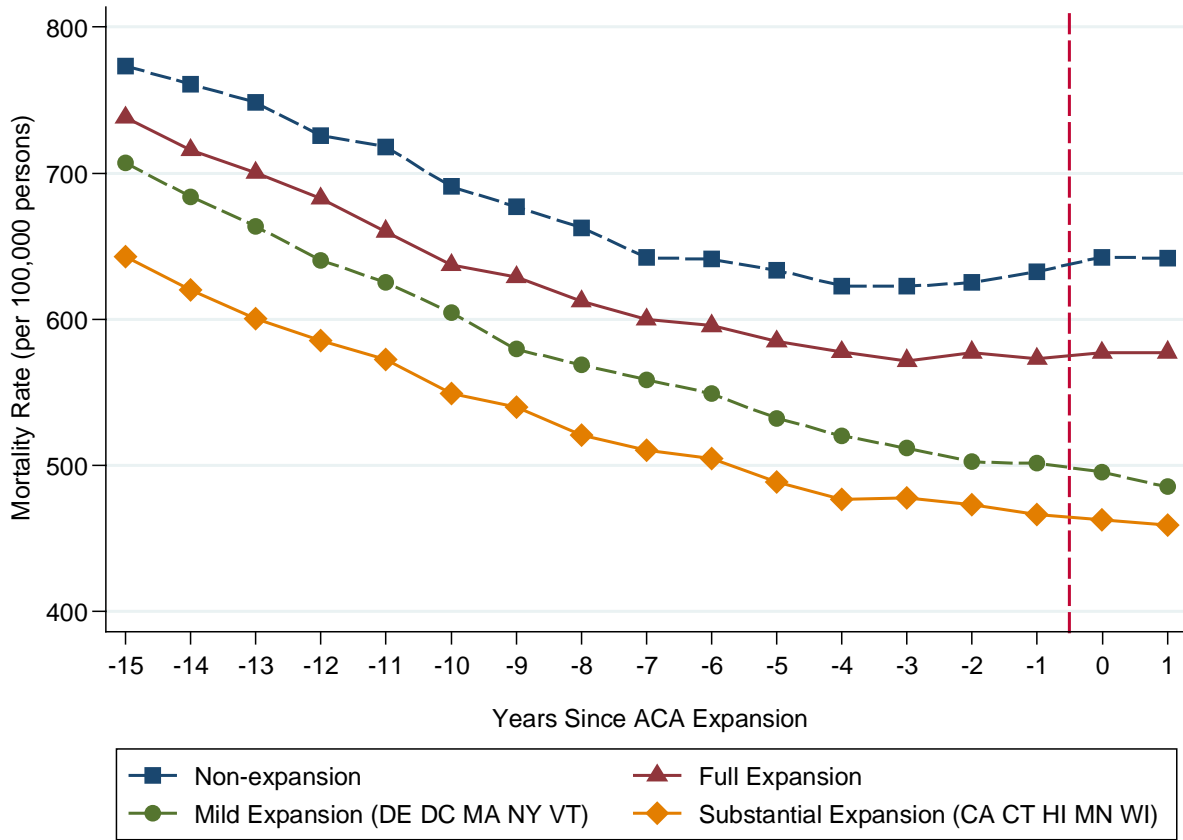
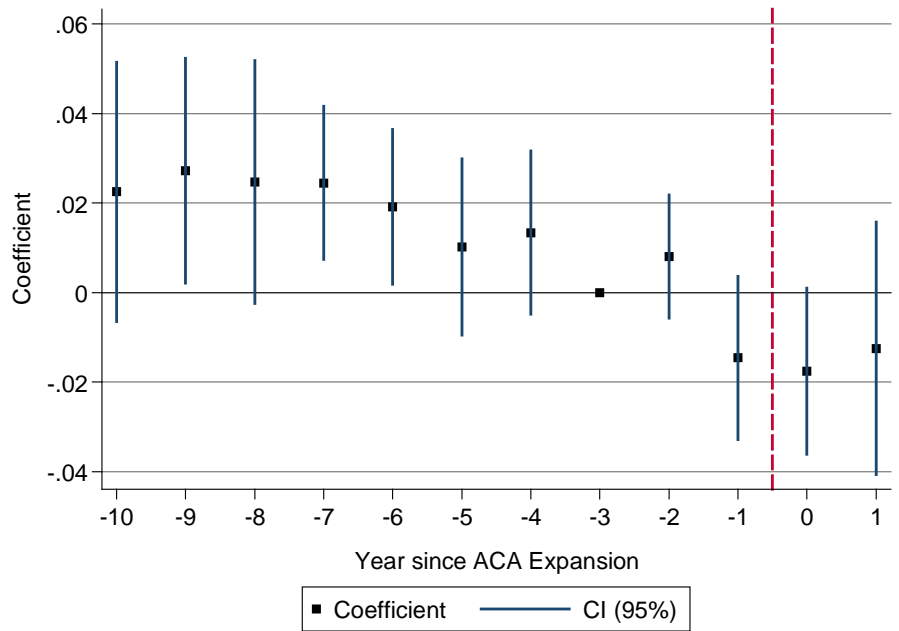


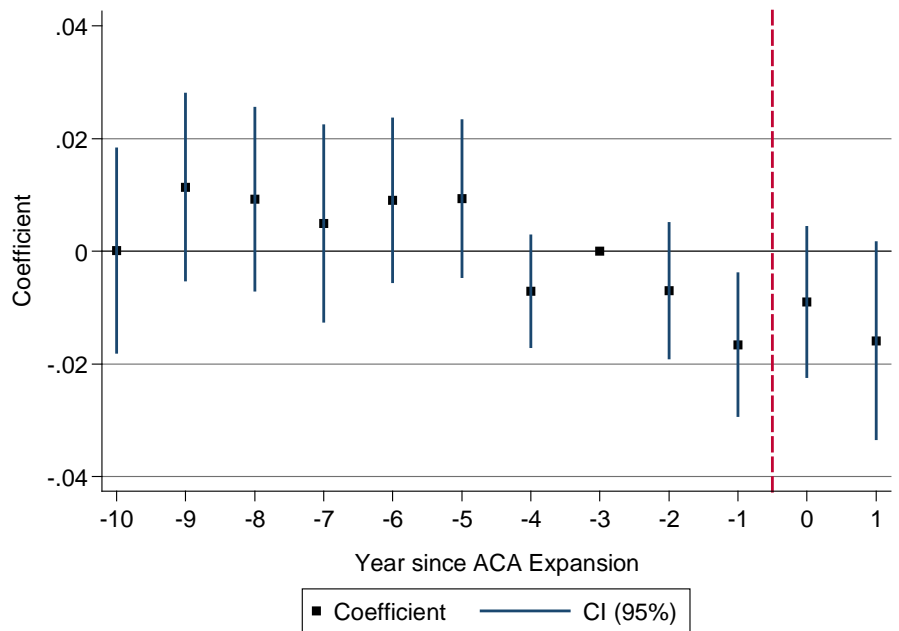
Figure 2. Leads-and-Lags Results for Ages 55-64 and 65-74, Amenable Mortality

Graphs from leads and lags regressions of $\ln((\text{amenable mortality}/100,000 \text{ persons})+1)$ for Full-Expansion States versus control group of Non-Expansion States, over 2004-2015. 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.

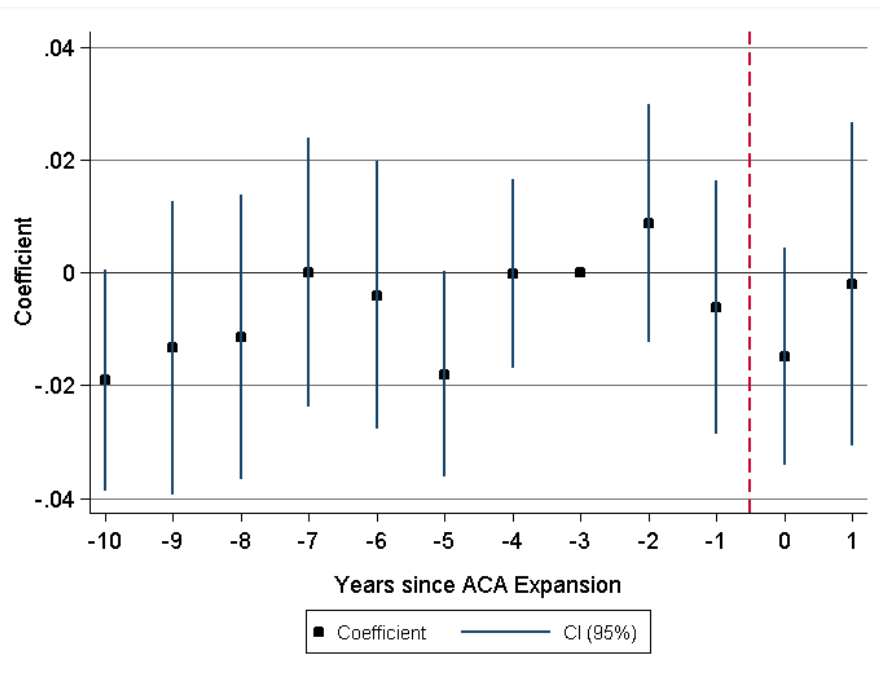


Figure 3. Synthetic Control Results for Near-Elderly Amenable Mortality

Synthetic control results for $\ln((\text{amenable mortality}/100,000 \text{ persons})+1)$ for Full-Expansion States (treated as a single treated unit) versus synthetic control drawn from Non-Expansion States, over 1999-2015. Covariates for constructing donor pool are same as in Figure 2, plus uninsurance rate in 2013. The y-axis shows $\ln((\text{amenable mortality}/100,000 \text{ persons})+1)$ 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.

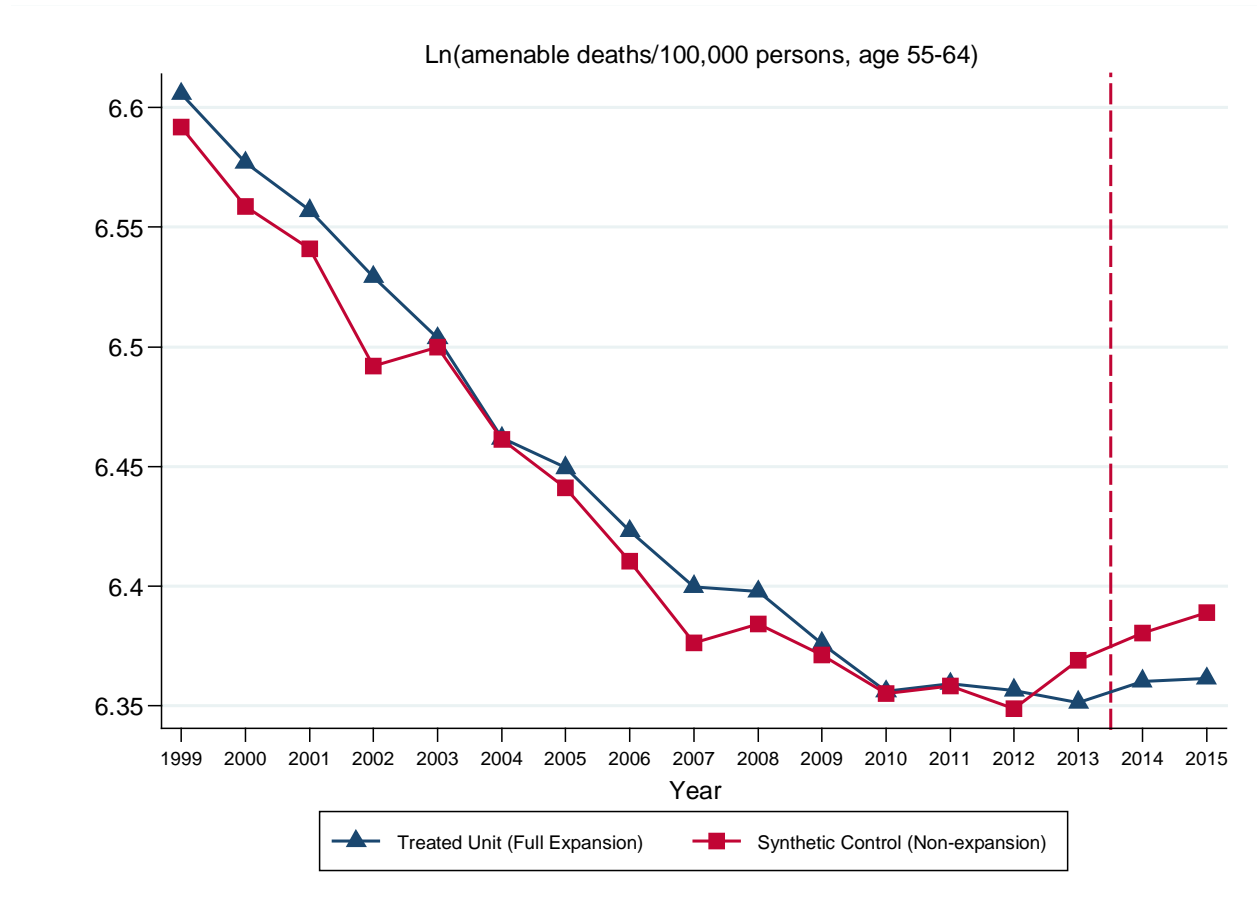


Figure 4. Triple Difference Leads-and-Lags Graphs: Demographic Groups

Graphs from leads and lags regressions of triple differences for indicated subsamples, of $\ln((\text{amenable mortality}/100,000 \text{ persons})+1)$ for persons aged 55-74, in Full-Expansion States versus No-Expansion States, over 2004-2015; 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.

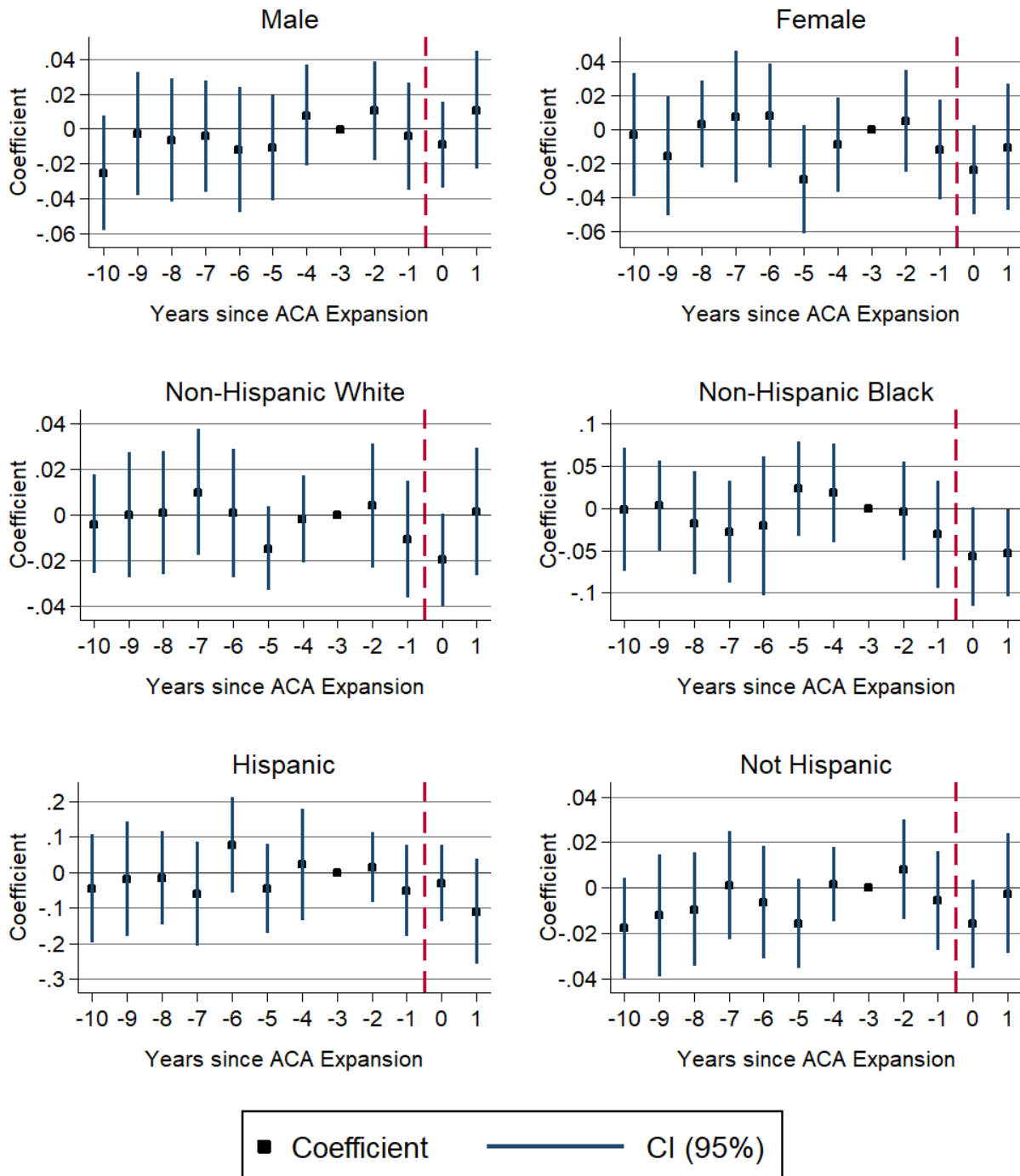


Figure 5. Triple Difference Leads-and-Lags Graphs: By Education Level

Graphs show leads and lags regressions of triple differences for indicated subsamples, of $\ln((\text{amenable mortality}/100,000 \text{ persons})+1)$ for persons aged 45+, in Full-Expansion States versus No-Expansion States, over 2004-2015; 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.

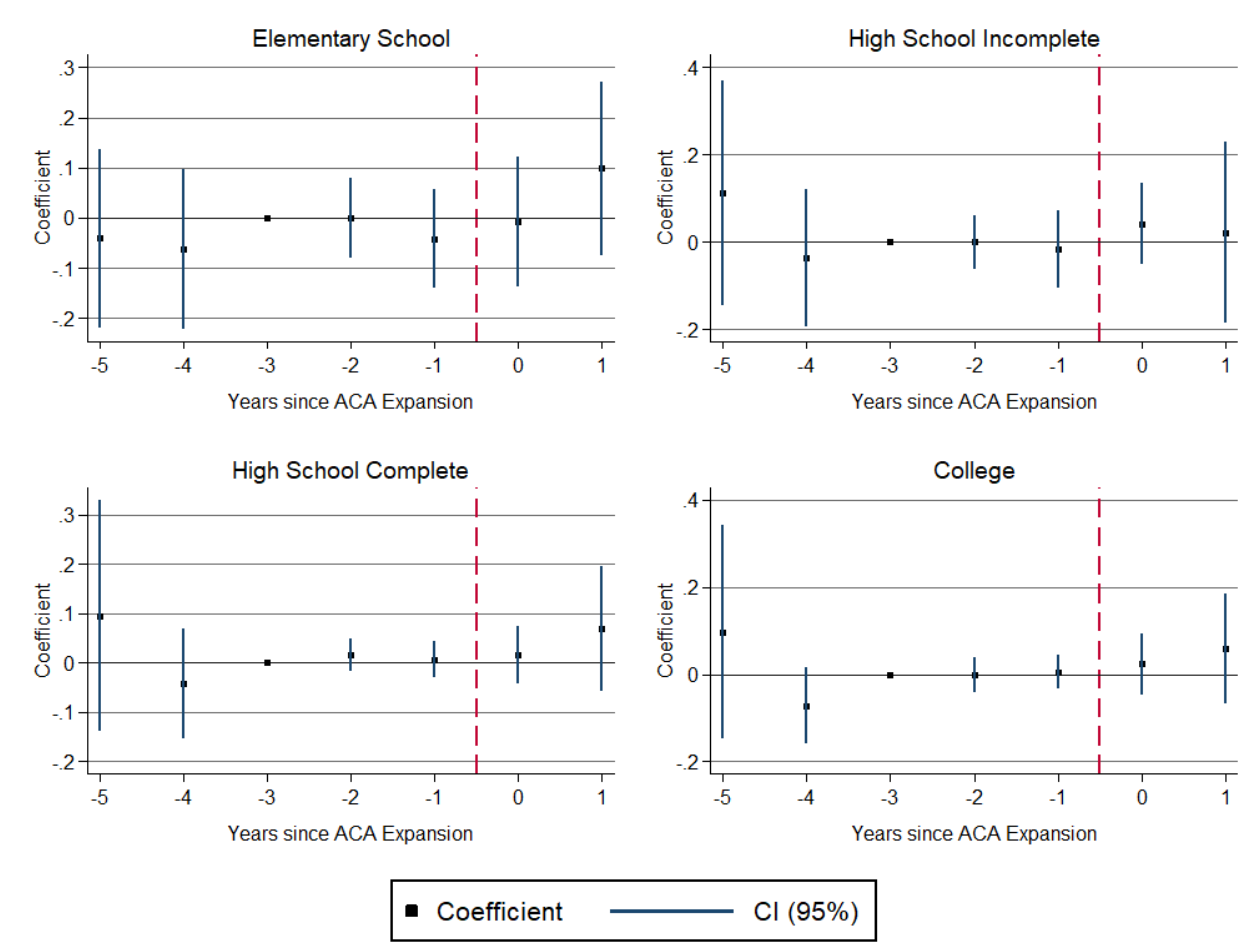
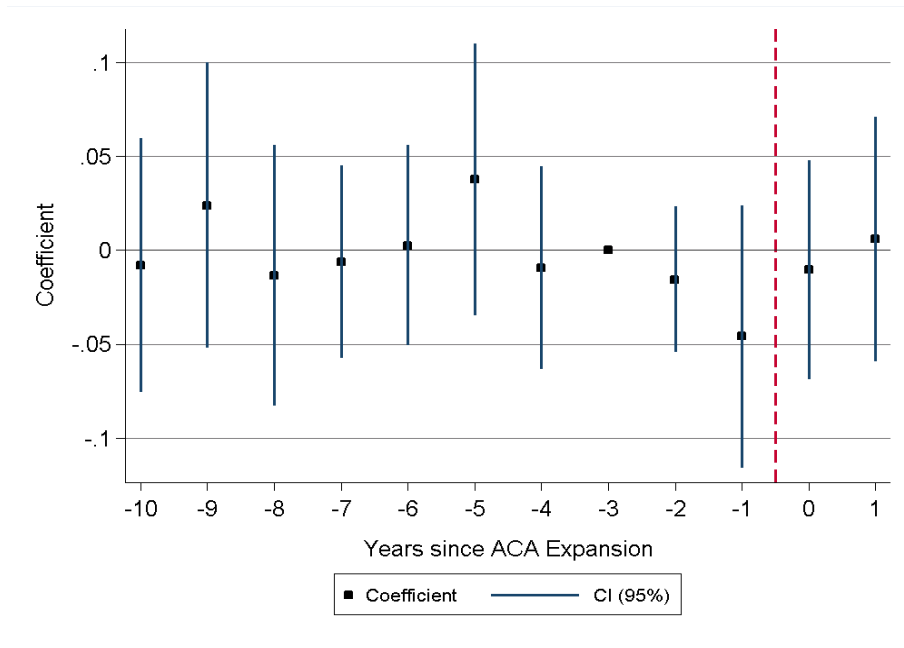


Figure 6: Leads and Lags Graphs for High-vs-Low Uninsurance and Poverty

Graphs show leads and lags regressions of triple differences for high versus low uninsured and high vs. low poverty counties, of $\ln(\text{amenable mortality}/100,000 \text{ persons})+1+$ for persons aged 55-64, in Full-Expansion States versus No-Expansion States, over 2004-2015. High (low) uninsured counties are those with highest (lowest) uninsured 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

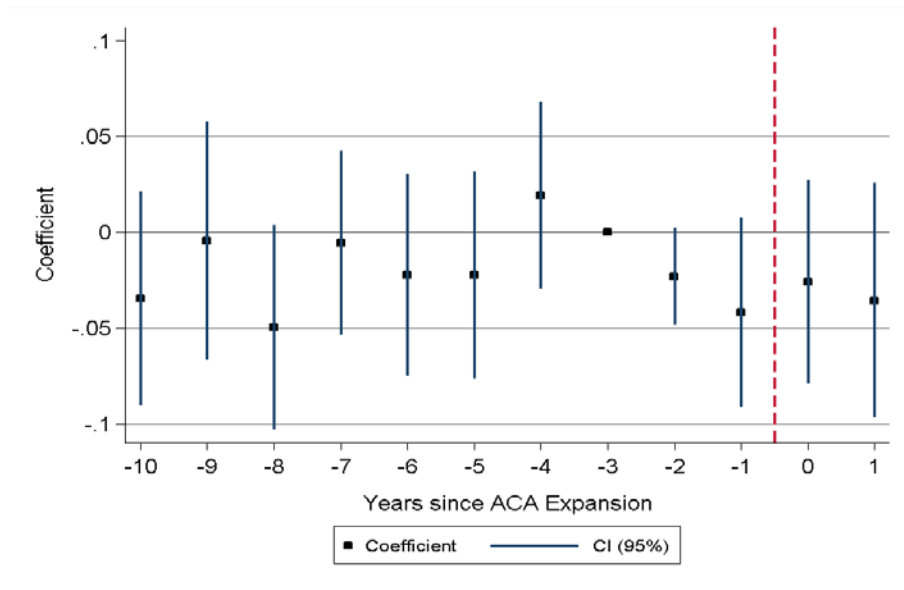
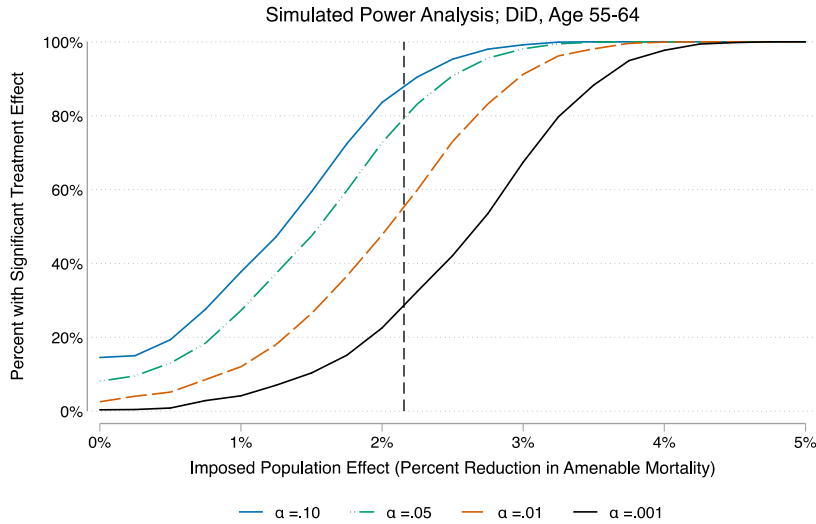


Figure 7: Simulation-Based Power Analysis

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 $\alpha = .10/.05/.01/.001$ correspond to 90%/95%/99%/99.9% confidence levels, respectively.

Panel A. DD Analysis



Panel B. Triple Difference Analysis

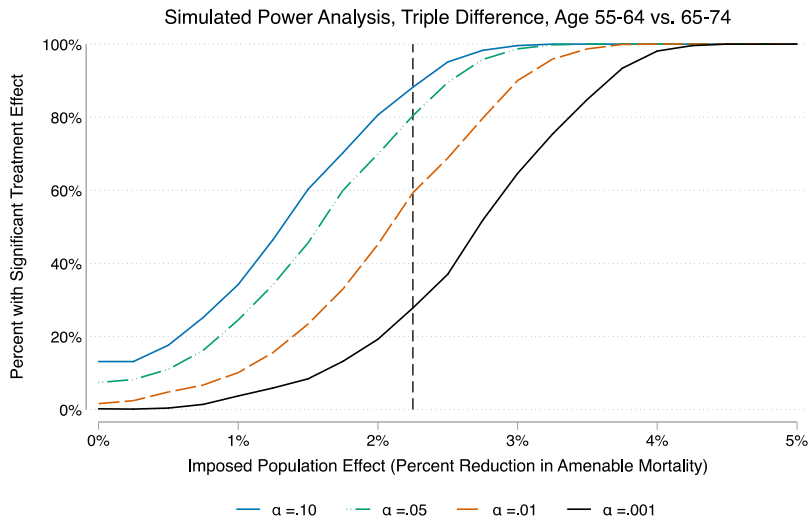
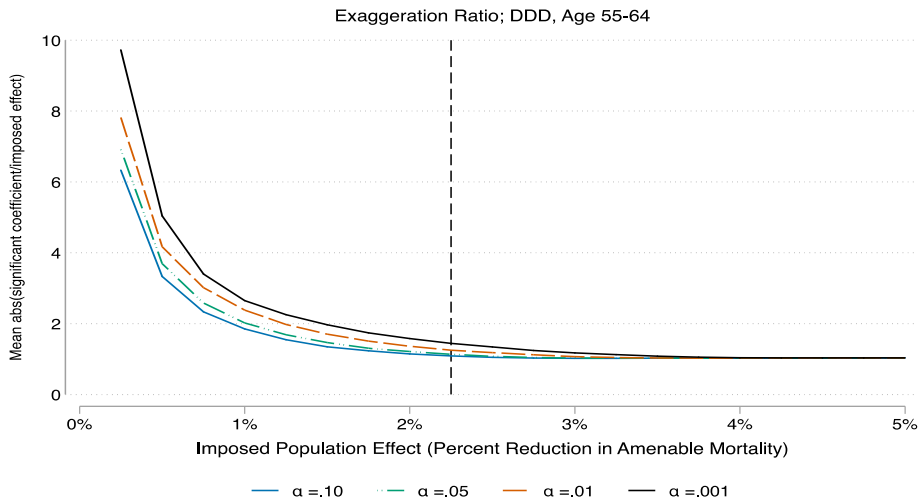


Figure 8. Power Analysis Extensions: Exaggeration Ratio and Likelihood of Wrong Sign

We conduct the same power analyses as in Figure 7 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 $\alpha = .10/.05/.01/.001$ correspond to 90%/95%/99%/99.9% confidence levels, respectively.

Panel A. Exaggeration Ratio



Panel B. Probability that Estimated Effect Has Wrong Sign

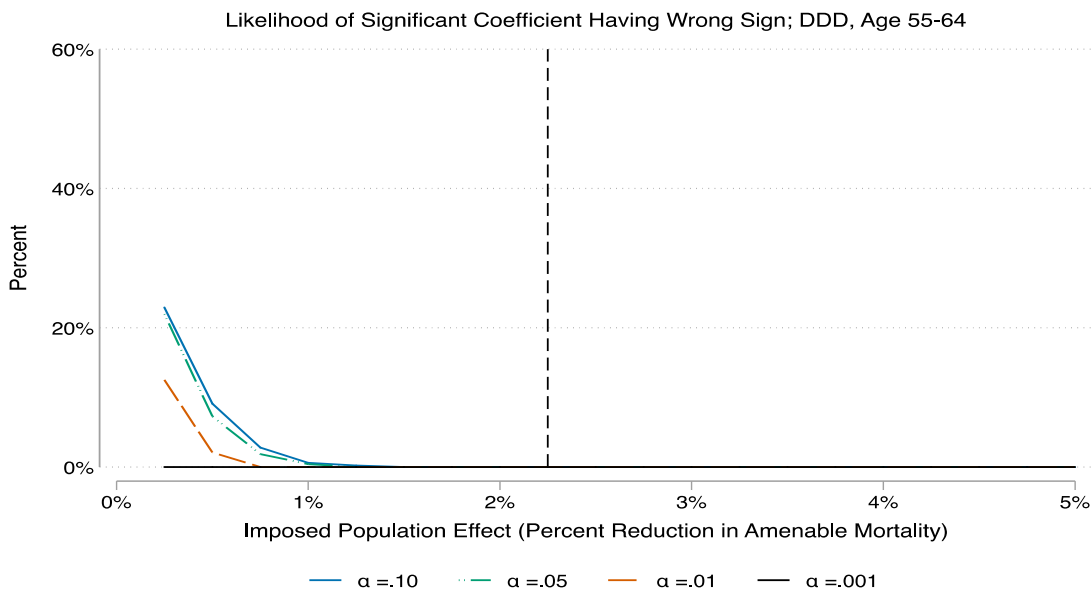


Figure 9. Pseudo-Shocks to Individual States in 2012-2013

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 $\ln(\text{state population in 2012})$ and constant term.

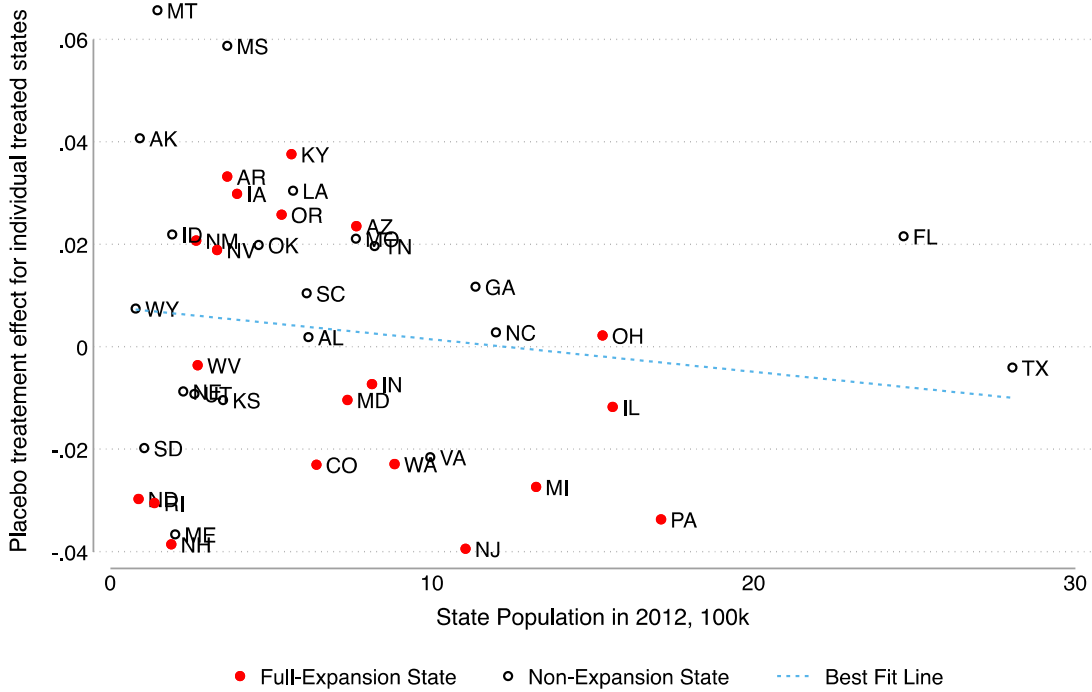


Figure 10. Simulation Based Power Analysis with Known Mortality Status of Decedent
[*to come, significance is at 95% confidence level]

