Damage Caps and Defensive Medicine: Reexamination with Patient Level Data

Ali Moghtaderi
George Washington University, School of Medicine and Health Science

Steven Farmer
George Washington University, School of Medicine and Health Science

Bernard Black
Northwestern University, Pritzker School of Law, Institute for Policy Research, and Kellogg School of Management

(Draft October 2017)

Northwestern University Law School
Law and Economics Research Paper No. 16-xx

Damage Caps and Defensive Medicine: Reexamination with Patient Level Data

Ali Moghtaderi, Steven Farmer, and Bernard Black*

Abstract

Physicians often claim that they practice "defensive medicine," including ordering extra imaging and laboratory tests, due to fear of malpractice liability. Caps on non-economic damages are the principal proposed remedy. Do these caps in fact reduce testing, overall healthcare spending, or both? Do they affect healthcare outcomes? We study the effects of "third-wave" damage caps, adopted in the 2000s, on specific areas that are expected to be sensitive to med mal risk: imaging rates, cardiac interventions, and lab and radiology spending, using patient-level data, with extensive fixed effects and patient level covariates. We find heterogeneous effects. Cardiac stress testing rates *rise*, as does spending on laboratory and radiology tests; these results are robust across a variety of specifications. CT scan rates also rise in most specifications. In contrast, cardiac intervention rates (catheterization, stenting, and bypass surgery) do not rise (and likely fall). We find some evidence that Medicare Part B rises, but variable results for Part A spending. We find no evidence that caps affect mortality.

JEL Classifications: I11, I18, K23, K32

Keywords: medical malpractice, tort reform, defensive medicine, Medicare, healthcare spending

* Moghtaderi is Assistant Research Professor at George Washington University, School of Medicine and Health Sciences, Department of Clinical Research and Leadership, Tel. 202-994-1966, email: moghtaderi@gwu.edu. Farmer is Associate Professor of Medicine and Public Health at George Washington University, School of Medicine and Health Science, Tel. 202-994-7388, email: safarmer@gwu.edu. Black is Nicholas J. Chabraja Professor at Northwestern University, Pritzker School of Law, Institute for Policy Research, and Kellogg School of Management. Tel. 312-503-2784, email: bblack@northwestern.edu. We thank participants in workshops at American Society of Health Economics 2016 Annual Meeting, American Law and Economics Association (2016), Conference on Empirical Legal Studies (2016), Conference on Empirical Legal Studies – Asia (2017); Southern Economics Association (2016), University of Wisconsin at Milwaukee, Economics Department, University of Paris II Law School, Michael Frakes and Kosali Simon (discussants), William Sage, and **Keith Telster** for comments and suggestions. The National Heart, Lung, and Blood Institute provided funding for this project (5 R01 HL113550).

1. Introduction

Physicians often claim that they practice "defensive medicine," notably ordering extra imaging and laboratory tests, due to fear of medical malpractice ("med mal") liability, which drives up healthcare costs. The concept of defensive medicine has no precise definition, but includes conducting tests and procedures with no (or even negative) clinical value, or whose value is too low to justify the associated cost. Imaging and laboratory tests are widely believed to be overused, partly for defensive reasons. An often proposed remedy is caps on non-economic damages.

We study whether damage caps affect imaging rates, cardiac interventions, and lab and radiology spending, using patient-level data, with extensive fixed effects and patient level covariates. Relative to prior research on defensive medicine, much of which studies principally overall spending, we innovate in two principal ways. First, we study specific areas that are likely to be sensitive to med mal risk. Second, we use a very large longitudinal dataset (the 5% Medicare random sample, covering around 2M patients), with zip code fixed effects (FE), plus extensive covariates.

We study "third-wave" damage caps, adopted during 2002-2005. We use a difference-in-differences (DiD) research design. We compare nine "new-cap" states which adopted caps during this period, to a narrow control group of 20 "no-cap" states, with no caps in effect during our sample period, and a broad control group that also includes the 22 "old-cap" states, with caps in effect throughout our sample period. We study rates for the principal cardiac stress tests (stress electrocardiogram (stress ECG), stress echocardiography (stress echo), and single-photon emission computed tomography (SPECT)); other computed tomography (CT) scans, and magnetic resonance imaging (MRI). We also study the principal invasive cardiac procedures: left-heart catheterization (LHC); percutaneous intervention (PCI, often called stenting), cardiac artery bypass grafting (CABG); and any revascularization (PCI or CABG). For spending, we study two categories expected to be sensitive to malpractice risk -- outpatient laboratory (below, "lab") and radiology spending (including stress tests, MRI and CT scans). We also study overall Medicare spending, for comparison to prior studies.

Our principal specification uses zip code and calendar year FE, plus extensive patient-level and county-level covariates, and compares new-cap states to no-cap state. Thus, we ask whether caps affect testing rates, cardiac intervention rates, and lab and imaging spending, in the same location, with controls for patient age, comorbidities, and other time-varying factors that could affect clinical decisions. In alternative models, we use patient*zip code FE (thus controlling for unobserved but time- constant patient health characteristics) or physician*zip code FE (thus controlling for unobserved but time-constant physician FE). As discussed below, using either patient FE or physician FE (we cannot feasibly use both together) has both advantages and costs; the balance between these depend on which outcome we study. The choice of whether to prefer the narrower or broader control group is also a close one. We also assess the sensitivity of our results to a number of other alternative specifications, including using more or fewer covariates, controlling for tort reforms other than damage caps, adding linear state trends,

Note that physicians may respond to malpractice risk in two distinct ways. They may order tests and other procedures with little or no health benefit, which reduce malpractice risk -- sometimes called "assurance behavior." Physicians may also avoid risky patients or risky procedures – sometimes called "avoidance behavior." If risk declines, physicians may engage in *both* less assurance behavior (hence fewer tests and lower spending), *and* less avoidance behavior (hence higher spending, and perhaps more testing as well). Providers may also order tests and perform procedures with limited clinical value for reasons other than liability risk, including economic incentives; patient preferences; desire to be thorough, and local norms. If physicians have multiple reasons to "do more," tort reform could have only a modest impact on clinical decisions and spending. Thus, the effect of caps on imaging rates and other clinical decisions is an empirical question. The balance between the effect of caps on assurance versus their effect on avoidance behavior could vary across physicians, patients, and procedures.

We find heterogeneous results, consistent with the balance between assurance and avoidance behavior varying across patients and procedures. Cardiac stress testing rates rise, and in most specifications MRI and CT rates also rise, with most point estimates around 4-6%. The rise in stress testing is statistically significant across specifications; the rise in MRI and CT scan rates is significant in most specifications. Combined lab and radiology spending rises by around 4-5% (significant in all specifications); radiology and lab spending rise separately (almost always significant for radiology). In contrast, cardiac intervention rates do not rise, and may fall – all point estimates are negative, typically around 4-6%, and are significant or marginally so with patient RE. For broad spending categories, Part B spending, all point estimates for Part B spending are positive and most are significant; we estimate a 2-4% rise in Part B spending. In contrast, the coefficients for Part A spending are never significant, are of variable sign, and are sensitive to specification.[1]

We find no evidence that damage caps affect mortality. This is expected given the modest effects of caps on utilization, and the likelihood that any effects are principally for patients at the margin for being tested or treated, versus not.

A core finding is that damage caps do not have a unidirectional effect on clinical decisions. Instead, procedure rates and spending appear to rise in some areas and fall in others. Determining how caps affect specific clinical decisions requires close examination of particular practice areas. We begin that effort in a companion paper (Farmer et al., 2017), where we focus on clinical decisions whether to conduct an initial ischemic evaluation for possible coronary artery disease (CAD), how to conduct that evaluation, and whether to engage in follow-up testing or treatment. We find that overall ischemic evaluation rates do not change -- the rise in stress test rates that we find here is offset by lower use of LHC as an initial diagnostic test. We also find a sharp drop in progression from an initial stress test to LHC (as a second, more precise test); and in progression from ischemic evaluation to revascularization. These results are consistent with physicians being more willing to tolerate clinical ambiguity and accompanying med mal risk when med mal risk falls. These nuanced results for CAD testing and treatment make sense of the otherwise puzzling finding that stress testing, if anything, *rises* after cap adoption. Clinically plausibly stories can also explain our counter-intuitive evidence for higher CT and MRI rates, and greater overall lab and imaging spending, after cap adoption.

The rest of the paper is organized as follows: Section 2 provides a literature review. Section 3 describes the data and estimation strategy. Section 4 presents results for procedure rates. Section

[1] Medicare uses administered pricing, with prices set largely on a national basis. Thus, when we study Medicare spending, we effectively study whether tort reform affects the *quantity* of medical services provided, not any effect of reform on prices. Medicare rates include a component that is related to the cost of med mal insurance, but this component is small and is revised only every five years. The med mal insurance component was revised most recently in 2000, 2005, and 2011. An example for Texas, which adopted a strict damage cap in 2003. The overall reimbursement to a cardiologist in Austin, Texas in 2011 for a transthoracic echocardiogram (CPT code 93350) was around $210. The med mal insurance component for Texas physicians rose by $0.96 (0.61% of the total payment for this procedure) in 2005 versus 2000, then fell by $0.26 in 2011. Note that any post-cap reductions in physician reimbursements after tort reform would cut against our finding of a rise in Part B spending.

5 presents results for Medicare spending. Section 6 provides results for mortality rates. Section 7 provides robustness checks. Section 8 discusses our findings and some study limitations, nd Section 9 concludes.

2. Literature Review

2.1. Effects of Damage Caps on Malpractice Risk

Many states have adopted a variety of tort reforms, which are intended to reduce med mal liability risk. Damage caps, adopted in 31 states as of the end of third reform wave in the early 2000s are the most important of the commonly adopted reforms. There is evidence that they significantly reduce both claim rates and payout per claim. In the principal datasets, which cover only closed claims, the effects on claim rates and payouts appear gradually during the post-reform period as pre-reform claims are closed (Paik, Black, and Hyman, 2013a, 2013b).[2]

2.2. Effects of Damage Caps on Health Care Spending

A principal policy rationale for cap adoption is the belief that caps will decrease defensive medicine and its associated costs. However, studies of the impact of damage caps on healthcare spending find mixed results. We discuss prior DiD studies here; Paik, Black, and Hyman (2017) provide a recent, more complete review.[3]

The best-known studies are by Kessler and McClellan (1996, 2002). They studied the effect of mid-1980s damage caps on Medicare spending for in-hospital care for heart disease (acute myocardial infarction or ischemic disease), and found that damage caps and other "direct" tort reforms reduced in-hospital spending over the next year by 4-5% without adverse health outcomes. The Congressional Budget Office (CBO, 2006) found a 5.2% drop in Part A (hospital services) spending. However, when the CBO controlled for hospital prices prior to the 1984 implementation of the Medicare prospective payment system for hospitals, the estimated post-cap drop in Part A spending fell to 1.6% and was statistically insignificant. The CBO also found a statistically insignificant 1.7% drop in Part B spending. Sloan and Shadle (2009) studied the effect of second-wave tort reforms on Medicare spending for hospitalized patients, in the year in of hospitalization occurs for heart attack, breast cancer, diabetes, and stroke over 1984-2000, with mixed results across these conditions.

Paik, Black, and Hyman (2017) study third-wave damage caps, using county-level spending data. They found that damage caps have no significant impact on part A spending, but predict 4-5% higher Medicare part B spending. Avraham and Schanzenbach (2015) study heart attack treatment after third-wave reforms and find a post-cap increase in medical management and a corresponding drop in combined PCI and CABG rates, but also substitution away from less-invasive PCI toward more-invasive CABG.

A notable non-DiD study by Baicker, Fisher and Chandra (2007) used med mal payments by physicians reported to the National Practitioner Database and med mal insurance premia from Medical Liability Monitor as a proxy for med mal risk. They found an insignificant overall

[2] Some earlier studies found mixed evidence that damage caps lead to lower post-cap claim rates and payouts. See, e.g., Avraham (2007); Donohue and Ho (2007); Waters (2007), and the literature review in Paik, Black and Hyman (2013b).

[3] A number of studies focus on Cesarean section rates, also with mixed results. See Currie and MacLeod (2008); Frakes (2013); Yang et al. (2009).

association with total Medicare spending, but found that higher premia predicted higher spending on imaging tests, but not other diagnostic tests.

2.3. Trends in Diagnostic Imaging Procedures

Rates for various imaging rates increased through around 2005-2010, depending on the test, but have since leveled off or even declined (e.g., Lucas and DeLorenzo, 2006; Andrus and Welch, 2012, and results for our sample in the Appendix). Physician's fear of malpractice has been cited as a reason for higher utilization of diagnostic testing for emergency department patients presenting with possible acute coronary syndrome (Katz and Williams, 2005; Kanzaria and Hoffman, 2015).

In response to rising imaging rates, several organizations launched initiatives aimed at reducing inappropriate testing. In 2005, the American College of Cardiology Foundation in conjunction with subspecialty societies and organizations developed Appropriate Use Criteria for echocardiography, nuclear cardiology and interventional cardiology (Hendel and Patel, 2013). The criteria are periodically revised as new evidence emerges. In 2009, the National Physicians Alliance launched the Choosing Wisely Campaign (Morden and Colla, 2014). There is some evidence that these campaigns have reduced cardiac imaging rates for low-risk patients (Rosenberg and Agiro, 2015). They could help to explain why, in our data, cardiac stress testing rates peak n the mid-2000s, and then begin to decline.

3. Data and Methodology

3.1. Datasets and Covariates

Our core dataset is the 5% Medicare fee-for-service random sample, covering Medicare part A (hospital services, both inpatient and outpatient) and Part B (physician services), for patients age 65+. Our cap adoption events are from 2002-2005. We report our principal, regression-based results use data for 1999-2011.[4] In univariate calendar time graphs, presented in the Appendix, we extend the sample period through 2013. This is patient level data, on roughly 2 million beneficiaries per year. As beneficiaries leave the sample (principally through death), the sample is "refreshed" with new beneficiaries, about 70% of whom are age 65 when they enter our sample.

Our principle outcome variables are 0-1 count variables for whether a beneficiary received a test or procedure in a given year, or Medicare spending in 1999 dollars.[5] The tests we study are cardiac stress tests (stress ECG, stress echo, and SPECT)); CT scans (other than SPECT), and MRIs. The cardiac interventions we study are LHC (an invasive diagnostic test) and revascularization through PCI or CABG. The Medicare spending categories we study are outpatient radiology and lab spending, and overall Part A, Part B, and "total" (Part A plus Part B) spending. The Appendix contains details on the diagnostic codes we use to define our outcome variables.

We include the following patient-level covariates: dummy variables for male, white, black (the omitted race category is "other"), for each of the 17 elements of the Charlson comorbidity index,

[4] The ending date is based on judgment as to when is "long enough" after the reform events so that any trends after that cannot reliably be attributed to the reforms.

[5] We convert nominal dollar amounts to real equivalents in 1999 dollars using the Consumer Price Index for all Urban Consumers, using year-end values. Source: www.bls.gov/cpi/. We used Healthcare Common Procedure Coding System (HCPCS), International Classification of Diseases (version ICD9-CM), and Diagnosis Related Group (DRG) codes to identify tests and procedures. See Appendix for coding details.

and for patient age in years.[6] We include the following time-varying county characteristics: Percent male, white, black, Hispanic, aged 65-74, 75-84, and 85+ and above, *ln*(population), active practicing non-federal physicians per capita, unemployment rate, *ln*(median household income), percent of Medicare enrollees receiving social security disability benefits, and managed care penetration (fraction of Medicare recipients enrolled in Medicare Advantage plans; linear and quadratic).[7]

Table 1 provides a covariate balance table for 2002, just before the third reform wave. We compare the means for outcome variables and covariates in new-cap and no-cap states. Treatment intensity is generally higher in the new-cap states. Stress echo is an exception, but patients in new-cap states receive more stress ECGs, SPECT tests, and more total stress tests. All differences are statistically significant due to the large sample size. A better measure of whether the differences are large is the "normalized difference" measure defined in Table 1 – roughly, how many standard deviations apart the two groups are.

Spending differences between new-cap and no-cap states are smaller; the new-cap states have somewhat lower Part A spending but higher Part B spending; slightly higher radiology spending, and similar lab spending.

With regard to covariates, Medicare recipients in new-cap states are slightly younger and healthier (fewer comorbidities) than those in no-cap states. New-cap states have a lower share of white population, and higher black and Hispanic shares. New-cap states are somewhat poorer (higher percent in poverty; lower median household income) and have lower managed care penetration than no-cap states.

3.2. Treatment and Control States

We identify treatment and control states relying on Avraham's (2014) Database of State Tort Reforms. We use the "exact" year in which a cap is adopted; in contrast, Avraham's spreadsheet accompanying his database time-shifts caps forward by 6 months (with some errors in coding the time shifting). Our treatment group is patients in nine treatment states adopted non-econ caps during the third reform wave of 2002-2005. This includes Florida (2003); Georgia (2005), Illinois (2005), Mississippi (2003), Nevada (2002), Ohio (2003), Oklahoma (2003), South Carolina (2005), and Texas (2003). The Georgia and Illinois caps were invalidated by state supreme courts in 2010; we consider these states as treated through 2009, but drop them from the sample for 2010 and after. Appendix Table App-1 provides additional information on each cap.

Our principal "narrow" control group is patients in 20 states which had no damage caps in place during our sample period Tennessee and North Carolina adopted caps in late 2011. We treat these

[6] Patient age is measured at the end of each year. In each calendar quarter, we measure comorbidities over the preceding four quarters. For the first four quarters in which a patient appears in the sample, we impute comorbidities back from the fifth quarter.

[7] We control for both managed care penetration and (managed care penetration)^2, because prior research (Frakes, 2013; Paik, Black and Hyman, 2017) finds evidence of a non-linear association between this covariate and Medicare spending. We obtain demographic characteristics from the Census Bureau. Source: http://www.census.gov/popest/. We use mid-year estimates of resident population -- inter-census estimates for 1991-2009, and post-census estimates for 2011-2013. We obtain data on number of physicians, unemployment rate, median household income, percent of Medicare enrollees receiving disability benefits, and managed care penetration (ratio of Medicare advantage enrollees/all Medicare eligible people over age 65) from the Area Health Resource File (AHRF). Source: http://arf.hrsa.gov/. We use the 2013-2014 Release. For physician counts for 2009, we interpolate between 2008 and 2010.

states as no-cap states in 2011. We also compare the new-cap states to a broader control group that also includes 22 "old cap" states, which had damage caps in effect throughout the sample period.

Many states have adopted a number of separate med mal reforms, often in packages. We view the results for damage caps as estimating the average effect of "serious" reform, with a damages cap as the central element, but often not the only element, of a reform package. Some studies of the effects of med mal reform either estimate the separate effects of a number of reforms, including damage caps, by including dummy variables for each reform in a single regression, or group different types of reforms together. We control separately for the principal reforms in sensitivity checks, but prefer our main specification because: (i) other reforms do not significantly affect med mal litigation outcomes (Paik, Black, and Hyman, 2013b); (ii) given this, we should not expect them to significantly affect our outcomes; and (iii) including other reforms in a regression model may provide misleading inferences for the impact of damage caps. We summarize our concerns in a note.[8]

3.3. Methodology

We study two main sets of outcome variables, using several variants on a difference-in-differences (DiD) research design. First, we study rates for diagnostic imaging tests. We study the three main diagnostic imaging tests: cardiac stress tests (any stress ECG, stress Echo, or SPECT), CT scans (other than SPECT) and MRI.[9] Second, we study rates for the most common interventional cardiac procedures: LHC, PCI and CABG, and for any revascularization procedure (PCI or CABG). Third, we study Medicare spending. We study the specific Part B subcategories for radiology and lab spending, which are especially likely to be influenced by assurance behavior. We also study Part A, Part B, and total Medicare spending.

DiD methods are a standard way to estimate the causal impact of legal changes, including adoption of damage caps. Atanasov and Black (2016) summarize the core requirements for DiD and other "shock-based" designs as: shock strength; shock exogeneity; "as-if random" assignment of patients to treated versus control states; covariate balance between treated and control states; and the only-through condition: the apparent effect of the shock on the outcome must be due only to the shock, not any other shock at around the same time.

[8] We do not control for other reforms for several reasons; see Paik, Black, and Hyman (2013b) for additional discussion. First, the number of usable reform events is often small. Also, suppose that a state adopts a damage cap and smaller reforms at or near the same time. States that adopt stricter damage caps may be more likely to adopt other reforms. This would bias the estimated effect of the damage cap (toward zero) and confound any effort to estimate the separate impact of the other reforms. Third, a number of reforms are not coded in Avraham (2014), the standard source that researchers rely on. An effort to code more reforms would make clear their massive colinearity, given the limited number of adopting states. In the Appendix, we study 7 other reforms with at least 3 reform events during our sample period, but omit 13 others, due to too few usable events or no one has coded them. We rely principally on Avraham's coding but hand-code certificate of merit laws.

[9] We consider all three types of stress tests together because there is often functional substitution between them. There is a further substitution possibility. In screening for possible coronary artery disease (CAD) ("ischemic evaluation"), physicians can begin with a stress test, and then progress to left-heart catheterization, a more accurate but invasive test, if the stress test is positive or ambiguous for CAD. Alternatively, physicians can proceed directly to LHC, without an initial stress test. In separate work (Farmer et al., 2017), we find that after cap adoption, overall ischemic evaluation rates are stable, but physicians substitute away from LHC toward stress testing as an initial screening test.

The DiD model makes the "parallel trends" assumption that the treated and control groups would have evolved in parallel, *but for* the treatment. This assumption is not directly testable, but one can assess whether trends appear parallel during the pre-treatment period. If pre-treatment trends are parallel, this makes it more likely that the parallel trends assumption is met, especially if there is also good covariate balance between treated and control states. Conversely, lack of covariate balance increases the risk that the parallel trends assumption would be violated in the post-treatment period, even if it is met in the pre-treatment period. Below, we provide graphs showing the year-by-year evolution treatment effects in event time, using the leads and lags models discussed below. The Appendix provides univariate graphs in calendar time of sample means for the three main groups of states: new-cap; no-cap; and old-cap.

The core innovations in our study include: (i) use of a very large, longitudinal, patient-level dataset which allows us to follow the same patient over time; (ii) use of extensive fixed effects and time-varying covariates to control for background factors that can affect outcomes; (iii) studying rates for specific procedures that are often believed to be sensitive to med mal risk; and (iv) careful assessment of whether any after-minus-before differences can be explained by non-parallel trends between treated and control states.

We use several graphical and regression approaches: (i) calendar-time graphs comparing treated and control groups; (ii) leads and lags graphs showing pre- and post-treatment trends; (iii) "simple DiD" regressions, that assume the cap effect "turns on" in the year after cap adoption; and (iv) "distributed lag" regressions, which allow the treatment effect to appear gradually over time. The simple DiD model allows for a one-time post-reform change in outcomes, is specified in Equation (1):

$$Y_{izt}=\delta_z + \gamma_t + \lambda\,(X_{it}) + \theta\,(X_{ct}) + \beta{*}(cap_{st}) + \varepsilon_{izt} \qquad (1)$$

Here *i* indexes patients, z indexes the zip-code in which the service was rendered (the δ_z are zip-code FE), and *t* indexes year (the γ_t are year FE). Y_{izt} is either a 0-1 dummy variable for tests or procedures (did patient *i* received that test or procedure in zip-code z in year *t*), or spending in one of the spending categories. X_{it} is a vector of patient characteristics and X_{ct} is a vector of time-varying county characteristics, with *c* indexing county. The treatment variable $cap_{st} = 0$ in control states for all *t*. In treated states, $cap_{st} = 0$ for years before the adoption year, =1 in years after the adoption year. For treated states, year 0 does not fit cleanly into either the pre- or post-reform period. For the cap adoption year, we drop that year for treated states. We use a linear probability model (LPM) rather than a logit or probit model because the extra computational demands of logit or probit estimation are prohibitive for our very large sample. Angrist and Pischke (2009, § 3.4.2) discuss why LPM, logit, and probit should, and in practice do, provide very similar estimates. Standard errors are clustered on state.[10] We also consider the patient*zip code fixed effects (FE) model in Equation (2), and a similar physician *zip code FE model:

$$Y_{izt} = \alpha_i + \gamma_t + \lambda\,(X_{it}) + \theta\,(X_{ct}) + \beta{*}(cap_{st}) + \varepsilon_{izt} \qquad (2)$$

To investigate whether pre-treatment trends differ between treatment and control states, we use a leads and lags model in event time, with the reform year set to zero, following Equation (3), and similar models that include patient or physician FE.

$$Y_{izt} = \delta_z + \gamma_t + \lambda\,(X_{it}) + \theta(X_{ct}) + \sum_{k=-4}^{6}(\ \beta^k * D_{zt}^{\ k}) + \varepsilon_{izt} \qquad (3)$$

[10] In robustness checks, standard errors are similar if we cluster on county.

Here, k indexes "event time" *relative* to the cap adoption year. $D_{zt}^{k} = 0$ for control states for all t and k. For treatment states, $D_{zt}^{k} = 1$ for the k^{th} year relative to the adoption year, 0 otherwise. For example, D_{zt}^{-4} takes the value of one four years before the non-econ cap adoption year, 0 otherwise; $D_{zt}^{+2} = 1$ two years after cap adoption year, 0 otherwise. Therefore, β^0 provides the estimated effect at the year when caps are enacted. β^1 provides the effect of reform one year after the enactment, and β^{-1} is the estimated effect one year before the reform's adoption. We include 4 leads (as many as our data will permit) and 6 lags in our specification. We combine years 6 and after into a single "lag 6+" dummy variable. We adjust the coefficients by subtracting β^{-3} from each, so that the *reported* $\beta^{-3} \equiv 0$.

We also report results from a "distributed lag" model, which allows for a different treatment effect in each post reform year. Without patient or physician FE, this model is:

$$Y_{izt} = \delta_z + \gamma_t + \lambda\,(X_{it}) + \theta(X_{ct}) + \sum_{k=0}^{5}(\ \beta^k * D_{zt}^{k\text{-lag}}) + \varepsilon_{izt} \qquad (4)$$

Here the first treatment lag D_{zt}^{0-lag} equals 1 in for a patient in a treated state in the cap adoption year and all subsequent years; D_{zt}^{1-lag} turns on in the year after reform, and stays on, D_{zt}^{2-lag} turns on in the second year after reform, and stays on, and so on for additional lags. Thus, the coefficient on D_{zt}^{0-lag} estimates the impact of reform in the year of reform; the coefficient on D_{st}^{1-lag} estimated the additional impact in the first full year after reform; the coefficient on D_{zt}^{2-lag} estimates the *additional* impact in the second year after reform, and so on. One can sum the lagged effects to obtain an overall treatment effect ($\sum_{k=0}^{n} \beta^k$), 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.

3.4 Benefits and Costs of Patient and Physician FE

We can follow patients over time, and thus potentially use patient FE to control for unobserved but time-invariant patient characteristics. That has important value in limiting the potential sources of omitted variable bias (OVB), and for our spending outcomes, is a preferred specification. However, for tests and procedures, patient FE have an associated cost. With patient FE, the causal effect we estimate is based only on patients in new-cap states who had repeat tests or procedures, one before and one after the time of tort reform, versus repeat patients in control patients. That is a much smaller sample of patients, which could lead to larger standard errors, depending on the balance between smaller sample size and greater potential precision due to the patient FE.

A second cost of using patient FE is subtle but important: The patients who receive repeat tests are, on average, likely in worse health than patients who do not, and we are estimating an effect for this non-random subset of the full patient population. Moreover, for revascularization, the first procedure changes the patient, likely in ways that our covariates do not effectively control for. The loss in sample size is also much greater for LHC, PCI, and CABG: the effective sample of treated patients varies from 65-84 thousand for the three imaging outcomes, but falls to 16 thousand for LHC, 3,500 for PCI, and only 64 for CABG.

We respond to these considerations as follows: (i) we report results for spending outcomes, where the gain from patient FE is clear, *with* patient FE; (ii) we report results for stress tests, MRI, and

CT scans, both with and without patient FE, and also report the effective sample size for patient FE regressions; and (iii) for surgical procedures (LHC, PCI, and CABG), report results without patient FE. Physician FE are an alternative to patient FE. We use them as an alternative specification for LHC, PCI, and CABG. The Appendix contains the FE results we do not report in the text: patient FE results for LHC, PCI, and CABG, and physician FE results for stress tests, MRI, and CT scans. Our results are consistent across all three specifications; we discuss the differences below.

4. Results for Diagnostic Imaging Tests and Cardiac Procedures

4.1. Imaging Tests

Physicians often cite fear of malpractice liability as an important driver of overuse of diagnostic testing (Katz and Williams, 2005; Kanzaria and Hoffman, 2015). We examine here the effects of damage cap adoption on rates for the three main cardiac stress testing (stress ECG, stress ECHO, and SPECT), and two major non-cardiac imaging tests: CT scans, and MRI.

4.1.1. Leads and Lags Graphs for Imaging Tests

Figure 1 presents leads-and-lags graphs of the treatment effects in event time for Any Stress Test, CT scan, and MRI, both without patient FE (left-hand graphs) and with patient*zip FE (right-hand graphs). The Appendix present results with physician*zip FE, as well as separate results for the three types of cardiac stress tests: SPECT, stress echo, and stress ECG. In Figure 1, the y-axis shows the coefficients on the annual lead and lag dummies; vertical bars show 95% confidence intervals (CIs) around each coefficient. We peg the coefficient for year -3 to zero, so there is no associated CI.

Consider first Any Stress Test. In both graphs, post-cap rates are higher than pre-cap rates. However, in both (more clearly without patient FE), this could simply reflect a continuation of a pre-treatment trend. We cannot tell with confidence whether the post-reform rise in rates that we observe in the graph is a true increase in response to reform, versus continuation of a pre-treatment trend. We can, however, conclude that there is no evidence of a drop in stress testing rates after cap adoption.

Figure 1 also shows a post-reform increase in the MRI and CT rates. For MRI, there is some evidence of non-parallel pre-treatment trends, which could continue after reform, with patient FE, but no similar trend without patient FE. For CT scans, the pre-treatment (lead) coefficients are reasonably flat in both graphs. Taken as a whole, there is no evidence that imaging rates fall after cap adoption and some evidence – strongest for CT scans – that imaging rates *rise*. We return in the discussion section to what might cause these increases, assuming they are real.

4.1.2. Regression Results for Imaging Tests

Table 2, Panel A, presents the results of simple DiD regressions, using no-cap states as the control group, without patient FE (first three regressions and with patient FE (next three regressions). These regressions assume a one-time change in outcomes, due to cap adoption. The coefficients on the damage cap variable can be interpreted as the change in the probability of receiving a stress test in a given year, due to a state adopting a damage cap. Without patient FE, the predicted effect for Any Stress Test is 4 additional tests per 1,000 patients, and is statistically significant ($t = 2.81$). However, as noted above, there is evidence of non-parallel pre-treatment trends. The point estimates are also positive for both MRI and CT scans, and are statistically significant for CT scans 3.9 additional scans per 1,000 patients (t = 2.70). However, here too, the positive coefficient for MRI could reflect continuation of non-parallel pre-treatment trends.

In Table 2, Panel B, we present distributed lag estimates, which allow the treatment effect to vary over time. We include the cap adoption year plus 3 lags; the last lag captures the average effect for year 3 and later years. This approach can better capture an effect that appears gradually over time – which from Figure 1 appears to be the case for all three outcomes. The principal cost of this approach is larger standard errors for the sum of coefficients estimate than for the simple DiD estimates. For Any Stress Test, the sum of coefficients estimate is 4.9 additional tests per 1,000 patients and is statistically significant (t = 2.74). The slightly higher sum of coefficients, relative to the simple-DiD estimate, is consistent with the treatment effect growing over time.

The sum of coefficients strengthens and becomes marginally significant for MRI, at 2.6 additional tests per 1,000 patients (t = 1.77) and strengthens and remains significant for CT scans, at 6.0 additional tests (t = 3.48). For both tests, the sum of coefficients estimates are substantially larger than the simple DiD estimates, consistent with the gradual response to reform suggested by the leads-and-lags graphs. In economic magnitude, the post-treatment rises, compared to base rates in the new-cap states of about 96 stress tests, 88 MRIs, and 190 CT scans, imply testing rate increases of about 5%, 3%, and 3% respectively -- "economically" meaningful, but not huge.

When we add patient*zip FE in regressions (4)-(6), the coefficients increase and are statistically significant for all three tests, despite much larger standard errors for CT scans. The distributed lag sums of coefficients are 6 additional stress tests, 4 additional MRIs, and 12 additional CT scans compare to a base testing rate for the population which receives two or more of the indicated test, of xx, yy, and zz per 1,000 patients, respectively. Thus the percentage increase is **[*to come]**

4.2. Results for Cardiac Procedures

The results in § 4.1 for imaging tests seem contrary to simple models of assurance and avoidance behavior, which suggest that a drop in malpractice risk should reduce screening tests and other forms of assurance behavior. We consider next the three most common invasive cardiac procedures: LHC, PCI, and CABG. LHC is a minimally invasive diagnostic test, which provides a more accurate assessment of coronary artery blockage than a noninvasive stress tests considered above. It can be ordered following an ambiguous stress test. It can also be ordered directly, if other evidence of CAD is seen as strong enough to justify this. LHC also a necessary precursor to revascularization through PCI or CABG. PCI is also minimally invasive – generally about a one-hour procedure, with no significant recovery period. CABG is open heart surgery -- a major operation, with significant operative mortality and a lengthy recovery.

One prior study (Avraham and Schanzenbach, 2015, below A&S) examines the effect of damage caps on treatment of heart attack patients. They find a post-cap increase in medical management and, for patients that receive revascularization, less PCI but more CABG. They interpret this relative change as physicians substituting a more remunerative, but risker procedure (CABG) for a less remunerative, safer procedure (PCI). However, their assumption that physicians have financial incentives to prefer CABG over PCI is problematic; we summarize our concerns in a note.[11]

---

[11] The statements in this footnote about cardiologist incentives reflect the judgments of Dr. Farmer, who is a non-invasive cardiologist. A&S assume that physicians have financial incentives to perform CABG rather than PCI, and that these incentives are constrained by malpractice risk. However, physician choice between PCI and CABG is

As we did for imaging tests, we present leads-and-lags graphs in event time, simple DiD regressions, and distributed lag regressions. However, in the regressions, we use physician FE rather than patient FE, for the reasons discussed above.

4.2.1. Leads and Lags Graphs for Cardiac Procedures

Figure 2 provides leads-and-lags graphs for cardiac procedures. Graphs without physician FE are on the left; graphs with physician FE are on the right. **[*to be rewritten once I see both graphs, buth what is here is wrong, there is a drop in LHC rates year +3]** For LHC, there is no obvious post-reform trend. However, there is evidence of a rising pre-reform trend between years -5 to 0. If that trend would have continued without reform, then the flattening in relative rates after year 0 might be interpreted as causal effect of damage caps, in reducing LHC rates.

For PCI, there is also some evidence of a rising relative trend, prior to reform, followed by a drop through year +3, and then partial recovery by year +6. A competing explanation is cardiologist response to the Courage trial, whose results were released in calendar 2007 (Boden et al., 2007), thus in event years +2 to +5, depending on state. This trial compared PCI to medical management for stable CAD, and found that PCI did *not* reduce either cardiovascular event rates or mortality from myocardial infarction (heart attack), when combined with optimal medical therapy. The response to this trial could plausibly vary by state, although we know of no reason to expect a larger response in the New-Cap states. If we assume that the pre-treatment trend would have continued after reform, the PCI graph provides some evidence for a post-cap drop in PCI rates in treated states. Our overall judgment is that the post-reform decline in relative PCI rates could be a response to tort reform, but that one cannot have much confidence in that attribution.

For CABG, there is some year-to-year bouncing, in both the pre-treatment and post-treatment period, but no strong visual evidence of a post reform change, in either level or trend. There is a possible trend toward lower relative CABG rates beginning in year +3.

The "any revascularization" results are driven by PCI, which is more common than CABG. These results should be interpreted with caution, because they implicitly assume that the clinical choice is to revascularize or not, with PCI and CABG as substitutes. Actual decision making is more complex. For patients without acute symptoms, the principal choice will often be medical management versus PCI; for others, with stronger need for revascularization, but no acute heart attack, the principal choice may be PCI versus CABG; while for patients with acute heart attack, the immediate intervention will be PCI, which may later be followed by CABG.

Overall, there is evidence of a modest drop in interventions, in contrast to the rise in non-invasive testing we saw in Section 4.1.

4.2.2. Regression Results for Cardiac Procedures

Table 3, for cardiac procedures, is similar in structure to Table 2, for imaging tests. It presents the results for simple DiD (Panel A) and distributed lag regressions (Panel B) for the three procedures and for any revascularization. Regressions (1)-(4) do not use either beneficiary or physician FE;

---

made (patient advice provided) by cardiologists, not cardiac surgeons. Non-invasive cardiologists should be neutral between the two approaches. Invasive cardiologists perform PCI but not CABG, and thus have financial incentives to perform PCI; while their malpractice risk concerns point toward referring patients to cardiac surgeons for CABG, so that someone else bears the malpractice risk. Thus, reducing malpractice risk should strengthen the incentives of interventional cardiologists to prefer PCI.

regressions (5)-(8) add physician FE. The simple DiD point estimates are uniformly negative, and are statistically significant for LHC and marginally significant for all revascularization. However, the point estimate are economically modest – the implied percentage drops are 2.5% for LHC 3% for PCI, and 3% for CABG.

These results strengthen in the distributed lag regressions, which allow for a gradual effect of cap adoption on clinical practice. The sums of coefficients are always larger in magnitude than the simple DiD coefficients, and are now statistically significant for any revascularization and marginally significant for PCI. The point estimates remain economically modest at around 4% for both LHC and any revascularization. Taken the leads-and-lags graphs and the regressions together, we have evidence – less than definitive – of a modest post-cap decline in procedure rates.[12]

5. Results for Medicare Spending

5.1. Laboratory and Radiology Spending

We turn next to spending and first consider the subcategories of Part B spending for radiology (including SPECT, MRI, and CT scans), laboratory tests, and combined lab and radiology spending. These spending categories are expected to be sensitive to a drop in assurance behavior. Our findings for spending in these categories is consistent with our assessment in Part 4 that damage caps, if anything, predict higher imaging rates.[13]Figure 3 provides leads and lags graphs for laboratory spending, radiology spending, and both categories combined. Radiology spending shows reasonably flat pre-treatment trends, and thus support the parallel trends assumption. The individual year point estimates are positive and statistically significant for each year from +1 on. There is a decreasing trend for laboratory spending during the pre-treatment period which is reversed around the time of damage cap adoption. Although all the point estimates are small and statistically insignificant, this trend reversal can be interpreted as causal effect of damage cap. Combined lab and imaging spending *rises* gradually after reform, but it is not statistically significant for each individual year. This result is robust to multiple alternative specifications.

Table 4, columns (1)-(3) presents the results of simple DiD and distributed lag regressions for lab, radiology, and all testing spending. In the simple DiD regressions, we find a small positive, but statistically insignificant coefficient on damage cap dummy for lab spending – the point estimate is $3.5 (around 1.6% of base spending). For radiology spending, we find a larger and statistically significant coefficient of around $10.6 (5.3% of base spending). The point estimate for combined lab and radiology spending is statistically significant and implies a spending rise of around 3.3%. The distributed lag results are similar. Taking the graphical and regression results together, there

[12] Our point estimates for PCI, CABG, and revascularization can usefully be compared with A&S. We estimate similar percentage drops in PCI and in total revascularization rates (A&S do not report percentage changes; we construct them from other data that they do report). But they find that CABG rates rise; we find that they fall; their estimate is well outside our 95% confidence bounds. The differing results can have a number of sources, including differing samples, differing reform events (they include West Virginia, which reduced its cap level in 2003, but exclude Illinois and lack data for Mississippi and pre-reform data for Oklahoma). Their CABG estimates also appear fragile – they do not appear in the partial leads-and-lags results that A&S report.

[13] We study radiology spending rather than the slightly broader category of imaging spending because spending data for tests that would be considered imaging but not radiology is not consistently captured in the early years of our dataset.

is evidence supporting higher combined lab and radiology spending following cap adoption, driven principally by higher radiology spending.

5.2. Overall Medicare Spending

We turn next, and last, to an assessment of the effect of damage caps on overall Medicare spending, a topic studied in several prior papers. Our principal contribution is to re-examine this issue with a large, patient-level dataset. We examine separately Part A spending, Part B spending, and total (Part A + Part B) spending.

5.2.2. Leads and Lags Results for Overall Spending

Figure 3 provides leads and lags graphs for these broad spending categories. The Part A graph shows a possible declining trend in relative spending prior to reform –actual spending increases in both new-cap and no-cap states, but increases a bit faster in no-cap states. The downward trend flattens out in year -1 and remains flat through year +1, followed by a drop through year +4, and then a recovery through year 6+. This delayed decline could be a temporary response to cap adoption. Overall, the Part A graph provides little evidence that cap adoption meaningfully affects Part A spending.

Part B spending increases after damage cap adoption through year +5, followed by a recovery by year 6+. However, there is some evidence of a pre-treatment trend toward higher spending which begins around year -5. The post-cap increase in part B spending could reflect continuation of this pre-treatment trend, rather than a true response to cap adoption.

The total spending graph is a blend of the Part A and Part B graphs: There is a declining trend between years -5 to -2, then an increasing period over years -2 to +1, a slight decrease in years +1 to +4, well short of statistical significance, and then a recovery by year 6+. We do not find strong evidence of decline in Medicare spending after damage cap adoption. Part B spending graph offers mild evidence of higher post-cap spending, and Part A graph provides evidence of a temporary decline. It is also troubling that there is evidence of non-parallel trends for both Part A and part B spending, even though these trends roughly offset each other, leading to a reasonably flat pre-treatment trend for total spending.

5.2.2. Regression Results for Overall Spending

Table 4, columns (3)-(5) presents simple DiD and distributed lag regressions for Part A, Part B, and total Medicare spending. Simple DiD point estimates for Part A spending is small and negative (at $17) and statistically insignificant. In contrast, the point estimate for Part B spending is $24 in additional spending (about 1% of base spending of $2,400) and is statistically insignificant.

The magnitudes of all coefficients slightly increase in the distributed lag regressions for radiology, lab, and part A spending, and it slightly decrease for Part B spending. consistent with the gradual rise we see in the leads-and-lags graphs. The Part A point estimate rises to 22 (around a 1% rise), but remains insignificant. The Part B point estimate drops to $19 and remains statistically insignificant. The total spending estimate is a negligible $2 decline and it is statistically insignificant. However, in the array of additional specifications we present in Table 5, the coefficient estimates for Part A spending are quite variable, and drive similar variability in total spending.

5.2.3. Comparison to Prior Results on Medicare Spending

Our results for overall Medicare spending can usefully be compared to Paik, Black, and Hyman (2017). They have only county-level data (rather than the patient-level data we rely on, but smaller standard errors, because they have data for the entire Medicare population; we have only a 5% sample). They found higher post-cap Part B spending, with no evidence for a change in Part A spending. In Figure 4, we illustrate the similarity in estimates for Part A and Part B spending by plotting both sets of results together (we convert our dollar estimates to percentages).

For Part B spending, our estimates and the Paik, Black and Hyman estimates are close to each other through year +2, and start diverging after, but both set of estimates are insignificant. For Part A spending, our estimates are similar to Paik, Black and Hyman for most years, they diverge in years +2 through +4.. The overall consistency of results across both papers, especially for part B spending, lends additional credibility to both sets of estimates.

6. Effects of Damage Caps on Mortality Measures

One of the consequences of damage cap adoption can deteriorate health outcomes if the policy leads to fewer test and procedures, which are essential for patients' health. We should not expect any impact on health outcomes since we only observe modest effect on utilization. In fact, the testing rates if anything increased for cardiac stress testing, CT, and MRI. We observe small decline in PCI and CABG rates, and it is likely that these marginal declines are the cases with little clinical values. We confirm this by looking at the various mortality rates as a measure of health outcomes. Figure 6 presents the leads and lags graph for overall mortality rate. There is some year-to-year bouncing, in both the pre-treatment and post-treatment period, but no visual evidence of a post reform change. We also estimate the effect of damage cap adoption on the mortality one year after hospitalization (presented in Appendix), and confirm that damage cap did not have any impact on mortality.

7. Additional Results and Limitations

We conduct an array of robustness and sensitivity tests on our main results reported above. We report selected results in this section and Table 5, and other results in the Appendix. In Table 5, Panel A presents results for imaging tests, and cardiac procedures; Panel B presents spending results.

7.1. Results with Varying Fixed Effects

We have a random 5% sample of Medicare fee-for-service beneficiaries that would allow us to follow beneficiaries over time. We can also observe the physicians who treated these beneficiaries over time. We exploit this unique feature of database for alternative specifications. First, we include physicians fixed effects in our specification, and in the second specification we include beneficiary fixe effects. This would allow us to capture some unobserved patient and physician characteristics that might derive health care utilization.

7.1.1. Results with Physician Fixed Effects

Including physician FE would allow us to investigate the change in probability of ordering specific tests by the same physician in the treatment states relative to physicians seeing similar patients in control states before and after damage cap adoption. We cannot conduct analysis with for Part A, and total spending because Part A claims consists of bundle of services that might be involved multiple healthcare providers which is unobserved to us. For imaging tests and procedures, we restrict the sample to those who have ordered 2+ of that test or procedure. For spending, we include all the physicians in our database. We then extract the entire cohort of patients who were visited by these physicians. We estimate

$$Y_{ijzt}=\alpha_i + \delta_z + \gamma_t + \lambda\,(X_{jt}) + \theta_1(X_{ct}) + \theta_2\,(X_{ct}) + \beta_*(cap_{st}) + \varepsilon_{izt} \qquad (4)$$

Here i indexes physician (the α_i are physician fixed effects), j indexes patient cohort who were visited by physician i, z indexes zip-code (δ_z are zip-code fixed effects), and t indexes year (γ_t is year fixed effect). Y_{ijzt} is either number of imaging tests and procedures ordered per all patients visited by any physician in any given year, or the average spending per patients for any physician. X_{jt} is a vector of patient cohort characteristics who were visited by each physician. It includes average age, percent patients male, percent of patients white, percent patients black, percent patients Hispanic, and 17 variables for percent patients with any of the Charlson comorbidity index. The remaining variables have similar definition as those in equation 1

Leads and lags graphs for imaging tests are reported in Figure 6. Leads and lags graphs are very similar to those without physician fixed effects for stress test and CT-Scan. Leads and lags graphs for MRI shows an increase starting right after damage cap adoption when physician fixed effects are not included; however, increase in MRI rate appears with a delay when we include physician FE. Results for cardiac interventions are presented in Figure 7. Results are very similar to those without physician FE, one notable difference is that the standard errors are generally larger when physician Fixed effects are included. Figure 9 presents the results for spending, Consistent with the previous findings, we observe evidence of increase in radiology spending, and perhaps Part B spending. Differential pre-treatment trends is more evident in the specifications with physician fixed effects. Increase in laboratory spending is more evident in the specifications with physician fixed effects. Overall, inclusion of physicians Fixed effects do not change our conclusion about the effect of damage caps.

7.1.2. Results with Patient FE

We include patient fixed effects in an alternative specification for robustness check. The patient and zip-code FE convey the strong advantage of controlling for otherwise unobserved, time-invariant patient and geographic characteristics. Patient fixed effects allow us to capture any time-invariant patient characteristics, including health status at the time of treatment, and health behaviors. The extensive fixed effects and patient and county covariates let us assess – in effect -- whether damage cap adoption predicts a change in imaging rates, or Medicare spending, for the same patient, in the same location, controlling for overall time trends and changes in patient and county characteristics.

Inclusion of patient fixed effects has a cost of effective sample size. They imply that we are studying the effect of reform only for patients who receive the same test or procedure both before and after reform, in the same zip code. This can be a restrictive assumption for less common procedures. We present a measure of effective sample size for our outcomes in the appendix. The effective sample sizes for patients in new-cap states remain large for spending and for the more common tests (any stress test, CT scan, MRI) but drop to around 16,000 for LHC, 3,500 for PCI and only 64 for CABG. We present the results with patient fixed effects for more common tests in the appendix. Results are in general agreement with the previous findings but there is less evidence of differential pre-treatment trends when we include patient fixed effects.

7.2. Results with Fewer or More Covariates

In addition to our basic patient, zip code and year FE, we use extensive patient- and county-level covariates, to control for time varying factors that could affect our outcome variables. However, no set of covariates can completely control for patient health. We therefore sought to assess whether our results were sensitive to the covariates we included, by rerunning our regressions with either (i) no covariates at all; (ii) all covariates except the dummy variables for the 17 elements of the Charlson comorbidity index; and (iii) our main covariates plus additional county-level health covariates for percent of the population obese, with diabetes, inactive, rate of death from heart disease, and percent of the population smokes daily (state-level).[14]

If we remove all time-varying covariates, inference is again similar (see Table 5). The coefficients and standard errors for the outcomes that are significant in our main specification rise; all outcomes remain significant except LHC. MRI, and Part B spending coefficients become statistically significant. Part A spending switches sign but remain statistically insignificant. .

Our main specification includes Charlson dummy variables to account for differences in patient health, which may affect treatment. We expect sicker beneficiaries, with more comorbidities, to consume more healthcare. However, causation could also run from local practice norms → healthcare utilization → more reported comorbidities. If so, controlling for comorbidities could mask the effect of tort reform. Coefficients become larger after excluding Charlson comorbidity index, and previously statistically significant results remain significant, and Part B spending becomes statistically significant in the new specifications. The coefficient for Part A spending switches sign

Motivated by the differences in estimates between the all- and no-covariates results, we searched for additional potential covariates, and found data on several county-level health characteristics (listed above). Results reported in the Appendix, are similar to our main specification.

7.3. Controlling for Other Tort Reforms

Notwithstanding our doubts, discussed above, about whether this is a sensible approach, we reran our regressions, controlling for seven other tort reforms (punitive damages cap, punitive damages evidence reform, collateral source reform, split recovery reform, periodic payment reform,

---

[14] Data are available for selected years; we use interpolation to fill in missing years. See Appendix for details.

certificate of merit requirements, and removal or weakening of joint and several liability). We present results for non-econ cap dummy in Table 5, and results for all reforms in the Appendix.

Coefficients and standard errors for the outcomes that are significant in our main specification become larger, and estimates for Any stress test becomes only marginally significant due to larger standard errors. Part B spending becomes statistically significant due to coefficient.. There is more variability in the coefficients for other outcomes.

7.4. Results with Broad Control Group

We reran all results using the broad control group, which includes both the 20 no-cap states and 22 old-cap states. We present leads-and-lags graphs with the broad control group in the Appendix. Table 5 presents our simple DiD results with the broad control group. The principal difference is that the negative coefficients for CABG strengthen and become statistically significant. Results with the two control groups are generally larger and more precisely estimated. This is consistent with leads and lags graphs (see Appendix), year by year coefficients are larger when we use the broad control group.

7.5. Adding State-specific Trends

A common robustness check, when non-parallel state trends may exist, is to add state-specific linear time trends. This model assumes that any non-parallel trends would have continued in the post-treatment period. We present simple DiD results with these trends in Table 5, for both the narrow and broad control groups. The largest change is for CT scans; the coefficient reverses sign and is statistically insignificant, with narrow control group, and becomes much smaller with broad control group and statistically insignificant. The coefficients switch sign for LHC, CABG, and PCI/CABG with both control groups. PCI coefficients becomes nearly zero with narrow control group, and switches sign with broad control group. .

Inference is similar for lab and radiology spending. However, the results for broader spending categories move quite a bit. The Part B coefficient strengthen and becomes statistically significant, and the Part A coefficients switch sign.

7.6. Equal State Weights and Leave One Out Regressions

We assess whether our results are driven by a particular state by: (i) including analytical weights in our regressions, which give equal weight to each treated state (our main specification gives equal weight to each beneficiary, and hence greater weight to larger states)[15]; and (ii) conducting "leave-one-out" regressions, in which we reran our results after removing one new-cap state from the sample. We present results in the Appendix. Coefficients are very similar for the results that are robust in Table 5 (Any Stress Test, radiology spending, and combined lab and radiology spending).

7.7.. Results with Synthetic Controls

We conducted synthetic control analyses, following Abadie, Diamond and Hainmueller (2010), for all results. The synthetic controls method does not provide analytical standard errors, but does

[15] We weight each state by inverse of the ratio of beneficiaries in that state over total number of beneficiaries in year 2002.

permit qualitative examination of whether the trends we find for the new-cap states, taken together, are consistent across states. The results are mixed. For Any Stress Test, for example, rates rise in 6 treated states relative to their synthetic controls, but fall in the other three; for CT scans, rates rise in 5 treated states, fall in one state, and are similar in the other three states**.**

7.8. Overview of Additional Tests

Which of our results seem strong, and which more fragile, after this array of additional tests? The results for Any Stress Test, and Radiology spending are robust across specifications. The CT scan results are otherwise strong, but disappear entirely with linear state trends. This concern is muted by the flat pre-treatment trends for CT scans (see Figure 1). MRI results are always positive but only statistically significant in some specifications.

The other results are less strong. Lab spending is always positive, but is significant only for the broad control group. Part B spending is significant in most specifications, but is not statistically significant in the main specification. Part A spending is never significant, and changes sign in some alternative specifications. Total Medicare spending is only occasionally significant.

8. Discussion

8.1. Plausibility: Why Might Imaging Rates Rise?

Our results for imaging rates are puzzling. The usual defensive medicine story posits that physicians overtest to protect against liability, and predicts that testing rates will fall after reform. In an alternative story above, tort reform could lead to both less assurance behavior (hence less testing and associated spending) but also less avoidance behavior (hence more spending), with no prediction for overall spending. But this alternative story does not explain why testing rates would rise. Nor can a third story, in which physicians test for multiple reasons, and don't react much to a change in malpractice liability.

Can a post-reform rise in testing rates be explained on theoretical grounds? We believe that it can, at least in part, based on close assessment of clinical context. That assessment has not been pursued in the defensive medicine literature. Consider the rise in stress testing rates. Physicians can assess the likelihood of CAD, and the possible need for intervention, by (i) starting with a stress test, and then proceeding to LHC if the stress test is positive or ambiguous for CAD; or (ii) starting with LHC. In Farmer et al. (2017), we use a physician FE specification, and find that initial screening rates do not change. Moreover, progression from an initial stress test to a followup LHC falls. Progression from initial ischemic evaluation to revascularization through PCI or CABG also falls. These results are consistent with physicians being more willing to tolerate clinical ambiguity when med mal risk falls, by: (i) accepting the less precise results from stress testing; and (ii) intervening less often, and relying instead on medical management.

A similar story can be told for other imaging tests. Consider the common case in which a patient comes to the emergency department (ED) with ambiguous symptoms. The emergency physician can either admit the patient for more careful evaluation, or conduct a "rule-out" MRI or CT scan, and then release the patient if the test is negative. In a lower-med-mal-risk environment, the physician may be more willing to "test and release" instead of admit. Testing rates would rise, but admissions from the ED would fall. Further research is needed to assess whether this story fits the data.

In a similar vein, physicians often face a choice between inpatient and outpatient surgery. In a lower-risk environment, they may be more willing to opt for outpatient surgery, which is lower cost and more convenient for the patient, but conveys small risks of adverse operative events (for example, severe allergic reaction to anesthesia) which cannot be addressed in in an outpatient setting. Such a shift could explain why we find some evidence of differential trends for Part A and Part B spending.

8.2. Heterogeneous Responses to Tort Reform

Damage caps are prominent on the state and national reform agendas because they are seen as a simple policy lever, which proponents claim will reduce defensive medicine and thus healthcare spending. One core message from the array of results presented here is that clinical response to this crude policy level is nuanced, and depends on clinical context. Rates appear to rise for several common imaging tests, and for overall lab and imaging spending, yet likely fall for the common cardiac interventions.

A second core message from our findings is that writ large, the "adopt damage caps, reduce spending" story lacks empirical support. Instead, measures that seek to reduce overtreatment will need to be targeted to particular areas of concern.

8.3. Limitations

Our study, like most others in this literature, is limited to the Medicare fee-for-service population; that is where the data is. Our results may not generalize to other populations, such as the uninsured or commercially insured populations. We study only third-wave damage cap adoptions, during 2002-2005. The second-wave reforms of the mid-1980s could have had different effects on healthcare spending.[16]

As we note above, the new-cap states are not a random subset of all states. They have higher stress testing and LHC rates than other states, and somewhat higher Medicare spending that the narrow control group of no-cap states. This affects the reliability of any inference that the post-cap changes we observe are caused by the cap adoptions.

It is unfortunate that CMS generally provides researchers who want national Medicare data with only a 5% sample, which is around 2 million beneficiaries. We turn out to be underpowered to reliably find statistical significance for changes in LHC and revascularization rates in the range of our point estimates, of around 4-5%.

9. Conclusion

Damage caps are physicians' preferred remedy for med mal risk. Physicians have long claimed that fear of med mal risk leads them to practice defensive medicine, including ordering unnecessary tests. Many policymakers have accepted the argument that adopting damage caps will reduce defensive medicine and thus reduce healthcare spending. We report evidence, from a careful study with a large, patient level dataset, of a much more complex and nuanced response to caps. Cardiac stress testing rates rise, instead of falling, MRI and CT scan rates likely rise as well, and overall Medicare lab and radiology spending also increases. Yet cardiac interventions do not

[16] Cardiac treatment has changed dramatically over the last several decades. Nonetheless it is interesting that our point estimates of a 4-6% decline in revascularization rates are close to the Kessler and McClellan (2002) estimate, for second wave reforms, of a 4-5% drop in hospital spending following heart attack.

rise, and likely fall – consistent with physicians continuing to conduct rule-out testing in a low-med-mal risk environment, but then – at least for cardia patients -- tolerate more clinical ambiguity and engage in less revascularization. There is no evidence of a fall in overall Medicare spending and, consistent with a recent prior paper (Paik et al., 2017), some evidence of higher Part B spending.

References

Ahmed, B., Dauerman, H. L., Piper, W. D., Robb, J. F., Verlee, M. P., Ryan, T. J., . . . Malenka, D. J. (2011). Recent Changes in Practice of Elective Percutaneous Coronary Intervention for Stable Angina. *Circulation: Cardiovascular Quality and Outcomes, 4*(3), 300-305. doi:10.1161/circoutcomes.110.957175

Andrus, B. W., and Welch, H. G. (2012). Medicare Services Provided by Cardiologists in the United States: 1999–2008. *Circulation: Cardiovascular Quality and Outcomes, 5*(1), 31-36. doi:10.1161/circoutcomes.111.961813

Angrist, Joshua, and Jorn-Steffen Pischke (2009), *Mostly Harmless Econometrics: An Empiricist's Companion*.

Avraham, Ronen (2007), An empirical study of the impact of tort reforms on medical malpractice settlement payments, 36 *Journal of Legal Studies* s186-s229.

Avraham, Ronen (2014), Database of State Tort Reforms (5th), working paper, at http://ssrn.com/abstract=2439198.

Atanasov, Vladimir, and Bernard Black (2016), Shock-Based Causal Inference in Corporate Finance and Accounting Research, *Critical Finance Review*, forthcoming, working paper at http://ssrn.com/abstract=1718555.

Baicker K, Fisher ES, Chandra A. Malpractice liability costs and the practice of medicine in the Medicare program. *Health Aff (Millwood).* 2007;26(3):841-852.

Boden, William E., O'Rourke, R. A., Teo, K. K., Hartigan, P. M., Maron, D. J., Kostuk, W. J., Knudtson, M., ... COURAGE Trial Research Group (2007), Optimal medical therapy with or without PCI for stable coronary disease. 356 *New England Journal of Medicine* 1503-1516.

Congressional Budget Office (2006), *Medical Malpractice Tort Limits and Health Care Spending*, at http://www.cbo.gov/ftpdocs/71xx/doc7174/04-28-MedicalMalpractice.pdf.

Currie, Janet, and W. Bentley MacLeod (2008), First Do No Harm? Tort Reform and Birth Outcomes, 123 *Quarterly Journal of Economics* 795-830.

Donohue, John J. III, and Daniel E. Ho (2007), The Impact of Damage Caps on Malpractice Claims: Randomization Inference with Difference-in-Differences, 4 *Journal of Empirical Legal Studies* 69-102.

Epstein, Andrew J., Daniel Polsky, Feifei Yang, Lin Yang, and Peter W. Groeneveld (2011), Coronary Revascularization Trends in the United States: 2001-2008, 305 *Journal of the American Medical Association 1769-1776.*

Farmer, Steven, Ali Moghtaderi, Samantha Schilsky, David Magid, William Sage, Nori Allen, Frederick Masoudi, Avi Dor, and Bernard Black, *Impact of Medical Liability Reform on Coronary Artery Disease Management* (working paper 2017).

Frakes, Michael (2012), Defensive Medicine and Obstetric Practices, 9 *Journal of Empirical Legal Studies* 457-481.

Frakes, Michael (2013), The Impact of Medical Liability Standards on Regional Variations in physician Behavior: Evidence from the Adoption of National-Standard Rules, *American Economic Review* 103: 257-276.

Hendel, R. C., Patel, M. R., Allen, J. M., Min, J. K., Shaw, L. J., Wolk, M. J., . . . Brindis, R. G. (2013). Appropriate Use of Cardiovascular Technology2013 ACCF Appropriate Use Criteria

Methodology Update: A Report of the American College of Cardiology Foundation Appropriate Use Criteria Task Force. *Journal of the American College of Cardiology, 61*(12), 1305-1317. doi:10.1016/j.jacc.2013.01.025

Kanzaria, H. K., Hoffman, J. R., Probst, M. A., Caloyeras, J. P., Berry, S. H., and Brook, R. H. (2015). Emergency Physician Perceptions of Medically Unnecessary Advanced Diagnostic Imaging. *Academic Emergency Medicine, 22*(4), 390-398. doi:10.1111/acem.12625

Katz, D. A., Williams, G. C., Brown, R. L., Aufderheide, T. P., Bogner, M., Rahko, P. S., and Selker, H. P. (2005). Emergency Physicians' Fear of Malpractice in Evaluating Patients With Possible Acute Cardiac Ischemia. *Annals of Emergency Medicine, 46*(6), 525-533. doi:http://dx.doi.org/10.1016/j.annemergmed.2005.04.016

Lucas, F. L., DeLorenzo, M. A., Siewers, A. E., and Wennberg, D. E. (2006). Temporal Trends in the Utilization of Diagnostic Testing and Treatments for Cardiovascular Disease in the United States, 1993–2001. *Circulation, 113*(3), 374-379. doi:10.1161/circulationaha.105.560433

Kessler, Daniel, and Mark B. McClellan (1996), Do Doctors Practice Defensive Medicine?, 111 *Quarterly Journal of Economics* 353-390.

Kessler, Daniel, and Mark B. McClellan, (2002), Malpractice Law and Health Care Reform: Optimal Liability Policy in an Era of Managed Care, 84 *Journal of Public Economics* 175-197.

Matlock, Daniel D., Peter W. Groeneveld, Steve Sidney, Susan Shetterly, Glenn Goodrich, Karen Glenn, Stan Xu, Lin Yang, Steven A. Farmer, Kristi Reynolds, Andrea E. Cassidy-Bushrow, Tracy Lieu, Denise M. Boudreau, Robert T. Greenlee, Jeffrey Tom, Suma Vupputuri, Kenneth F. Adams, David H. Smith, Margaret J. Gunter, Alan S. Go, and David J. Magid (2013), Geographic Variation in Cardiovascular Procedure Use Among Medicare Fee-for-Service vs Medicare Advantage Beneficiaries, 310 *Journal of the American Medical Association* 155-162.

Imbens, Guido W., and Donald B. Rubin (2014), *An Introduction to Causal Inference in Statistics, Biomedical and Social Sciences*.

Montanera, Daniel (2013), The Importance of Negative Defensive Medicine in the Effects of Malpractice Reform, working paper.

Morden , N. E., Colla , C. H., Sequist , T. D., and Rosenthal , M. B. (2014). Choosing Wisely — The Politics and Economics of Labeling Low-Value Services. *New England Journal of Medicine, 370*(7), 589-592. doi:doi:10.1056/NEJMp1314965

Paik, Myungho, Bernard Black, and David Hyman (2013a), The Receding Tide of Medical Malpractice Litigation Part 1: National Trends, 10 *Journal of Empirical Legal Studies* 612-638.

Paik, Myungho, Bernard Black, and David Hyman (2013b), The Receding Tide of Medical Malpractice Litigation Part 2: Effect of Damage Caps, 10 *Journal of Empirical Legal Studies* 639-669.

Paik, Myungho, Bernard Black and David A. Hyman (2017), *Damage Caps and Defensive Medicine Revisited*, *Journal of Health Economics*, forthcoming, working paper at http://ssrn.com/abstract=2110656.

Riley, Robert F., Creighton W., Don, Wayne Powell, Charles Maynard, and Larry S. Dean (2011), Trends in Coronary Revascularization in the United States from 2001 to 2009: Recent Declines in Percutaneous Intervention Volumes, 2011:4 *Circulation: Cardiovascular Quality and Outcomes* 193-197.

Rosenberg, A., Agiro, A., Gottlieb, M., and et al. (2015). Early trends among seven recommendations from the choosing wisely campaign. *JAMA Internal Medicine, 175*(12), 1913-1920. doi:10.1001/jamainternmed.2015.5441

Sclar, David and Michael Housman (2003), Medical Malpractice and Physician Liability: Examining Alternatives in Defensive Medicine, 4 *Harvard Health Policy Review* 75-84.

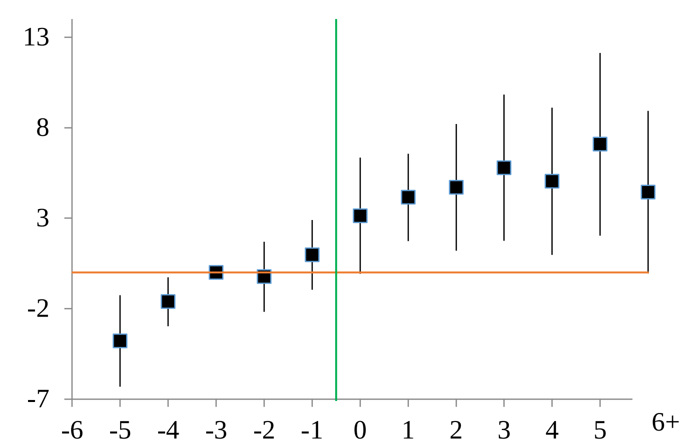
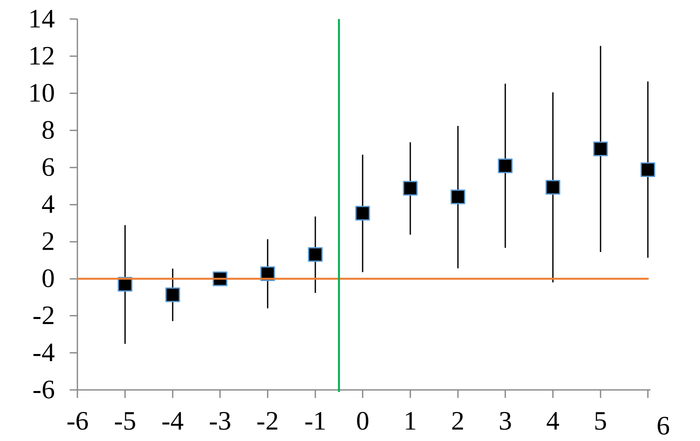
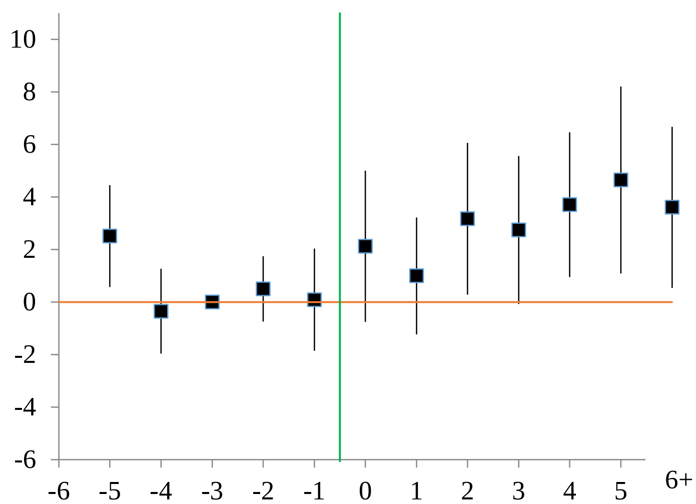
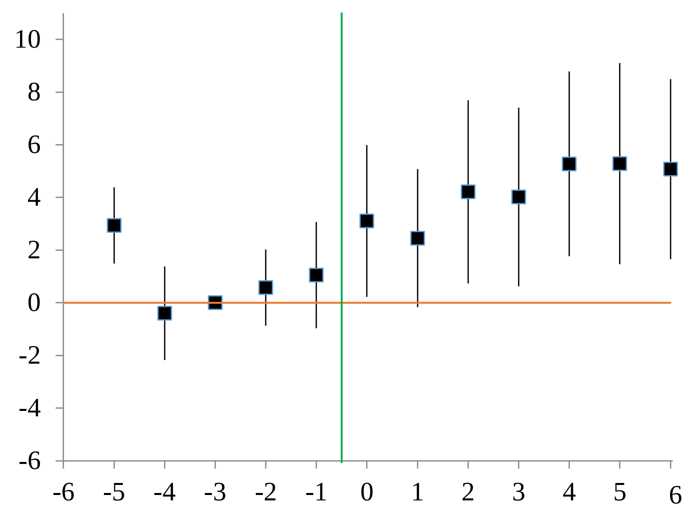
Sloan, Frank A., and John H. Shadle (2009), Is There Empirical Evidence for "Defensive Medicine"? A Reassessment, 28 *Journal of Health Economics* 481-491.

Waters, Teresa M., Peter P. Budetti, G. Claxton and JP Lundy (2007), Impact of State Tort Reforms on Physician Malpractice Payments, 26 *Health Affairs* 500-509.

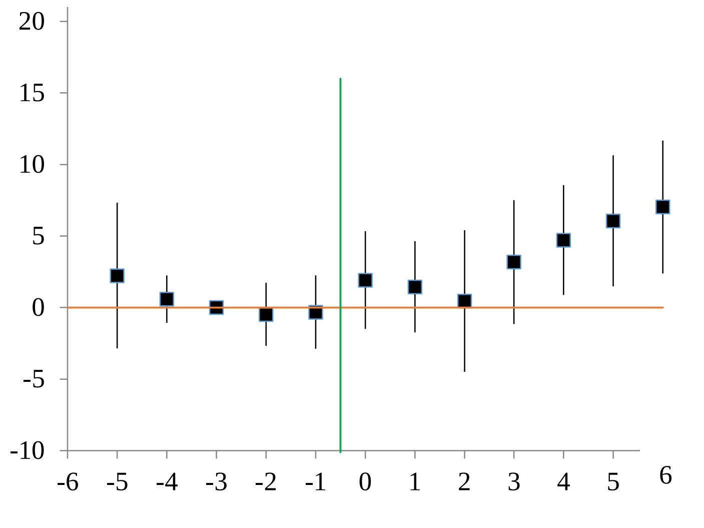
Yang, Y. Tony, Michelle M. Mello, S.V. Subramanian, and David M. Studdert (2009), Relationship Between Malpractice Litigation Pressure and Rates of Cesarean Section and Vaginal Birth After Cesarean Section, 47 *Medical Care* 234-242.

Figure 1. Imaging Rates: Leads and Lags Graphs of Effect of Damage Cap Adoption

Leads and lags regressions (linear probability model) of dummy variables for whether a patient had the indicated imaging test in a given year, for 9 new-cap states, versus narrow control group of 20 no-cap states , over 1999-2011. Leads and lags coefficients are multiplied by 1,000, so provide predicted effect of cap on annual rates per 1,000 patients. Regressions include zip-code, and year fixed effects, and covariates described in section 3.1. 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. Left hand graphs exclude, but right-hand graphs include, patient FE. Sample without patient FE: **[*to come]**. Sample with patient FE: 14,057,920 observations of **[*1,991,821]** distinct beneficiaries.

Without Patient FE	**With Patient FE**
Any Stress Test	
	
MRI	
	

CT-Scan



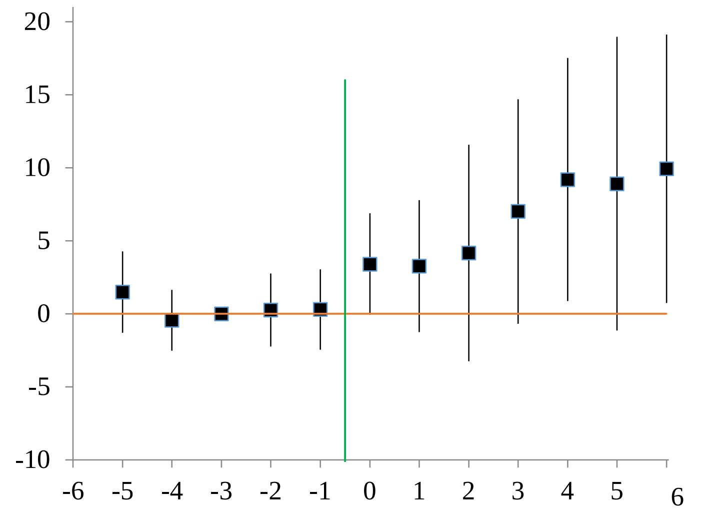


Figure 2- Cardiac Intervention Rates: Leads and Lags Graphs of Effect of Damage Cap Adoption

Leads and lags regressions (linear probability model) of dummy variables for whether a patient had the indicated procedure in a given year, for 9 new-cap states, versus narrow control group of 20 no-cap states, over 1999-2011. Coefficients on leads and lags are multiplied by 1,000, so provide predicted effect of cap on annual rates per 1,000 patients. y-axis shows the coefficients on the 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. Left hand graphs exclude, but right-hand graphs include, physican FE. Sample and covariates are same as in Figure 1.

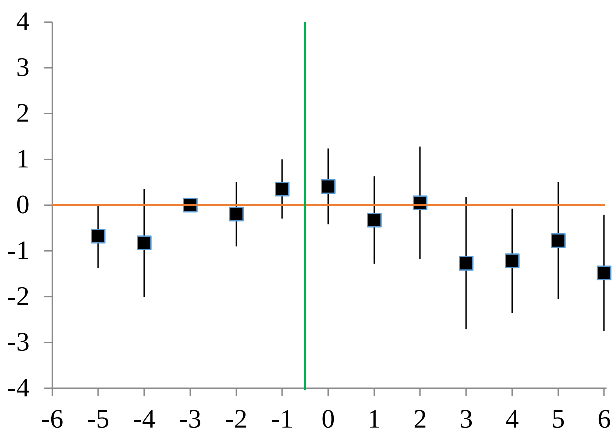
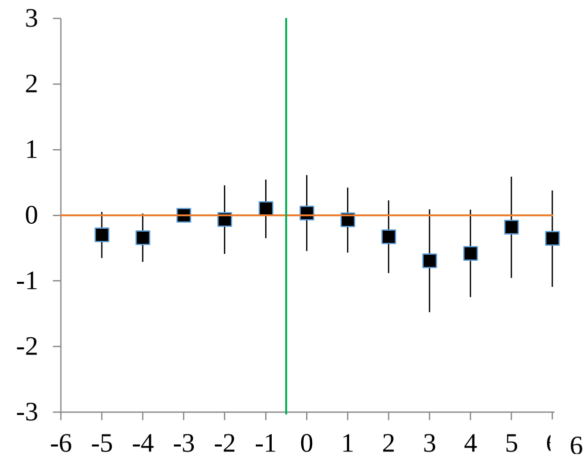
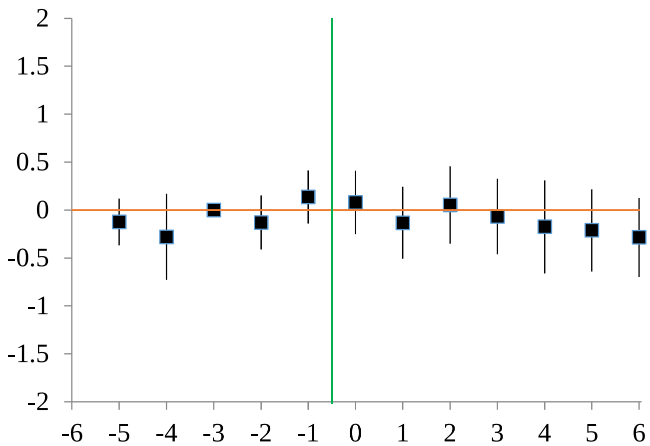
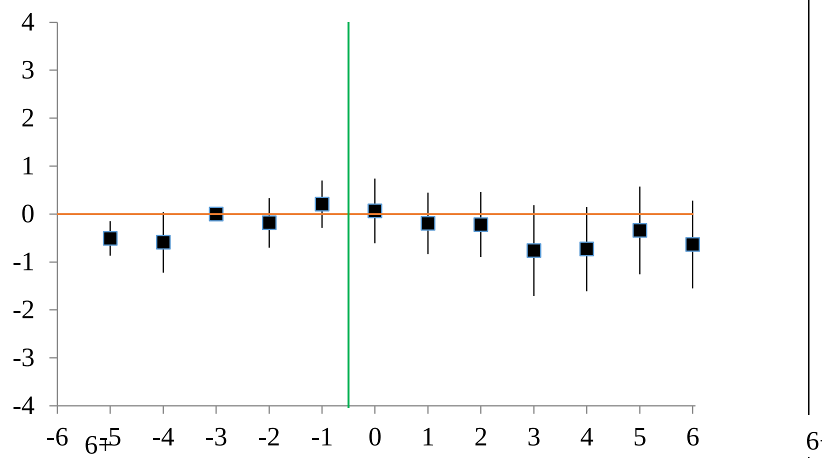
LHC	**PCI**
	
CABG	**Any Revascularization**
	

Figure 3- Medicare Spending: Leads and Lags Graphs of Effect of Damage Cap Adoption

Leads and lags regressions of outpatient laboratory, radiology spending, and combined (lab and radiology) spending per beneficiary over 2000-2011, and Part A, Part B, and total Medicare spending over 1999-2011, for 9 new-cap states versus narrow control group of 20 no-cap states. y-axis shows coefficients on the 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. Sample and covariates are otherwise same as in Figure 1. Amounts in 1999 $.

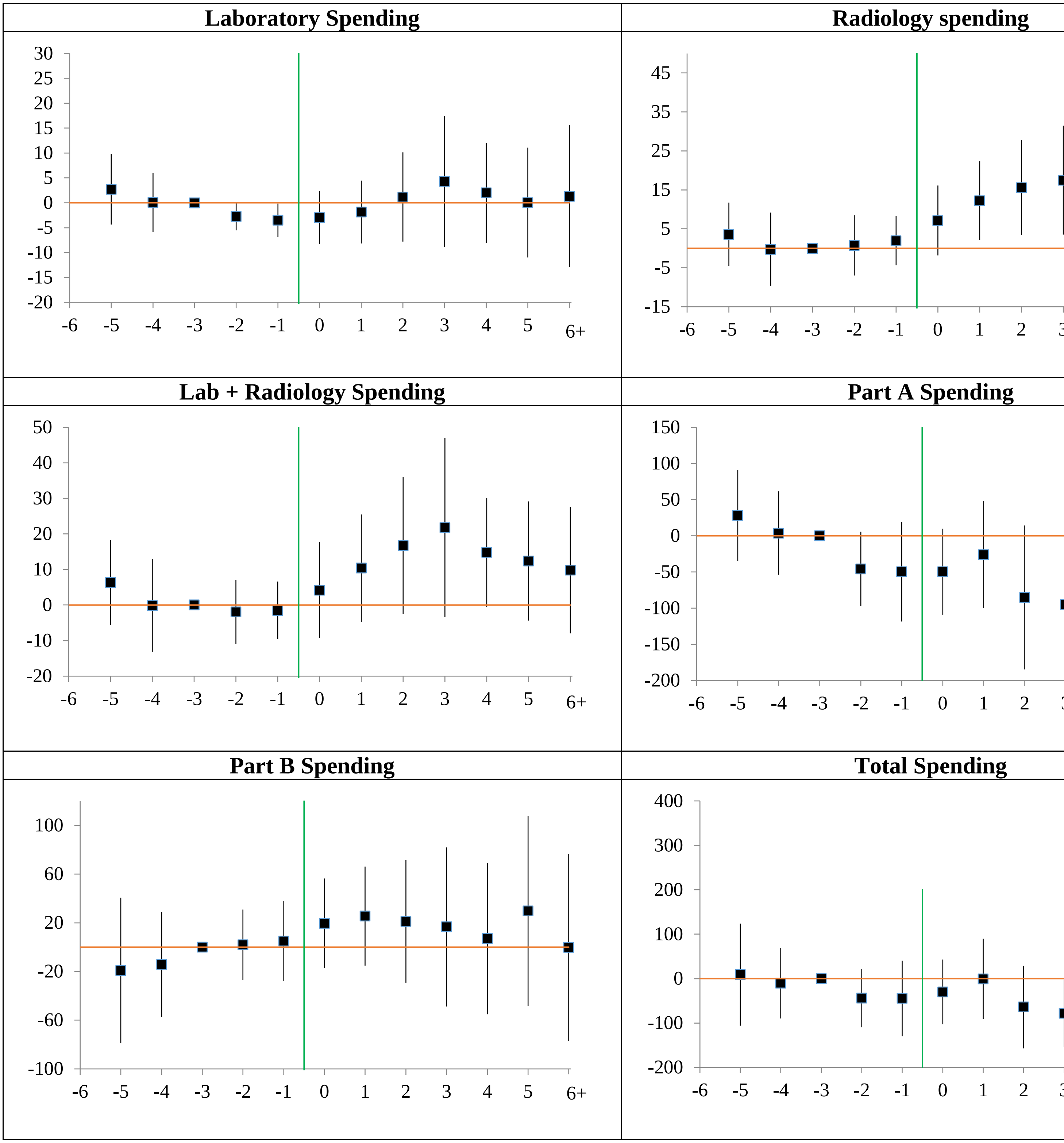
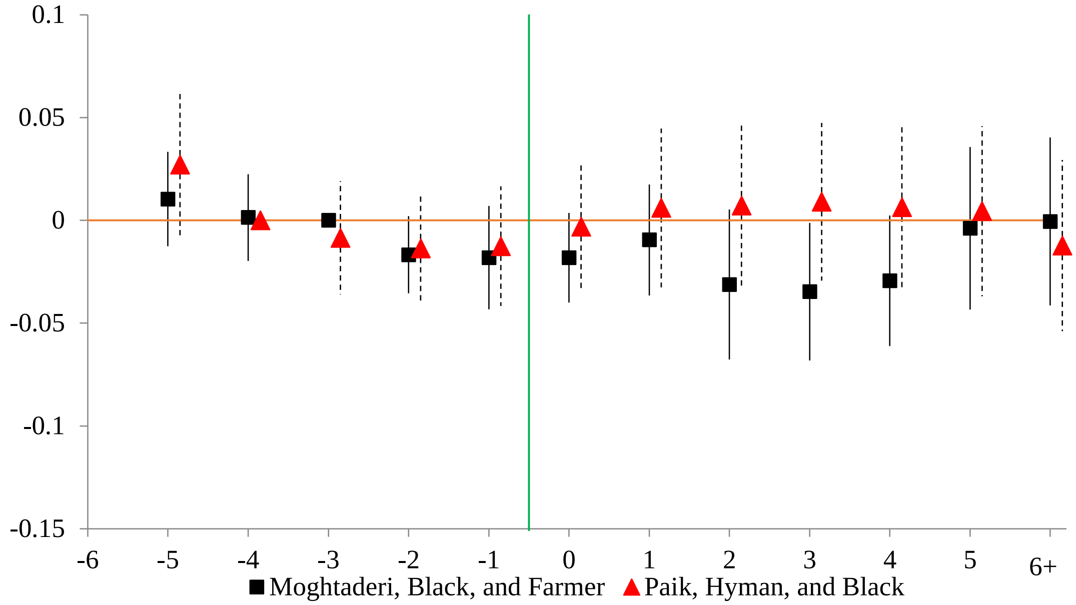


Figure 4 - County Level Spending: Leads and Lags

Figures compare our results for Part A and Part B Medicare spending from Figure 6 (converted from dollars to percent of 2002 spending) to results from Paik, Black and Hyman (2017), who have county-level data on Part A and Part B spending, and regress *ln*(Medicare spending per enrollee) on leads and lags relative to reform year, county and year fixed effects, covariates, and constant term, with weights based on average number of enrollees in each county over 1998-2011. Figure shows coefficients on leads and lags relative to year (*t* -4), which is set to zero. Vertical bars show 95% Cis around coefficients, using standard errors clustered on state.

Part A



Part B

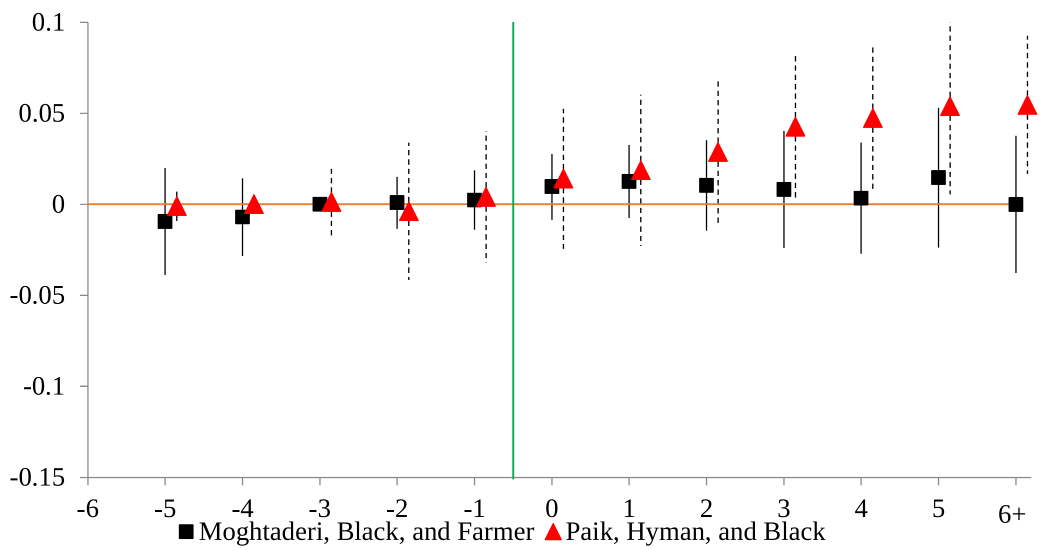


Figure 6- Mortality: Leads and Lags Graphs of Effect of Damage Cap Adoption

Leads and lags regressions (linear probability model) of dummy variables for whether a patient had died in a given year, for 9 new-cap states, versus narrow control group of 20 no-cap states, over 1999-2011. Coefficients on leads and lags are multiplied by 1,000, so provide predicted effect of cap on annual rates per 1,000 patients. Regressions include patient, zip-code, and quarter fixed effects, patient age dummies (for year of age, from 65 on), 17 dummy variables for elements of Charlson comorbidity index (measured annually using data from that year), and other covariates listed in Table 1. 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. Sample includes 12,020,886 observations of 1,991,821 distinct beneficiaries.

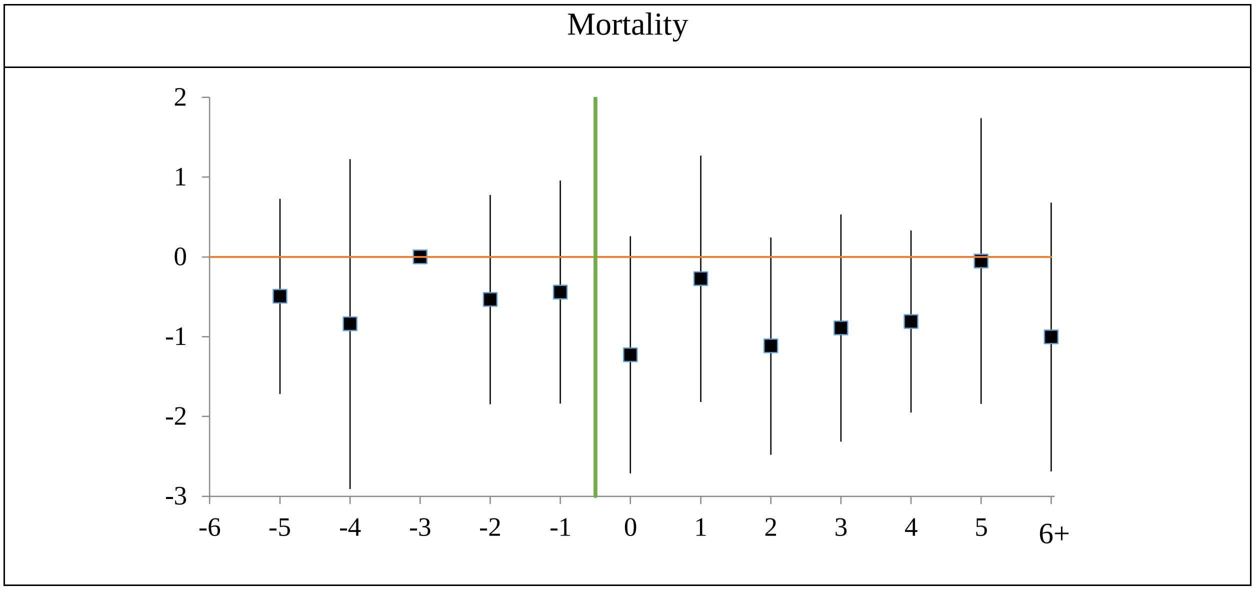
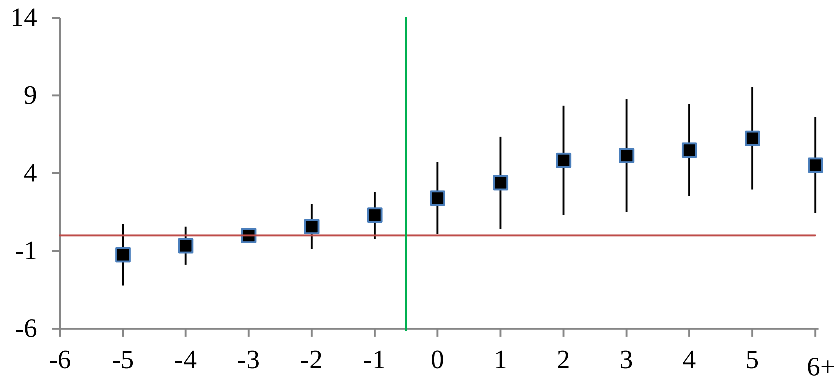


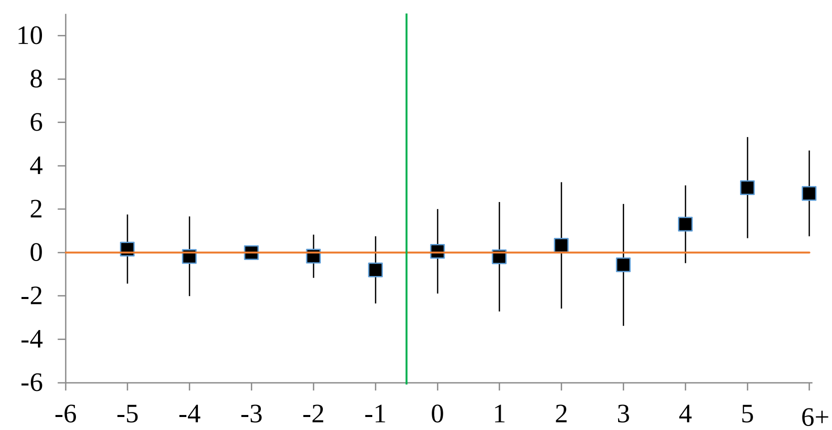
Figure 6. Imaging Rates: Leads and Lags Graphs with Physician Fixed Effects

Leads and lags regressions (linear probability model) of dummy variables for whether a patient had the indicated imaging test in a given year, for 9 new-cap states, versus narrow control group of 20 no-cap states , over 1999-2011. Leads and lags coefficients are multiplied by 1,000, so provide predicted effect of cap on annual rates per 1,000 patients. Regressions include physician, zip-code, and year fixed effects, and covariates described in section 3.1. 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. Sample includes 14,057,920 observations of XX distinct beneficiaries.

Any Stress Test



MRI



CT-Scan

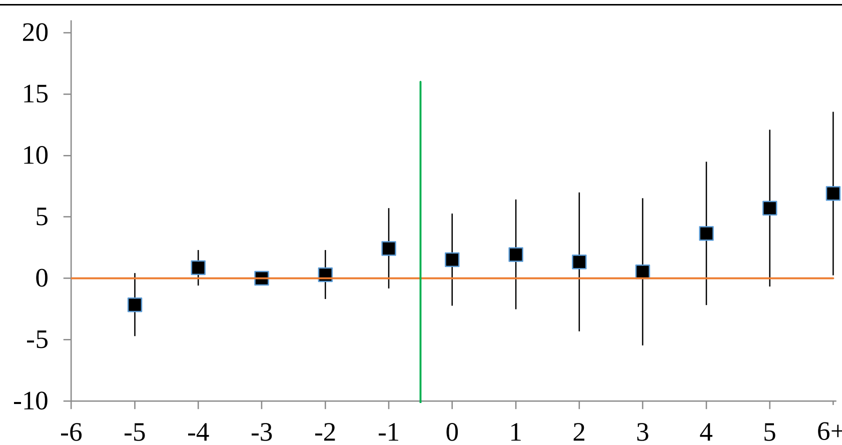
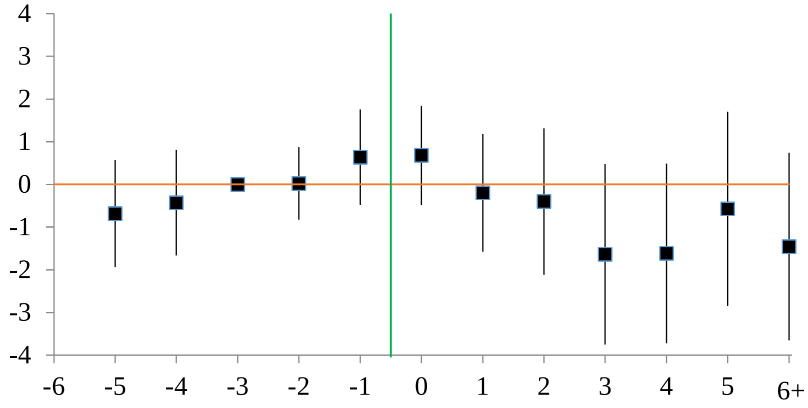
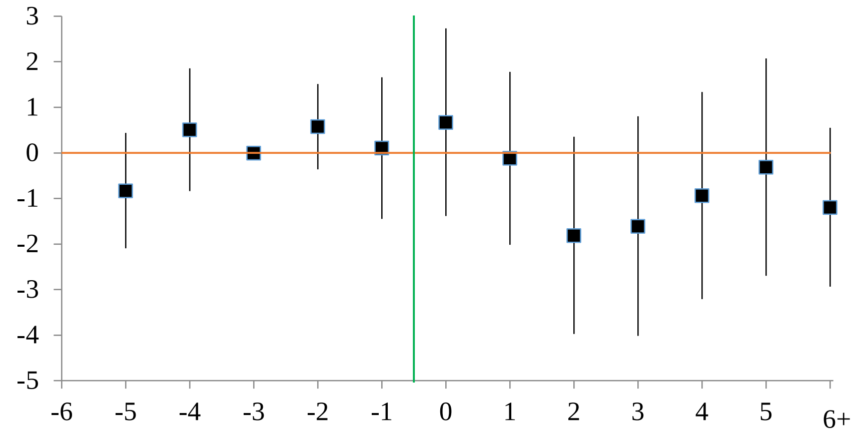
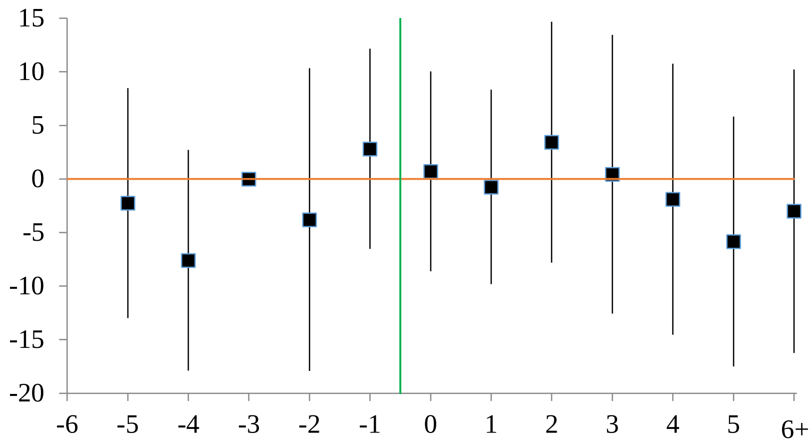


Figure 7- Cardiac Intervention Rates: Leads and Lags Graphs with Physician FE

Leads and lags regressions (linear probability model) of dummy variables for whether a patient had the indicated procedure in a given year, for 9 new-cap states, versus narrow control group of 20 no-cap states, over 1999-2011. Coefficients on leads and lags are multiplied by 1,000, so provide predicted effect of cap on annual rates per 1,000 patients. y-axis shows the coefficients on the 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. Sample and covariates are same as in Figure 1.

LHC	PCI
	

CABG



Any Revascularization

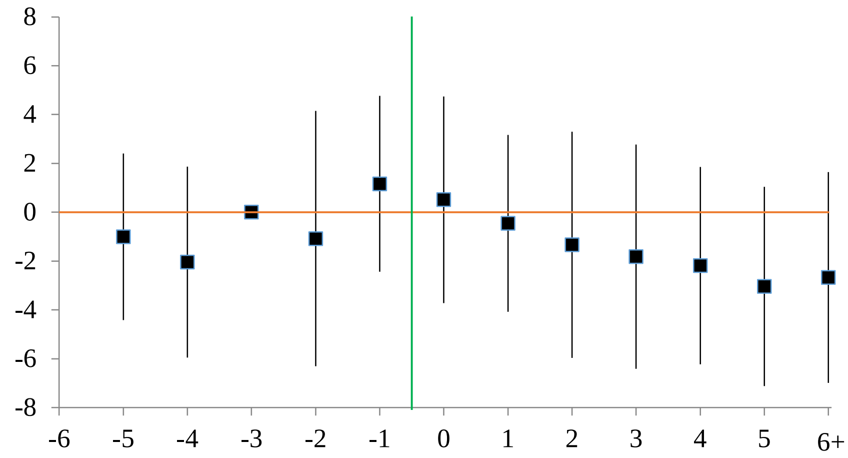
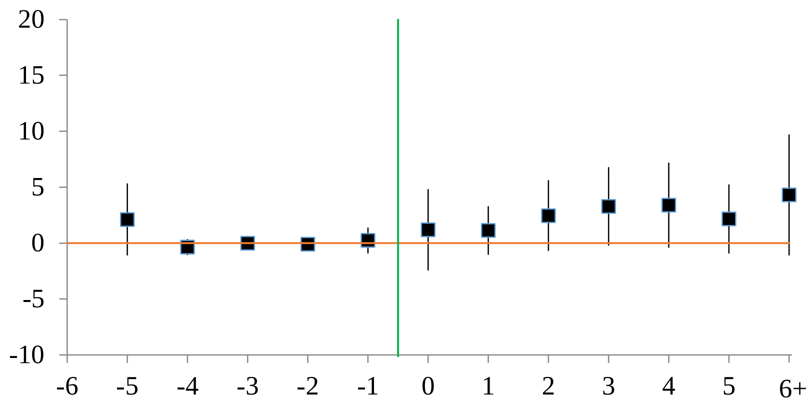
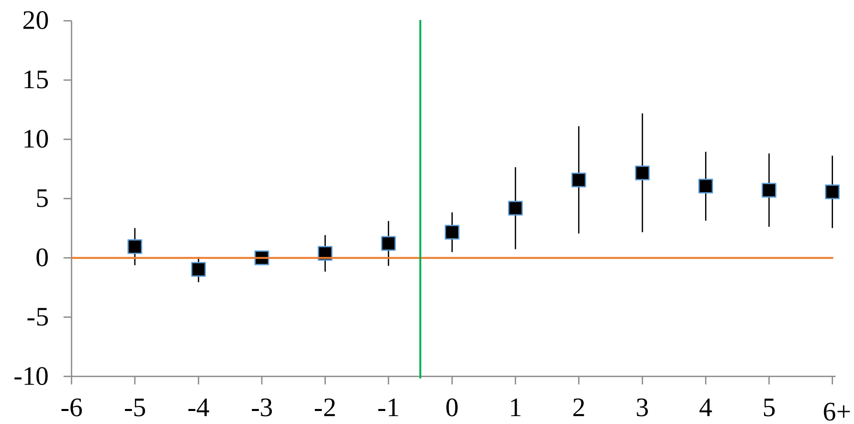
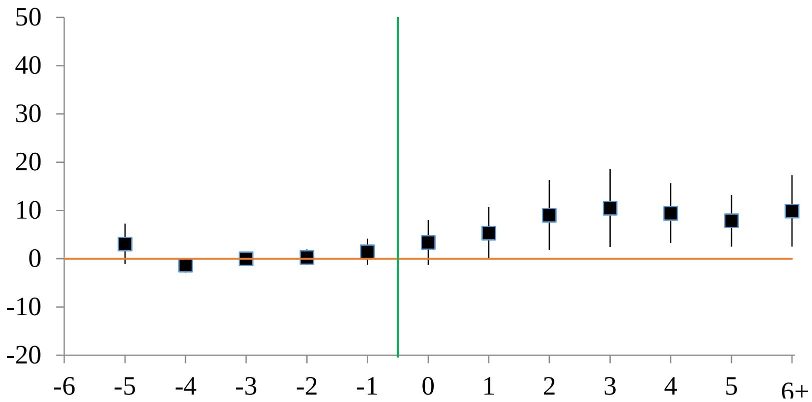


Figure 8- Spending: Leads and Lags Graphs with Physician FE

Leads and lags regressions (linear probability model) of dummy variables for whether a patient had the indicated procedure in a given year, for 9 new-cap states, versus narrow control group of 20 no-cap states, over 1999-2011. Coefficients on leads and lags are multiplied by 1,000, so provide predicted effect of cap on annual rates per 1,000 patients. y-axis shows the coefficients on the 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. Sample and covariates are same as in Figure 1.

Laboratory Spending	Radiology spending
	

Lab + Radiology Spending



Part B Spending

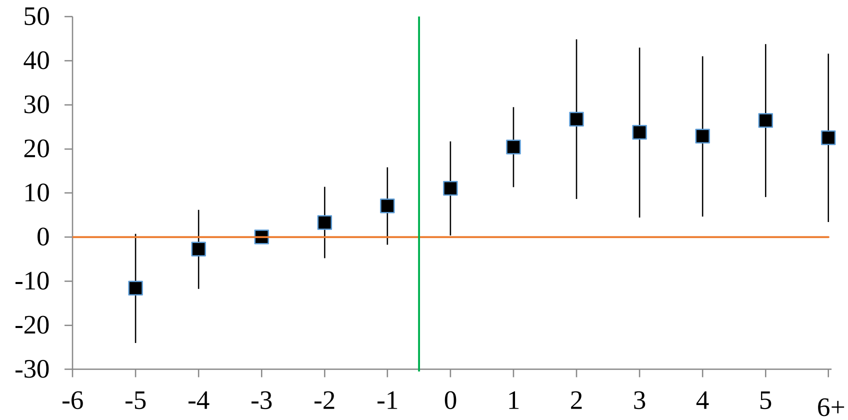


Table 1- Summary Statistics and Covariate Balance

Table presents summary statistics for 2002 (just before third reform wave), for outcome variables and averages for outcome variables and selected covariates, for 9 treated states versus 20 no-cap states, normalized difference, and two-sample *t*-test for difference in means. Amounts in 1999$. Normalized difference (ND) is defined as $ND_j = (\overline{x}_{jt} - \overline{x}_{jc}) / [(s_{jt}^2 + s_{jc}^2)/2]^{1/2}$ (see Imbens and Rubin, 2015). *t*-test is for two-sample difference in means. *, **, *** indicates statistical significance at the 10%, 5%, and 1% level; significant differences at 5% level are in **boldface**.

				New-Cap v. No-Cap	
States	**New-cap**	**No-cap**	**Old-Cap**	**ND**	**t-test**
Outcome measures					
Imaging (number per 1000 patients)					
stress echo	10.81 (0.41)	12.18 (0.42)		0.11	**2.34****
SPECT	74.88 (1.07)	61.64 (0.87)		0.47	**9.59*****
stress ECG	14.58 (0.30)	12.95 (0.28)		0.19	**3.91*****
any stress test	95.78 (1.09)	82.76 (0.92)		0.44	**9.10*****
MRI	87.73 (0.97)	81.07 (0.89)		0.25	**5.06*****
CT scan	190.23 (1.59)	185.88 (1.51)		0.10	**1.98****
Cardiac procedures (number per 1,000 patients)					
LHC	32.22 (0.53)	26.42 (0.41)		0.42	**8.68*****
LHC or stress test	110.02 (1.16)	94.56 (0.99)		0.49	**10.10*****
PCI	10.26 (0.18)	9.00 (0.18)		0.24	**4.95*****
CABG	4.99 (0.12)	4.59 (0.11)		0.12	**2.40****
PCI or CABG	14.82 (0.24)	13.28 (0.21)		0.23	**4.83*****
Medicare Spending per Enrollee (in 1999 $)					
Imaging	193.56 (2.38)	187.55 (1.93)		0.10	**1.96****
Laboratory spending	219.46 (2.53)	220.97 (2.24)		0.02	0.44
All testing spending	413.02 (4.74)	408.52 (3.96)		0.03	0.72
Part A	2732.96 (27.19)	2863.21 (41.68)		0.13	**2.62****
Part B	2033.04 (19.46)	1976.70 (18.17)		0.10	**2.11****
Total spending	4765.99 (42.63)	4839.91 (56.06)		0.05	1.05
Patient covariates					
Mean age	75.65 (0.046)	75.957 (0.037)		0.256	**5.24*****
Number of Charlson comorbidities	1.09 (0.01)	1.12 (0.01)		0.15	**3.03*****

Covariates (state averages, with population weights)					
Percent of population age 65-74	6.53 (0.08)	6.70 (0.05)		0.09	1.88*
Percent of population age 75-84	4.42 (0.06)	4.73 (0.04)		0.20	**4.13*** **
Percent of population above age 85	1.45 (0.02)	1.65 (0.01)		0.35	**7.21*** **
Percent white	80.16 (0.46)	82.41 (0.52)		0.16	**3.23*** **
Percent black	16.41 (0.45)	13.11 (0.47)		0.25	**5.05*** **
Percent Hispanic	16.30 (0.65)	8.92 (0.34)		0.50	**10.10*** **
Percent male	49.11 (0.04)	48.77 (0.04)		0.279	**5.72*** **
Percent below poverty line	13.24 (0.17)	11.75 (0.17)		0.30	**6.20*** **
Unemployment rate	5.95 (0.06)	5.82 (0.07)		0.07	1.40
Managed care penetration	10.78 (0.42)	13.04 (0.45)		0.18	**3.68*** **
Physician per capita	2.01 (0.04)	2.51 (0.06)		0.34	**7.08*** **
Percent of Medicare Enrollees who are disabled	14.24 (0.13)	14.73 (0.14)		0.12	**2.53** **
Population (millions)	0.11 (0.22)	0.11 (0.29)		0.01	**1.07**
Median household income ($ thousands)	41.01 (0.33)	43.78 (0.39)		0.26	**5.42*** **

Table 2. Regression Analyses: Effect of Damage Caps on Imaging Test Rates

Panel A: Simple DiD. Difference-in-differences regressions of dummy variables for whether a patient had the indicated test in a given year. Damage cap dummy =1 in new-cap states, in years with a cap in effect. We drop cap adoption year. There are 11,559,309 observations for 1,991,561 distinct beneficiaries.

Panel B: Distributed lags. Distributed lag regressions of dummy variables for whether a patient had the indicated test in a given year. Variable for "cap adoption year and after" =1 for treated states in cap adoption year and after; 0 otherwise. Variable for "cap year n and after" = 1 is similar but turns on in year n after cap adoption. Coefficients on covariates are suppressed. There are 12,020,886 observations for 1,991,821 distinct beneficiaries.

Both panels: Regressions use linear probability model. Coefficients on cap-related variables are multiplied by 1,000, to provide predicted effect of cap on annual rates per 1,000 patients. Regressions include indicated covariates, patient age dummies (for each year of age, from 65 on), and 17 dummy variables for elements of Charlson comorbidity index. Regressions (4)-(6) include patient FE (which absorb gender, race, and ethnicity). Sample period is 1999-2011. We drop IL and GA from treatment group for 2010 on, due to cap reversals in 2010. Standard errors, clustered on state, in parentheses. *,**, *** indicates statistical significance at the 10%, 5%, and 1% level. Significant results, at 5% level or better, in **boldface**.

Effective FE sample: Effective FE sample f is patients who receive the indicated two or more times. Effective *treated* FE sample is patients who receive the indicated test at least once before and once after reform.

Patient*zip FE	**No**	**No**	**No**	**Yes**	**Yes**	**Yes**
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Simple DiD						
Dependent variable	**Any Stress Test**	**MRI**	**CT**	**Any Stress Test**	**MRI**	**CT**
Damage cap dummy	**4.040*** (1.440)**	1.480 (1.210)	**3.920** (1.450)**	**5.300*** (1.320)**	**3.050** (1.350)**	**6.380** (2.720)**
Male	0.022*** (0.001)	-0.020*** (0.001)	-0.01*** (0.001)			
White	0.011*** (0.001)	0.016*** (0.001)	0.018*** (0.001)			
Black	-0.005*** (0.002)	-0.002* (0.001)	0.013*** (0.002)			
Hispanic	0.006*** (0.002)	0.012*** (0.003)	0.027*** (0.003)			
Fraction of population age 65-74	-0.137* (0.081)	-0.187 (0.114)	-0.122 (0.115)	-0.078 (0.085)	**-0.284** (0.132)**	-0.301 (0.246)
Fraction age 75- 84	0.186 (0.110)	0.062 (0.139)	-0.044 (0.226)	0.062 (0.174)	0.101 (0.176)	0.429 (0.330)
Fraction age 85+	-0.161 (0.344)	0.293 (0.186)	0.265 (0.373)	-0.041 (0.429)	**0.400* (0.209)**	0.682 (0.605)
Fraction white	-0.084* (0.041)	0.145*** (0.047)	0.296*** (0.063)	-0.036 (0.056)	0.039 (0.065)	**0.541** (0.220)**
Fraction Black	-0.202*** (0.072)	0.160*** (0.049)	0.341*** (0.088)	-0.178* (0.090)	0.091 (0.071)	**0.676*** (0.244)**
Fraction male	-0.006 (0.112)	-0.303*** (0.064)	-0.482*** (0.128)	-0.002 (0.137)	**-0.445*** (0.086)**	**-0.582*** (0.190)**
Fraction Hispanic	0.105*** (0.033)	-0.019 (0.024)	-0.035 (0.034)	**0.065** (0.029)**	-0.002 (0.024)	-0.006 (0.061)
Fraction below poverty line	0.011 (0.020)	0.002 (0.015)	0.040* (0.021)	0.009 (0.015)	-0.0003 (0.017)	0.055** (0.021)
Unemployment rate	0.041 (0.026)	0.018 (0.021)	0.059 (0.036)	0.028 (0.027)	-0.0005 (0.024)	0.081* (0.045)
Fraction of population disabled	-0.028 (0.049)	-0.031 (0.042)	0.090* (0.051)	-0.064 (0.040)	-0.013 (0.047)	0.183* (0.106)
Ln (population)	0.015*	0.024***	0.038***	0.014	**0.023***	**0.063***

Patient*zip FE	No	No	No	Yes	Yes	Yes
	(1)	(2)	(3)	(4)	(5)	(6)
	(0.009)	(0.007)	(0.010)	(0.010)	**(0.007)**	**(0.017)**
Physicians/1000 population	-0.003**	0.001	0.003	-0.003*	**0.001****	0.0005
	(0.001)	(0.001)	(0.002)	(0.002)	**(0.0006)**	(0.003)
Ln(household median income)	0.021***	0.001	0.001	**0.018*****	-0.009	-0.003
	(0.005)	(0.009)	(0.008)	**(0.005)**	(0.008)	(0.008)
Medicare penetration	-0.021*	-0.024***	-0.008	-0.017	**-0.027*****	-0.013
	(0.011)	(0.008)	(0.010)	(0.014)	**(0.009)**	(0.023)
(Medicare penetration)^2	0.007	0.007	-0.045**	0.003	0.016	-0.050
	(0.025)	(0.015)	(0.021)	(0.030)	(0.017)	(0.039)
Constant	0.060	0.050	-0.042	-0.037	**0.204*****	-0.068
	(0.054)	(0.046)	(0.082)	(0.076)	**(0.062)**	(0.215)
R^2	0.046	0.035	0.106	0.36	0.34	0.41
Effective FE Sample				310,610	337,918	750,306
Effective treated FE Sample				73,123	65,275	149,068
Panel B. Distributed Lags						
Cap adoption year or after	1.120	-1.020	2.740*	**3.16****	2.40*	**4.48*****
	(1.190)	(1.020)	(1.510)	**(1.160)**	(1.360)	**(1.500)**
Cap year 1 or after	2.420*	1.630	-0.529	1.220	-0.560	0.695
	(1.370)	(1.150)	(1.730)	(1.410)	(0.906)	(1.540)
Cap year 2 or after	0.292	0.564	-0.404	-0.281	0.911	0.889
	(0.794)	(0.702)	(1.700)	(0.881)	(0.695)	(1.650)
Cap year 3 or after	1.110	1.400	4.170*	1.47*	0.013	2.660
	(0.903)	(1.010)	(2.340)	(0.865)	(0.917)	(1.740)
Sum of coefficients	**4.943****	2.570*	**5.978*****	**6.082****	**4.278****	**12.483*****
	(1.802)	(1.449)	**(1.719)**	**(2.254)**	**(1.985)**	**(4.307)**
R^2	0.046	0.035	0.106	0.36	0.34	0.40

Table 3. Regression Analyses: Effect of Damage Caps on Cardiac Intervention Rates

Panel A: Simple DiD. Difference-in-differences regressions of dummy variables for whether a patient had the indicated procedure in a given year.

Panel B: Distributed lags. Distributed lag regressions of dummy variables for whether a patient had the indicated procedure in a given year.

Both panels: Regressions use linear probability model. Coefficients on cap-related variables are multiplied by 1,000 , to provide predicted effect of cap on annual rates per 1,000 patients. Sample, main variables, covariates, and fixed effects are same as in Table 2; coefficients on covariates are suppressed. Regressions (5)-(8) include physician*zip code FE. Standard errors, clustered on state, in parentheses. *,**, *** indicates statistical significance at the 10%, 5%, and 1% level. Significant results, at 5% level or better, in **boldface**.

Physician FE	**No**	**No**	**No**	**No**	**Yes**	**Yes**	**Yes**	**Yes**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable	**LHC**	**PCI**	**CABG**	**Any Revasc.**	**LHC**	**PCI**	**CABG**	**Any Revasc.**
Panel A- Simple DiD								
Damage cap dummy	**-0.803****	-0.297	-0.144	-0.409*				
	(0.391)	(0.194)	(0.109)	(0.232)				
R^2	0.026							
Panel B- distributed lags								
Cap adoption year or after	-0.111	0.093	-0.040	0.018				
	(0.379)	(0.376)	(0.155)	(0.362)				
Cap year 1 or after	-0.561	-0.207	-0.211	-0.342				
	(0.519)	(0.448)	(0.206)	(0.450)				
Cap year 2 or after	0.488**	-0.190	0.279*	0.112				
	(0.235)	(0.161)	(0.151)	(0.189)				
Cap year 3 or after	-1.080**	-0.164	-0.202	-0.407*				
	(0.413)	(0.166)	(0.146)	(0.225)				
Sum of coefficients	**-1.264*****	-0.468*	-0.175	**-0.619****				
	(0.444)	(0.244)	(0.138)	**(0.290)**				
R^2	0.026	0.011	0.007	0.015				

Table 4. Regression Analyses: Effect of Damage Caps on Medicare Spending

Panel A: Difference-in-differences regressions for indicated Medicare spending categories per enrollee, per *calendar year*, on damage cap dummy and covariates. Coefficient on damage cap dummy are multiplied by 4, so coefficients provide estimated effect of cap on annual spending per patient. Coefficients on covariates are suppressed, but presented in the Appendix (Table App 3).

Panel B: Distributed lag regressions for indicated Medicare spending categories per enrollee, per *year*. Distributed lag analysis in a given year. Coefficients on covariates are suppressed.

Both panels: Sample, main variables, covariates, and fixed effects are same as in Table 2; coefficients on covariates are suppressed, but presented in the Appendix for simple DiD.. Amounts in 1999 $. Standard errors, clustered on state, in parentheses. *,**, *** indicates statistical significance at the 10%, 5%, and 1% level, respectively. Significant results, at 5% level or better, in **boldface**.

VARIABLES	(1) **Laboratory**	(2) **Radiology**	(3) **Lab + Radiology**	(4) **Part A**	(5) **Part B**	(6) **Total**
Panel A. Simple DiD						
Damage cap dummy	3.585	10.63***	14.21**	-17.52	24.99	7.471
	(4.039)	(2.927)	(5.923)	(36.25)	(20.45)	(35.92)
R^2	0.196	0.117	0.192	0.088	0.175	0.134
Panel B- distributed lags						
Cap year adoption and after	11.57***	-2.300	9.273**	7.159	57.66**	64.82
	(2.128)	(2.483)	(4.006)	(32.83)	(22.21)	(44.27)
Cap year 1 or after	-11.43***	12.98***	1.546	-5.202	-38.59	-43.80
	(3.369)	(4.020)	(6.386)	(41.86)	(26.87)	(56.72)
Cap year 2 or after	2.123	2.964	5.087	-44.33	11.52	-32.80
	(2.318)	(2.508)	(3.997)	(34.06)	(17.25)	(43.53)
Cap year 3 or after	2.134	-1.598	0.536	19.98	-10.71	9.273
	(2.083)	(2.080)	(2.874)	(25.62)	(18.47)	(40.50)
Sum of coefficients	4.395	12.046***	**16.442****	-22.387	19.880	-2.507
	(5.560)	(3.586)	**(7.157)**	(45.040)	(28.264)	(42.271)
R-squared	0.196	0.117	0.192	0.088	0.175	0.134

Table 5. Additional Regression Results

Difference-in-differences and distributed lag regressions. Specification and sample is same as in Tables 2-4, except as indicated. Broad control group includes 20 no-cap states and 22 old-cap states; sample is 21,466,430 observations of 2,912,005 distinct beneficiaries over 1999-2011. Dollar amounts are in 1999 \$. Standard errors, clustered on state, in parentheses. *,**, *** indicates statistical significance at the 10%, 5%, and 1% level. Significant results, at 5% level or better, in **boldface**.

Panel A: Imaging and Cardiac Procedures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep. Variables	**Any stress test**	**MRI**	**CT**	**LHC**	**PCI**	**CABG**	**PCI or CABG**
Results from Tables 2-3 (narrow control group)	**4.040*** (1.440)**	**1.480 (1.210)**	**3.920** (1.450)**	**-0.803** (0.391)**	-0.297 (0.194)	-0.144 (0.109)	-0.409* (0.232)
Remove all time-varying covariates	**6.080** (2.310)**	**3.820** (1.660)**	**8.410*** (2.160)**	-0.474 (0.369)	-0.280 (0.227)	-0.157 (0.142)	-0.397 (0.310)
Remove Charlson dummies as covariates	**4.480*** (1.550)**	1.920 (1.210)	**5.610*** (1.600)**	-0.568 (0.359)	-0.217 (0.190)	-0.126 (0.104)	-0.315 (0.221)
Narrow control group and control for other reforms	4.510* (2.260)	2.000 (1.510)	**5.560*** (1.470)**	-0.495 (0.433)	-0.122 (0.168)	-0.185 (0.134)	-0.279 (0.253)
Broad control group (w. main specification)	5.280*** (1.210)	2.660** (1.170)	5.190*** (1.160)	-0.448 (0.338)	-0.042 (0.133)	-0.241** (0.093)	-0.264 (0.177)
Narrow control group with state trends	3.940*** (1.100)	1.060 (1.020)	-0.188 (1.450)	0.400 (0.401)	0.005 (0.252)	0.068 (0.129)	0.076 (0.332)
Broad control group with state trends	3.990*** (1.120)	1.960* (1.140)	0.448 (1.440)	0.360 (0.429)	-0.043 (0.261)	0.064 (0.128)	0.030 (0.344)

Results from Tables 2-3 (narrow control group)							
Use patient RE and state FE							
Broad control group (w. main specification)							

Panel B: Medicare Spending

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Variables	**Laboratory**	**Radiology**	**Lab + Radiology**	**Part A**	**Part B**	**Total**
Results from Table 4 (narrow control group)	3.585 (4.039)	**10.63***** **(2.927)**	**14.21**** **(5.923)**	-17.52 (36.25)	24.99 (20.45)	7.471 (35.92)
Same but remove all time-varying covariates	7.540 (6.722)	**14.04***** **(2.998)**	**21.58**** **(8.218)**	20.43 (42.19)	**86.87**** **(34.92)**	**107.3**** **(49.88)**
Remove Charlson dummies as covariates	6.113 (4.536)	12.30*** (3.107)	18.41** (6.698)	28.38 (40.32)	57.00** (24.36)	85.38* (44.04)
Add controls for other reforms	8.798* (4.755)	**12.17***** **(2.886)**	**20.96***** **(6.954)**	-45.87 (49.84)	**71.92***** **(20.34)**	26.05 (40.07)
Broad control group (w. main specification)	9.865** (3.926)	9.831*** (2.422)	19.70*** (4.803)	-34.49 (30.98)	47.19** (22.63)	12.71 (30.89)
Narrow control group with state trends	4.415 (3.171)	13.92*** (4.988)	18.33** (7.498)	5.103 (29.07)	38.72** (16.58)	43.83 (34.09)
Broad control group with state trends	5.676* (2.974)	15.69*** (4.965)	21.37*** (7.111)	29.99 (30.60)	44.91** (17.47)	74.90* (40.19)
Distributed Lags: Sum of coefficients						
Results from Table 4 (narrow control group)						
Use patient RE and state FE						
Broad control group (main specification)						