Institute for Policy Research
Northwestern University
Working Paper Series

WP-15-04

Does Medical Malpractice Law Improve Healthcare Quality?

Michael Frakes
Associate Professor of Law
Faculty Fellow, Institute for Policy Research
Northwestern University

Anupam Jena
Harvard Medical School, Massachusetts General Hospital, and NBER

Version: March 2015

DRAFT
Please do not quote or distribute without permission.

2040 Sheridan Rd. ◆ Evanston, IL 60208-4100 ◆ Tel: 847-491-3395 Fax: 847-491-9916
www.northwestern.edu/ipr, ◆ ipr@northwestern.edu

Abstract

The researchers estimate the relationship between medical liability forces and healthcare quality using clinically validated measures of treatment quality rather than the coarse measures previously relied upon by scholars. Drawing upon remedy-focused tort reforms—e.g., damage caps—they estimate that current liability rules only modestly improve quality. They contend that this limited independent impact may reflect the structural nature of the present system of rules, which hold physicians to standards determined according to industry customs. They find evidence suggesting, however, that physician practices might respond more significantly to substantive reforms, which alter the standards against which physicians are judged.

The imposition of liability under tort law is sometimes said to serve a purely private function—to correct the injustice created by a wrongdoer and/or to provide compensation to those harmed by that wrongdoer. Tort law is also often viewed, especially by economists, as serving a broader public function—to deter potential wrongdoers from committing costly and harmful errors in the first place.[1] Despite the prominent role that deterrence plays in the theoretical justification for tort law, very little evidence has been put forth to date suggesting that it fulfills this promise.[2] With the substantial direct costs that society expends in administering a system of tort liability,[3] it is crucial for scholars to continue to challenge the empirical foundations of this regulatory premise behind the law. In this paper, we address this challenge in the specific context of medical malpractice law, exploring whether medical liability forces induce (or at least have the potential to induce) the provision of higher quality care.

Surprisingly, medical malpractice scholarship has deemphasized considerations of medical quality and medical errors and has paid significantly more attention to the relationship between malpractice pressure and measures of treatment utilization and health care costs, i.e. "defensive medicine." Although studies of defensive medicine generally demonstrate that greater health care spending, prompted by greater malpractice pressure, is not associated with improvements in broad quality measures such as overall mortality rates of the affected population, it is nearly universally recognized in the medical literature that such broad-based mortality

---

[1] Of course, those viewing tort law through the lens of its deterrence rationale do not contend that tort law is the sole institution in place that is designed to encourage safer practices.

[2] Dewees and Trebilcock (1992) and Schwartz (1994) provide a review of early investigations into deterrence in a range of tortious contexts: automobile accidents, products liability, workers compensation and medical malpractice. Mello and Brennan (2002) likewise review the limited evidence of deterrence put forth to date in the medical malpractice context. Polinsky and Shavell (2010) provide a recent review of studies on deterrence in products liability contexts, generally finding no convincing evidence of improvements in the level of care arising from liability fears (as opposed to market forces that otherwise compel the provision of sound care).

[3] In the context of medical malpractice jurisprudence alone, the United States spends over $7 billion annually in litigation expenses for defendants and plaintiffs and for the overhead incurred by medical liability insurers (Mello et al, 2010). Surveying the relevant literature on the costs of the tort system, Polinsky and Shavell (2010) claim that for every dollar received by plaintiffs from defendants in tort cases, roughly a dollar is spent in legal or administrative expenses.

measures themselves are poor surrogates for more direct measures of health care treatment quality. In testing for the deterrent impact of the law, it is beneficial to employ measures that are more reflective of the care actually received in clinical encounters rather than of background environmental factors that likewise shape health outcomes.[4] In this paper, we attempt to achieve this separation using direct, clinically validated measures of health care quality.

In particular, using data from the National Hospital Discharge Surveys from 1977 to 2005 and the Behavioral Risk Factor Surveillance System from 1987 to 2008, we analyze the effect of medical malpractice liability on several comprehensive inpatient and outpatient health care quality metrics including: (1) risk-adjusted mortality rates during inpatient stays for selected events (e.g., acute myocardial infarction), which have been argued to specifically reflect the quality of inpatient care, (2) avoidable hospitalization rates and cancer screening rates, which reflect outpatient care quality, and (3) adverse-event rates to mothers during childbirth, which reflect an alternative, patient-safety-focused indicator of inpatient quality.[5] Collectively, these indicators account for four of the five domains of quality targeted by the OECD's Health Care Quality Indicator's project and for each of the three domains of quality promulgated by the Agency for Health Care Research and Quality (AHRQ).[6]

---

[4] Lakdawalla and Seabury (2009), for instance, find that higher county-level malpractice pressure leads to a modest decline in county-level mortality rates. Other studies have focused on slightly more targeted populations. For instance, numerous studies have estimated the impact of malpractice reforms on infant mortality rates (or infant Apgar scores, which are seen as predictive of infant mortality), generally finding no relationship (Klick and Stratmann 2007; Frakes 2012; Currie and MacLeod 2008; and Dubay, Kaestner, and Waidmann 1999). Similarly, Kessler and McClellan (1996) estimate a trivial relationship between liability reforms and survival rates during the one year period following treatment for a serious cardiac event (e.g., acute myocardial infarction). Sloan and Shadle (2009) undertake a similar analysis. In each of these cases, however, the mortality measures of interest are broad enough in temporal or clinical scope that they may be driven by many factors other than the care actually delivered at particular outpatient and inpatient encounters (McClellan and Staiger 1999). These factors may include health care access, socio-economic status, individual risky behaviors, living conditions, social support networks, etc. The possibility of a multitude of determinants of one's health status raises statistical concerns over the ability to reliably identify the link between variations in the malpractice environment and the indicated health outcomes.

[5] We note, of course, that particular quality measures may implicate greater liability fears among physicians than others, given the frequency with which lawsuits arise in the associated medical contexts.

[6] A key advantage of focusing the malpractice inquiry on those quality measures promulgated by such institutions as the AHRQ is that such measures are, by their very design, better reflective of the influence of the delivered health care itself. Few malpractice studies have investigated the link between malpractice law and metrics of this nature. Two exceptions are

Physicians are only liable under tort law when their harmful actions fail to comply with an operable standard of care. Empirical investigations into the effects of the malpractice system on physician behavior often ignore this fundamental fact. They speak about liability "pressure," without necessarily asking: "pressure to do what?" In evaluating the potential for malpractice law to improve physician practice quality, we move beyond this abstract treatment of the law and center our analysis on a consideration of the particular expectations placed upon physicians by the liability system.

Primarily, we ask how the quality of physician care is affected when the clinical standards against which they are judged in court are themselves altered. We explore this question by drawing upon the one significant liability reform of this substantive variety that the majority of states have implemented over time: the abandonment of rules requiring that physicians be judged according to the customary practices of local physicians and the contemporaneous adoption of rules requiring that physicians be judged according to national (or non-geographically-limited) standards of care (Frakes 2013). In light of the rampant regional disparities in care that have persisted across regions for decades and that have been the subject of a massive literature in health economics and medicine (see, for example, Chandra and Staiger 2007), one can view the move from a local to a national-standard rule as a meaningful and large alteration of the standards clinically expected of physicians (Frakes 2013).

---

perhaps provided by Iizuka (2013) and Currie and MacLeod (2008). The latter study finds that damage cap adoptions increase preventable complications of labor and delivery, suggesting that higher liability pressure improves patient safety. Iizuka (2013) finds that certain tort reforms—e.g., collateral source rule reforms and punitive damage caps—increase labor and delivery-related complications. Interestingly, Iizuka finds no such relationship with non-economic damage-caps, despite the fact that such caps arguably amount to the most significant reduction in liability pressure out of the four traditional reforms that he explores (Paik et al, 2013). While obstetrics has formed the canonical example of research in empirical malpractice, obstetricians themselves account for less than 3 percent of U.S. physicians. The health care quality processes that we study form the 'bread-and-butter' practices of generalist physicians which form the largest group of practicing physicians. Moreover, we analyze the quality of health care provided in outpatient settings, a setting which accounts for over 20% of the nation's total health care dollars (CMS 2011) and has received no special attention by the malpractice deterrence literature. One exception is perhaps Baicker and Chandra (2005), which documents notable sensitivity in mammography screening to changes in malpractice premiums.

For each measure of treatment quality explored, we find that when states modify their standard-of-care rules to expect physicians to provide higher levels of quality—e.g., when initially low quality regions adopt national standard rules—observed levels of quality increase substantially in the direction of such new expectations. However, when states modify their rules so as to condone the provision of lower quality care, physicians do not respond by reducing the quality of their practices. Collectively, these findings suggest that that medical liability forces—under the right structural framework—hold the potential to elevate the quality floor.

Focusing on liability-standard changes holds two key advantages over the traditional approach taken by the literature to tease out the influences of medical liability on physician practices. This typical route estimates the impacts of reforms that do not affect the standards against which physicians are judged, but rather reduce the expected damages awarded when a physician is found liable—e.g., the adoption of caps on awards for non-economic damages (pain and suffering damages). If one views liability-standard reforms as akin to reducing driving speeds by lowering the speed limit, adopting / removing damages caps is akin to changing driving behaviors by altering the associated speeding fines. One important limitation of this traditional, remedy-focused approach, however—and thus one key motivation for the primary liability-standards approach of this paper—is the mixed set of findings to date bearing on the necessary first-stage to this damage-cap inquiry. That is, scholars have not consistently found a link between damage cap adoptions and the frequency of malpractice suits. Ultimately, it is not clear that damage caps have systematically represented a sizeable enough change in liability pressure (or in the perception of liability pressure) from which to tease out the law's influence on behavior.

A second limitation of using damage-cap adoptions to illuminate the law's influence on health care quality stems from the fact that caps were generally adopted in the face of a system that determined liability standards by deferring to customary industry practices. One might predict that damage cap adoptions would not have substantial impacts on *average* prevailing practice patterns when malpractice standards are simply designed to reign in errant practitioners and to reinforce those practices that the industry has already decided it wishes to perform in the first place. We find evidence consistent with this prediction. For each measure of treatment quality, the estimated effect of malpractice pressure within our current system, as identified by the adoption of non-economic damage caps, is both statistically insignificant and small in magnitude, with a 95% confidence interval that is relatively tightly bound around zero.

The danger of relying solely upon damage-cap adoptions to understand liability's influence is that one might infer from such modest impacts of cap adoptions that physicians are only weakly sensitive to the parameters of the liability system. Any such inference would be misguided, however, insofar as the modest effects of caps might simply arise from the particular structure of the industry-custom-focused system in which they were enacted. These findings do not suggest that physicians would fail to respond to a more substantive alteration of that system altogether, as some modern reform proposals aim to do—e.g., retreating from a reliance on industry custom and instead setting liability standards with reference to evidence-based clinical practice guidelines. By demonstrating that physicians respond to a meaningful shift in liability standards, our analysis provides insight into the possible impacts of next generation liability reforms of this nature and broadly demonstrates the potential for tort liability to improve treatment quality.

The paper proceeds as follows. In Section I, we provide a background on liability-standard rules and discuss their potential impacts on physician practices. Section II discusses the data and empirical methodology. Section III presents the results of the empirical deterrence analysis. Finally, Section IV concludes.

I. MALPRACTICE LAW AND PHYSICIAN BEHAVIOR

A. The effect of substantive malpractice reforms

The clinical decisions that physicians make, including decisions to perform particular treatments and to deliver certain levels of quality, are likely to be shaped by a number of forces. One such force may stem from fear over liability for harming a patient through actions that fail to comply with the operable standard of care expected of physicians by law. For instance, if tort law expects that physicians perform surgical technique *X* when administering a particular treatment, physicians may feel pressured to adhere to technique *X* despite an inclination to otherwise deviate. Accordingly, depending on how physicians weigh liability considerations against other determinants of clinical practices, the imposition of a liability system may compel at least some physicians to follow the legally expected standards inherent in that system. It stands to reason that if the liability system is reformed so as to now expect that physicians follow a different set of practices—e.g., to perform technique *Y* in the hypothesized context—a new equilibrium will be reached in which at least some subset of physicians adjust their practices in the direction of *Y*. In order to gauge the sensitivity of physician practices, and of the quality delivered, to the standards expected of physicians under the law, we explore the impacts of a liability reform of just this nature.

To understand the quasi-experimental framework that we construct, it is first important to note that liability systems in the United States have generally deferred to the customary practices of the industry itself in order determine the standards it imposes. Essentially, parties in U.S. malpractice suits call upon expert witnesses to testify as to customary physician practices in order to determine the benchmarks to which the defendant physicians are held. Where the law has varied, however, is with respect to which physicians it looks in order to assess customary practices. Historically, state malpractice laws judged physicians against customary practices of physicians working in the same locality, essentially expecting physicians to follow the practices applied by those around them. Deviations in care from these customary standards that led to adverse medical events were judged as negligent. Over the latter half of the Twentieth Century, however, the majority of states amended their substantive malpractice laws to abandon locality rules in favor of rules requiring physicians to follow national standards of care, thereby geographically harmonizing clinical expectations under the law.[7]

These distinctions in liability rules are of particular relevance in light of the substantial geographical variations that have persisted in clinical practices over time, a phenomenon that has been the subject of a large literature in medicine and health economics (Chandra and Staiger, 2007; Skinner 2011). Different theories have been set forth as to why clinical practices have developed along distinct regional pathways. One reason could be that physicians operate off of a geographically-limited set of information (Wennberg and Gittelsohn 1973). A more nuanced story of productivity spillovers could also be present such that local conditions in one region lead to specialization in one practice style while local conditions in other regions call for specialization in a different type of practice (Chandra and Staiger 2007). Regardless of the precise mechanism, locality rules may cement these regional pathways, either by discouraging physicians from deviating from local customs or by providing comfort to physicians wishing to maintain such customs. The adoption of national standard rules may break these forces and lead local physicians to follow the vastly distinct practices followed elsewhere in the nation. The fundamental hypothesis that we test in this paper is

[7] Online Appendix B provides further details on the evolution of malpractice-standard rules. Roughly 16 stated abandoned the use of local standards in favor of national standards over the sample period. As discussed in Frakes (2013) earlier reforms (i.e., in the mid-20th Century) to the initial locality rules focused on making it easier for outside experts to testify as to local standards—e.g., allowing a New York physician to testify as to local South Carolina standards. The substantive reforms of interest in this study pick up after these earlier procedural changes and focus instead on changes in the expected practices—e.g., requiring that South Carolina physicians now follow the practices applied elsewhere.

whether prevailing rates of health care quality shift in the direction of such altered expectations.[8]

Importantly, this analysis affords us the opportunity to separately test how physicians respond to changes in malpractice standards which in some instances expect higher levels of quality and in other instances lower levels. For each of the quality metrics explored, a number of treatment states began the sample period with high quality levels while a number of others began with low quality levels, in which event the move towards a national standard represented a change in legal expectations in both directions depending on the pre-reform level of quality. In light of the possibility that physicians may respond differently to an elevation of what is expected of them relative to a slackening of what is expected of them, we test for asymmetrical responses to the adoption of national-standard rules.

B. The effect of remedy-focused liability reforms

A key conceptual point of this paper is that a medical liability system is largely characterized—at least on a substantive level—by an underlying set of standards imposed on physicians. As such, exploring the impacts of a liability system on physician behavior, as the malpractice literature endeavors to do, requires understanding how physicians respond to the standards imposed by that system. In the previous subsection, we set forth one way to approach this analysis, asking what happens when those clinical standards themselves are modified. One may shed further light on this general inquiry by evaluating the marginal impact of a *given* set of liability standards, as distinct from a changing set of standards. In other words, for a particular set of standards, what happens to practice patterns when we diminish the influence of such standards on the margin?

---

[8] Frakes (2013) similarly explores the impact of national-standard rule adoptions on physician practices, though focusing on treatment and diagnostic utilization patterns (in obstetric and cardiac contexts) rather than health care quality metrics.

A straightforward way to confront this alternative analysis is to explore the impacts of those penalty-reducing reforms most traditionally emphasized by the malpractice literature—e.g., dollar caps on non-economic damage awards.[9] Such reforms diminish the penalties associated with liability without altering the basis for liability.[10] A reduction in observed levels of quality upon the adoption of a damage cap might suggest that the present liability structure is otherwise encouraging higher levels of care—i.e., is deterring poor medical practices.

The validity of this latter approach of course rests on the assumption that damage caps do indeed meaningfully reduce the expected harm to be imposed upon physicians. At first blush, one may doubt this considering that physicians generally face limited immediate financial risk from associated damage awards insofar as they are insured against such losses with coverage that is typically not experience rated (Sloan 1990, Zeiler et al. 2007). However, despite this limited immediate financial risk, physicians may face a number of uninsurable costs as a result of malpractice liability – e.g., reputational and psychological damage (Jena et al. 2011).[11] Damage caps may therefore reduce the extent of such uninsurable harms to the extent that caps decrease the probability of suit—e.g., by leaving plaintiffs and/or plaintiffs' attorneys less inclined to file suit (Shepherd 2014). The estimated impact of caps on claims frequency / likelihoods has varied significantly across

[9] The seminal example of this approach is provided by Kessler and McClellan (1996). Non-economic damages generally represent over half of the typical malpractice award (Hyman et al. 2009). Furthermore, caps on such damages represent the tort-reform measure that has been most commonly associated with an observed change in certain malpractice outcomes: claims severity, physician supply and malpractice premiums. See Mello and Kachalia (2011) for a comprehensive review of relevant studies. Paik et al. (2013) provides a recent example. Nearly thirty states currently have non-economic damage cap provisions in place, most of which were adopted during the malpractice crisis of the 1980's. In most specifications, we also explore the association between observed health care quality and certain additional types of tort reforms, including reforms of the collateral source rule, caps on punitive-damages awards and other "indirect" reforms. Further descriptions on all reforms are provided in Online Appendix A.

[10] Not all studies that explore the marginal impacts of liability pressure do so by evaluating the impacts of damage caps and related reforms. Some evaluate the marginal influence of liability forces in the present liability system by drawing on variations in malpractice premiums or liability payments per physician. See, for example, Baicker et al. (2007).

[11] Subject to certain exceptions, payments made on behalf of physicians to settle claims or to satisfy judgments must, under federal law, be registered in the National Practitioner Data Bank (NPDB), an electronic repository which is made available to hospitals and certain other health care entities. The NPDB was established by the Health Care Quality Improvement Act of 1986, as amended (42 U.S.C. 11101 et seq.). This repository may reinforce any reputational consequences of malpractice liability.

studies (Mello and Kachalia 2011), casting some doubt on the validity of using damage caps to assess the influence of the law on practices and thereby bolstering the merits of exploiting liability-standard reforms in the alternative. Nonetheless, reinvigorating the promise of a damages-cap-based approach, a recent study by Paik et al. (2013) draws upon significant variation in state laws to find a roughly 29 percent drop in claims frequency upon a damage-cap adoption. This uncertainty in the first-stage evidence surrounding the damage-cap approach to a deterrence analysis lends support to the alternative approach embraced below that focuses instead on meaningful variations in standard-of-care formulations.

Moreover, these first-stage concerns aside, one may doubt that damage-cap adoptions will substantially influence prevailing quality levels in light of the fact that caps have generally been implemented in the United States in the face of a liability system that sets standards according to customary industry practices. In particular, if one merely diminishes the penalty facing physicians under a customary-standards system, physicians may on average simply continue to adhere to the customary industry practices that are the bedrock of those standards in the first place. After all, under a liability structure of this nature, the law exerts little independent force of its own to redirect practice patterns (assuming, of course, that the law is not altering the set of physicians it is looking to in order to assess custom, as discussed above). Systematically, under such rules, physicians may only alter their practices in response to liability fears due to uncertainty in their beliefs as to how courts will assess customary practices—i.e., they may aim to deliver higher quality than otherwise customarily desired over fear that courts will misjudge customary practices to entail such higher practices. Any such channel of influence may not be expected to induce substantially higher levels of quality insofar as it largely entails guesswork on the part of physicians, rather than the delivery of clear signals as to how to improve quality. Damage caps may therefore only induce changes in physician practices to the extent that they reduce the cost of uncertainty to physicians about whether their practice patterns deviate negatively from customary market practices. Caps otherwise do not alter the clinical expectations being placed upon physicians. In the empirical analysis below, we test the hypothesis that damage cap adoptions have modest impacts on levels of treatment quality chosen by physicians.

II. DATA AND METHODOLOGY

A. Data and Quality Measures

We employ two data sources which allow us to study clinically validated measures of the quality of care provided by physicians in both inpatient and outpatient clinical settings. First, we construct several quality measures using the 1977 to 2005 National Hospital Discharge Surveys (NHDS). The NHDS data, supplemented with geographic identifier codes, provides rich inpatient discharge records over a long enough period of time to allow us to draw on an extensive set of within-state variations in both liability-standard rules and damage-cap provisions. Second, we use data from the 1987 to 2008 Behavioral Risk Factor Surveillance System (BRFSS) to capture various rates of cancer screening.[12]

Foremost among those organizations promulgating health care quality indicators is the Agency for Healthcare Research and Quality (AHRQ). AHRQ measures are particularly useful for the present study insofar as they are designed for use with administrative inpatient databases such as the NHDS. In constructing quality indicators, we build off of the three domains of quality indicators developed by the AHRQ, supplementing the AHRQ measures with those capturing cancer-screening practices. We provide a brief overview of each metric below, with additional

[12] Not all screening measures are available over this entire time period, however. While longer time periods are available for some measures—e.g., mammograms—others are only available over the 2000s—e.g., PSA testing.

details regarding the construction of the resulting quality metrics provided in Online Appendix A.

Inpatient mortality for selected conditions—Following the AHRQ's Inpatient Quality Indicators, we first construct a composite inpatient mortality rate for selected acute medical conditions using NHDS data. Unlike mortality rates computed over an entire area affected by a relevant legal regime—which may be influenced by a number of socioeconomic factors—IQI-inspired rates are designed to capture mortality events associated with a clinical encounter itself. To rule out selection concerns—i.e., concerns regarding the liability regime impacting the probability of patients appearing in the inpatient environment in the first place—this measure focuses on mortality among a subsample of discharges in which the primary diagnosis code indicates select medical conditions (e.g., acute myocardial infarction, stroke, etc.) that uniformly entail hospitalization upon their occurrence.

Avoidable hospitalizations as a measure of outpatient quality—Second, we use the NHDS to capture the rate of avoidable hospitalizations (AH) within each state-year cell, a measure that is inspired by the AHRQ's Prevention Quality Indicators. Though constructed using inpatient data, AH rates are thought to reflect the quality of care prevailing in the associated outpatient community. Such measures identify conditions—e.g., asthma—with respect to which proper outpatient care would have prevented (or at least reduced the likelihood of) hospitalization. To alleviate any concern that such metrics are confounded by physician choice as to whether to hospitalize patients with the indicated conditions (which would threaten the ability to infer outpatient quality), we construct an AH rate that focuses only on a subset of AHs with little physician discretion over the decision to hospitalize—e.g., ruptured appendix.[13]

[13] Nonetheless, in the Online Appendix, we demonstrate that the low-discretionary AH-rate findings mirror those based on AH-rate constructions that are not limited by the degree of discretion in the admission decision.

Maternal Trauma and Complications—The AHRQ's Patient-Safety Indicators (PSIs) capture complications and adverse events that take place in inpatient settings following surgeries, procedures and deliveries. Using NHDS data, we focus our analysis of PSIs on those related to delivery / childbirth. Many PSIs reflect the quality of care provided during surgeries, rates of which may be a function of the liability environment, implicating issues of selection. Rates of childbirth, on the other hand, are unlikely to be impacted by malpractice pressures. For the sake of simplicity and to maximize the sample size for this analysis, we group together cesarean trauma events with vaginal delivery trauma events (with and without instruments) and thus construct a composite obstetric trauma indicator. To look at a broader, but related set of obstetric-related complications, we follow Currie and MacLeod (2008) and also consider the incidence of preventable delivery complications—e.g., fetal distress, excessive bleeding, precipitous labor, prolonged labor, dysfunctional labor, etc.

Cancer Screening—To round out our assessment of outpatient care quality, we use patient self-reports from the Behavioral Risk Factor Surveillance System from 1987 to 2008 to compute incidences of mammography, physical breast exam, Prostate-Specific Antigen (PSA) testing, digital rectal exam, pap smear, and sigmoidoscopy / colonoscopy, used to screen for breast, prostate, cervical, and colon cancer, respectively. As explained in greater detail in Online Appendix A, we use national cancer screening guidelines to select the relevant age groups for the analysis and the window period of relevance for the exam—e.g., mammography within the previous two years for females. Cancer-screening has received little attention by the empirical malpractice literature to date, with the exception of Baicker and Chandra (2005), which observed sensitivity in mammography utilization to malpractice premium changes.

Notes regarding quality metrics—Of course, particular quality measures may induce greater liability fears among physicians than others, given the frequency with which lawsuits arise in the associated medical contexts. For instance, cancer screening rates may be especially good measures to study the link between malpractice and health care quality given that missed cancer diagnoses constitute a frequent basis for malpractice lawsuits (Schiff et al. 2013). Our goal, however, was not to pre-select quality metrics based on some priors as to which quality domains are more implicated in liability settings. Rather, we have attempted to paint a more general picture of the link between liability forces and quality and have thus endeavored to collect as many quality metrics as possible, bearing in mind data limitations and sample selection concerns inherent in some metrics—e.g., patient safety incidents occurring during surgeries. We note, in addition, the possibility of spillovers in effects. That is, one might observe a link between liability forces and avoidable hospitalization rates even if few lawsuits are based on such events to the extent that AH rates are proxying for more general changes in outpatient practices and to the extent that liability-induced changes in outpatient behavior in other contexts—e.g., cancer screening—spill over into the management of chronic care conditions underlying avoidable hospitalizations.

Descriptive statistics—On average, each NHDS state-year cell contains roughly 424 discharges associated with the selected conditions used in the composite inpatient mortality rate measure, our first quality indicator. The average inpatient mortality rate among this sub-sample is 8 percent, as presented in Table 1. Likewise, each state-year cell contains an average of roughly 600 low-discretionary avoidable hospitalizations. As explained in Online Appendix A, we form AH rates by normalizing AH counts by an index of hospitalizations for certain medical conditions—e.g., acute myocardial infarction, stroke, etc.—with respect to which there is virtually no discretion over whether or not to admit the patient. This denominator captures the size of the relevant state-year cell without itself being sensitive to legal or financial incentives. With this normalization, the average AH rate across state-year cells equals 1.0. Furthermore, each state-year cell in the NHDS contains on average roughly 600 deliveries. Within this delivery subsample, maternal trauma (third or fourth degree lacerations) occurs nearly 4 percent of the time and preventable complications occur nearly 16 percent of the time. Finally, cancer screening rates among the relevant BRFSS participants ranges, on average, from 40 to 73 percent.

To describe the variation in quality of care across regions, Column 2 of Table 1 provides, for each quality indicator, a measure of the average gap over the sample period between the mean state level and the associated mean national level. More specifically, following Frakes (2013), we summarize this gap by calculating the mean absolute deviation between the state and national indicator levels (for each year) and normalizing this rate by the national level. For instance, on average over the sample period, the mean maternal trauma rate within a state differed from the national mean trauma rate by an amount equal to roughly 26 percent of the national level. Because this measure is computed over the entire sample period, it somewhat understates the regional disparity measure that is most relevant to our analysis. In particular, in early years of the sample and among states which began the sample under a locality-rule regime, the average gap between the state and the national rate, for each of the listed indicators, is substantially larger than the figures provided in Table 1. For instance, in the pre-1982 period for those locality-rule jurisdictions with below-average levels of maternal trauma rates, the mean gap between the state and national rate of maternal trauma was nearly 68 percent of the national level. In the empirical analysis below, we explore whether these gaps are narrowed through the adoption of national-standard rules (approaching the inquiry separately from each side of the regional quality distribution).

Table 1. Descriptive Statistics

	Mean (Standard Deviation)	*Percentage Absolute Deviation between State and National Mean*
Panel A: Quality Measures (NHDS)		
Composite Inpatient Mortality Rate	0.08 (0.03)	0.16 (0.15)
Low-Discretionary Avoidable Hospitalization Rate (Scaled by Low-Variation Health Index)	1.00 (0.23)	0.15 (0.15)
Maternal Trauma Rate	0.04 (0.02)	0.26 (0.25)
Maternal Preventable Complications Rate	0.16 (0.06)	0.20 (0.20)
Panel B: Cancer-Screening Rates (BRFSS)		
Mammogram (within last year, female age 40-75)	0.73 (0.45)	-
Physical breast exam (within last year, female age 40-75)	0.64 (0.48)	-
Proctoscopic exam (sigmoidoscopy or colonoscopy within last 5 years, age 50-75)	0.40 (0.49)	-
PSA Testing (within last year, age 50-75)	0.53 (0.50)	-
Digital Rectal Exam for Prostate Cancer (within last year, age 50-75)	0.50 (0.50)	-
Pap smear (within last year, age 21+)	0.60 (0.49)	-

Notes: Standard deviations are in parentheses. Quality measures in Panel A are from a sample of 1190 state-year cells from the 1977 – 2005 NHDS files. Quality statistics in Panel A are weighted by the relevant denominator used in the state-year quality rate (e.g., the state-year delivery count or the state-year low-variation health index).
Source: Panel A: National Hospital Discharge Survey (1977-2005), Panel B: Behavioral Risk Factor Surveillance System (1987-2008).

B. Specifications

To explore whether the quality of health care provided by physicians is affected by the clinical malpractice standards expected of physicians under the law, we estimate the degree to which state mean rates for the relevant quality measures converge towards their respective national mean rates as states adopt national-standard rules. In this investigation, however, we allow for a differential convergent response from the top and the bottom of the regional quality distribution—that is, we allow for a different response when the law changes so as to expect a higher level of quality compared to when the law changes so as to condone a lower level of quality. Following Frakes (2013), we estimate:

$$(1) \qquad Log(Q_{s,t}) = \alpha + \boldsymbol{\gamma_s} + \boldsymbol{\lambda_t} + \boldsymbol{\varphi_s}\mathrm{t} + \boldsymbol{\beta_1 X_{s,t}} + \boldsymbol{\beta_2 Z_{s,t}} + \boldsymbol{\beta_3 O_{s,t}} +$$

$$\beta_4 HIGH\_QUALITY_s + \beta_5 NS_{s,t} + \beta_6 HIGH\_QUALITY_s * NS_{s,t} + \varepsilon_{s,t}$$

where *s* indexes state and *t* indexes year. State fixed effects, $\boldsymbol{\gamma_s}$, and year fixed effects, $\boldsymbol{\lambda_t}$, control for fixed differences across states and across years, respectively. $\mathbf{X_{s,t}}$ represents certain demographic characteristics of the patient population, along with certain mean characteristics of the represented hospitals. $\mathbf{Z_{s,t}}$ represents certain other state-year characteristics (HMO penetration rate, physician concentration rate, and median household income). $\mathbf{O_{s,t}}$ is a matrix representing a set of indicator variables for the incidence of additional tort reforms. In some specifications, we include state-specific linear time trends, $\boldsymbol{\varphi_s}$t, to control for slowly-moving correlations between the relevant quality measures in a state and the adoption of tort reforms by that state. $Q_{s,t}$ represents the relevant healthcare quality measure – e.g., the composite inpatient mortality rate or the avoidable hospitalization rate. For each of the relevant quality indicators, Online Appendix A provides additional details regarding the compositions of **X** and **Z**.

$NS_{s,t}$ represents an indicator for a national-standard law. *HIGH_QUALITY* is an indicator for a state that began the sample period with an initial rate below the national mean for the relevant quality indicator. Again, for all indicators other than cancer screening rates, high levels of the various indicators represent lower levels of quality (and vice versa). The coefficient of β_5 in this interaction specification can effectively be interpreted as the association between national-standard laws and quality indicator levels for states with initially low levels of quality (i.e., with initially above-average indicator levels). For states with initially higher than average levels of quality, this same association is captured by the sum of β_5 and β_6.

Next, we estimate the association between damage caps and the quality of care provided by physicians through the following specification:

$$(2) \qquad Log(Q_{s,t}) = \alpha + \boldsymbol{\gamma}_{\mathbf{s}} + \boldsymbol{\lambda}_{\mathbf{t}} + \boldsymbol{\varphi}_{\mathbf{s}} t + \beta_1 CAP_{s,t} + \boldsymbol{\beta}_{\mathbf{2}} \mathbf{X}_{\mathbf{s,t}} + \boldsymbol{\beta}_{\mathbf{3}} \mathbf{Z}_{\mathbf{s,t}} + \boldsymbol{\beta}_{\mathbf{4}} \mathbf{O}_{\mathbf{s,t}} + \varepsilon_{s,t}$$

where $CAP_{s,t}$ represents an indicator variable for the presence of a cap on non-economic damages in state s and year t. The coefficient of interest in each specification is captured by β_1.

III. RESULTS

A. Liability-Standards Analysis

AHRQ-Inspired Measures and Preventable Delivery Complications—In our first approach to exploring the link between malpractice forces and health care quality, we estimate the interaction specification indicated by equation (1) above and explore whether health care quality is influenced by reforms that directly alter the clinical standards of care expected of physicians, effectively separating the inquiry by whether the reform expects physicians to follow higher or lower standards of care. The results of this exercise are presented in Table 2.[14] Each of the 4 panels represents results for different AHRQ quality indicators.

[14] Reported standard errors in Table 2 and in all subsequent tables are clustered at the state level to allow for arbitrary within-state correlations of the error structure.

Table 2. The Relationship between National-Standard Laws and Various Health Care Quality Metrics

	(1)	(2)	(3)
Panel A. Dependent Variable: Inpatient Mortality Rate for Selected Conditions (Logged)			
National-Standard (NS) Law Dummy	-0.076**	-0.073*	-0.100
	(0.035)	(0.039)	(0.084)
NS Law * Initially High Quality State	0.128*	0.156**	0.200
	(0.073)	(0.064)	(0.128)
N	1104	1048	1048
Panel B. Dependent Variable: Low-Discretionary Avoidable Hospitalization Rate (Logged)			
National-Standard (NS) Law Dummy	-0.545***	-0.456***	-0.348***
	(0.075)	(0.067)	(0.061)
NS Law * Initially High Quality State	0.534***	0.451***	0.309***
	(0.126)	(0.112)	(0.082)
N	1139	1082	1082
Panel C. Dependent Variable: Maternal Trauma Rate (Logged)			
National-Standard (NS) Law Dummy	-0.396***	-0.311**	-0.302
	(0.103)	(0.134)	(0.191)
NS Law * Initially High Quality	0.446**	0.228	0.395
	(0.178)	(0.207)	(0.323)
N	1053	1005	1005
Panel D. Dependent Variable: Preventable Delivery Complication Rate (Logged)			
National-Standard (NS) Law Dummy	-0.403***	-0.415***	-0.419***
	(0.086)	(0.090)	(0.108)
NS Law * Initially High Quality	0.513***	0.441***	0.471**
	(0.135)	(0.122)	(0.183)
N	1089	1035	1035
Control Variables?	NO	YES	YES
State-Specific Linear Trends?	NO	NO	YES

Notes: robust standard errors corrected for within-state correlation in the error term are reported in parentheses. All regressions include state and year fixed effects. Regressions in Panel A are weighted by the number of admissions (for the relevant state and year) in the sub-sample of discharges associated with the relevant selected conditions (e.g., acute myocardial infarction). Regressions in Panel B are weighted by the low-variation health index (i.e., the sum of discharges for acute myocardial infarction, stroke, hip fracture or gastrointestinal bleeding) associated with each state-year cell. Regressions in Panels C and D are weighted by the number of deliveries associated with the relevant state-year cell. Inpatient mortality rates are risk-adjusted for the incidence (among the sub-sample) of each of the conditions comprising the sub-sample of selected conditions. Each regression also includes a separate dummy variable indicating whether the state has an initially below-average rate of the relevant quality metric (coefficient omitted).
Source: 1977 – 2005 National Hospital Discharge Surveys.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

The coefficients presented in the first row of each panel can be interpreted as the association between the given quality indicator and the adoption of a national-standard rule in those treatment states that began the sample period with initially low levels of quality (i.e., where the above-average quality variable equals zero), representing those states with initially high levels of the respective quality measure. In the case of inpatient mortality rates for selected medical conditions, the low-discretionary avoidable-hospitalization (AH) rate, the maternal trauma rate and the preventable complication rate, we estimate that the adoption of a national-standard rule in such states is associated with a substantial and statistically significant (across nearly every specification) decrease in the respective indicator measure and thus a substantial *increase* in health care quality. More specifically, in the basic difference-in-difference specifications with only state and year fixed effects, we estimate a 7.6, 54.5, 39.6 and 40.3 percent decrease in the respective quality indicator in connection with national-standard adoptions. With the inclusion of various state-year covariates and state-specific linear time trends, these estimates remain comparable, suggesting a 7.3, 45.6, 30.2, and 41.9 percent decline in the respective indicator.

Considering that a national-standard adoption in such initially-low-quality states entails a shift in clinical expectations in the direction of higher quality, the results from this exercise suggest that liability reforms that affirmatively elevate the standards expected of physicians may indeed succeed in inducing higher quality practices. These findings demonstrate a substantial closing of the gap between low-quality regions and other regions upon the retreat from a local-standard-of-care rule. For instance, considering the initial mean gap between maternal trauma rates in locality-rule states and national maternal trauma rates (equaling 68 percent after normalizing by the national rate, as stated above), the above findings imply that roughly half of this gap is closed upon the move from local to national-standard-of-care rules.

In Figures 1—4, we present results from dynamic versions of the specifications estimated in Table 2, which include a set of lead and lag indicators for the adoption of a national-standard reform (the underlying tabular regression results are presented in Tables B1 and B2 of the Online Appendix). The plotted line reflects a time trend (with confidence bounds) of the differential between treatment and control states in the respective quality indicator, where time is not measured in calendar years but instead with reference to the point of adoption of national-standard rules (the trend is normalized such that the differential equals zero at time zero). For each measure of health care quality, the estimated pattern of lead coefficients for the national-standard indicators do not suggest any increasing trends in the differential quality attainments between treatment and controls states prior to the reforms (i.e., any decreasing trends in the differential indicator levels between such states). Pre-treatment trends of that nature may have undermined the assumption inherent in the difference-in-difference specification that, but for the change in the law, the quality indicators would have trended in the same direction in the treatment and control groups alike. As such, the fact that the differential in quality appears to emerge only upon the adoption of the national standard rules themselves increases our confidence in a causal interpretation of the documented associations and increases our confidence that the primary results are not merely a reflection of mean reversion.

While practices appear to improve upon a shift in clinical standards expecting higher quality, the results do not overwhelmingly suggest a corresponding decline in quality upon a shift in legal standards arguably condoning lower quality care. To assess this reverse question, we explore what happens to initially high quality states (states with initially low quality indicator levels) when they adopt national-standard rules, which, in the case of such states, arguably lower operable standards by expecting that physicians follow the lesser-quality practices applied elsewhere. These results can be obtained from Table 2 by adding the baseline effect in the initially low-quality states (row 1 in each panel) to the coefficient of the interaction term (row 2 in each panel), which captures the marginal alteration of this national-standard effect from instead beginning in an initially high quality state. Across the various indicators, this addition suggests that a national standard adoption in the initially high-quality states is associated with a 5.2, -1.1, 5.0, and an 11 percent change in the respective quality indicator. We do observe a mean decline in quality—that is, an increase in the respective indicator—for three of the four indicators upon this change in standards arguably condoning a lower level of quality. Even in those cases, however, these responses are more modest than the responses indicated above for the initially low-quality states. In Table B5 of the Online Appendix, we allow for a more immediate estimate of the national-standard effect in initially high-quality states by estimating an identical specification to that set forth in equation (1), except replacing the *HIGH-QUALITY* indicator variable in the specification with an analogously specified *LOW_QUALITY* indicator. β_5 in this alternative specification now captures the relationship between national-standard adoptions and the relevant quality rates for initially high-quality regions. As can be observed from Table B5, not only are the coefficients smaller than those capturing the effect of national-standard adoptions in initially low-quality regions, but they are also generally not significantly different from zero, except perhaps with respect to the specifications using inpatient mortality rates for selected conditions as the quality measure of interest. Even in the case of these morality-rate specifications, however, it appears from Figure 5 that the inpatient mortality rate response emerges largely in the period of time prior to the national-standard adoption, suggesting that it may not even be a true response to the law itself.

Figure 1. Liability Standards Analysis, Dynamic Difference-in-Difference Regression Results: Inpatient Mortality Rate for Selected Conditions, Initially Low-Quality States

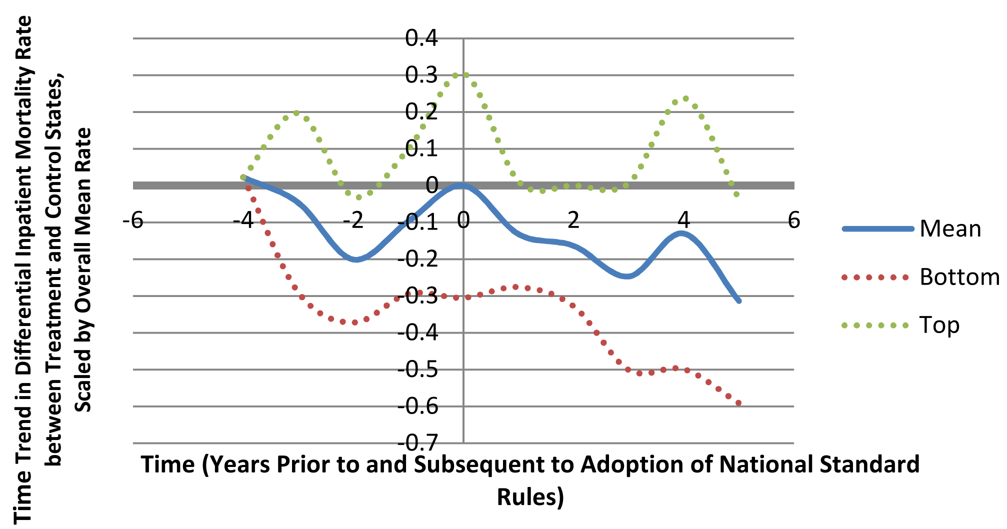


Figure 2. Liability Standards Analysis, Dynamic Difference-in-Difference Regression Results: Low-Discretionary Avoidable Hospitalizations, Initially Low-Quality States

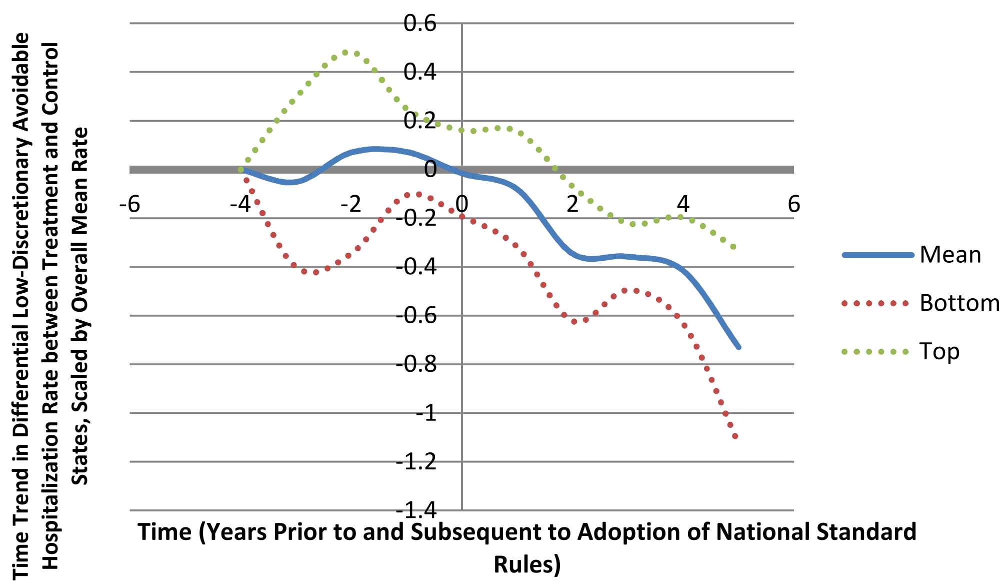


Figure 3. Liability Standards Analysis, Dynamic Difference-in-Difference Regression Results: Maternal Trauma, Initially Low-Quality States

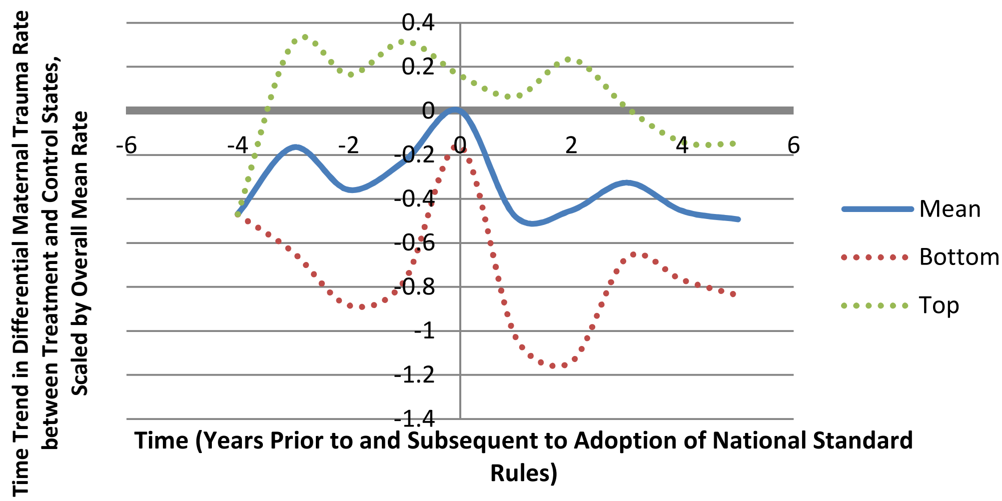


Figure 4. Liability Standards Analysis, Dynamic Difference-in-Difference Regression Results: Preventable Delivery Complications, Initially Low-Quality States

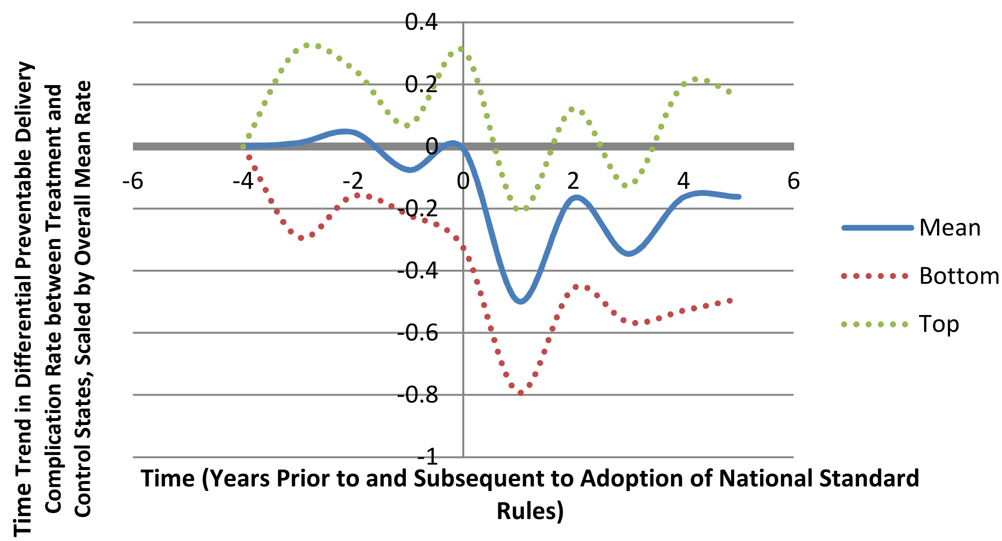


Figure 5. Liability Standards Analysis, Dynamic Difference-in-Difference Regression Results: Inpatient Mortality Rate for Selected Conditions, Initially High-Quality States

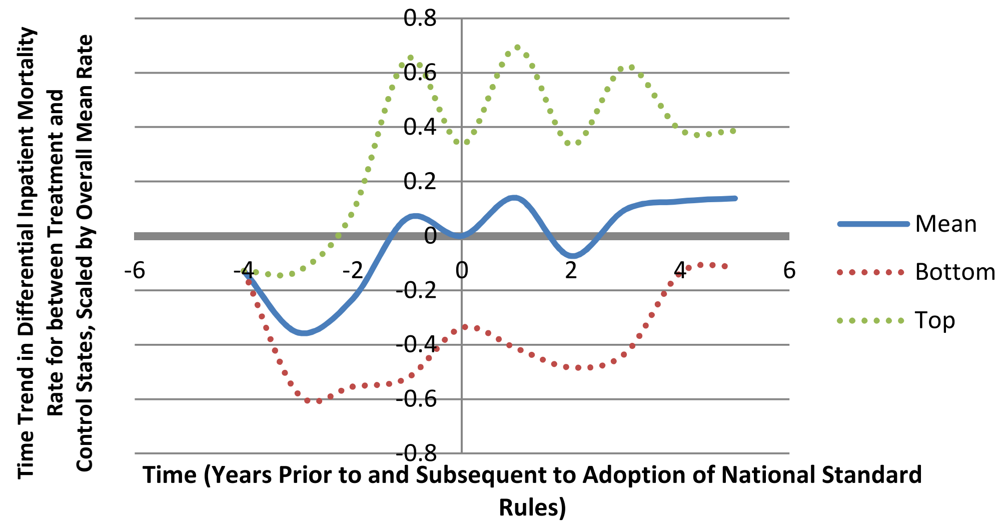


Figure 6. Liability Standards Analysis, Dynamic Difference-in-Difference Regression Results: Low-Discretionary Avoidable Hospitalizations, Initially High-Quality States

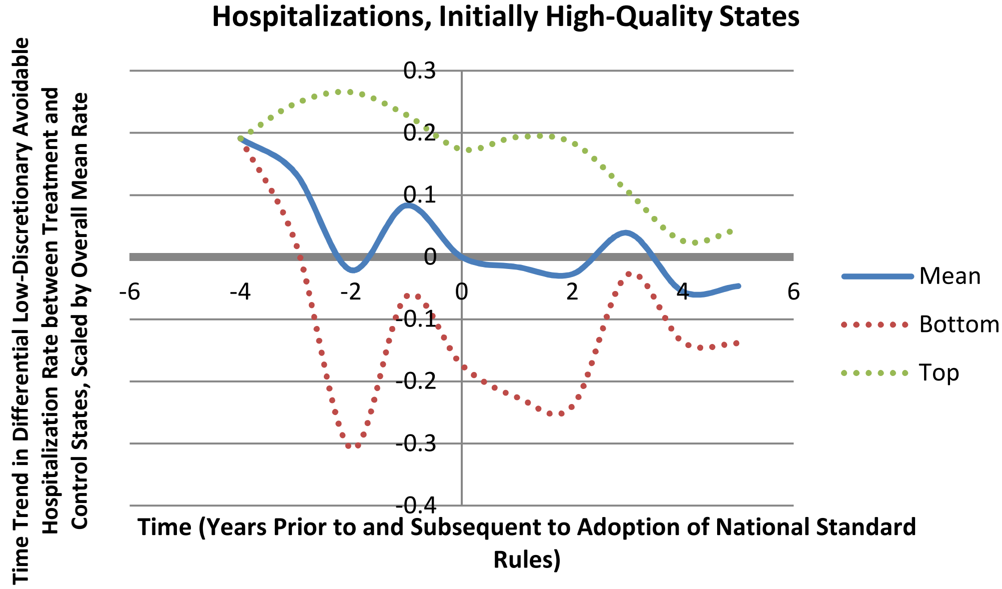


Figure 7. Liability Standards Analysis, Dynamic Difference-in-Difference Regression Results: Maternal Trauma, Initially High-Quality States

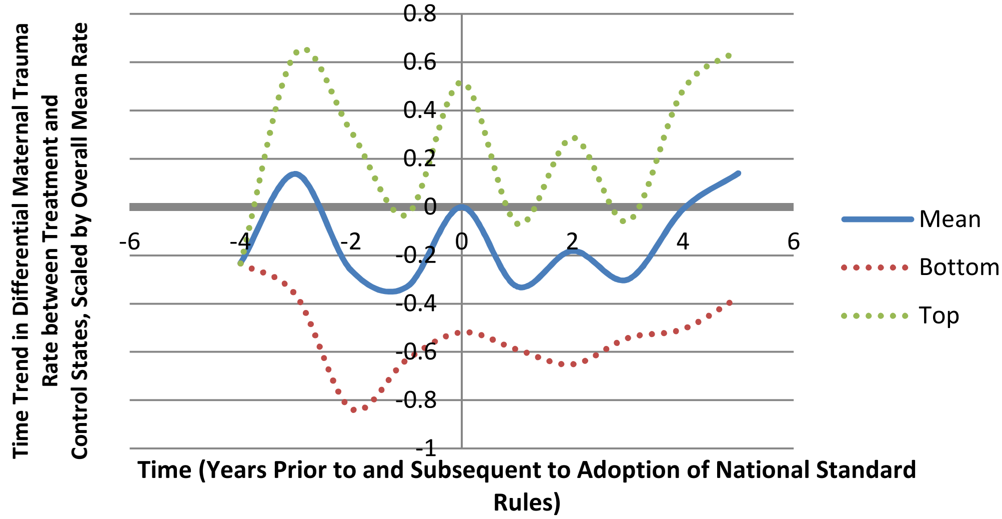
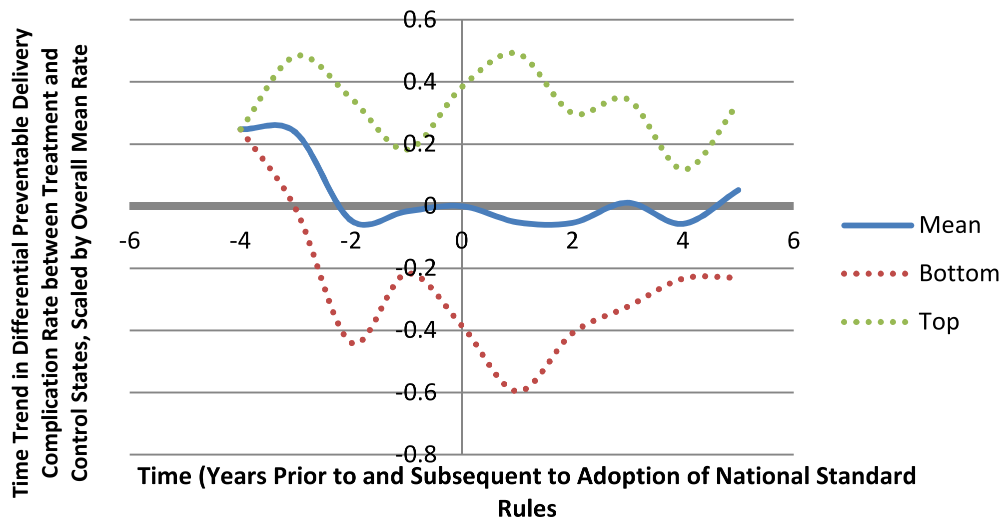


Figure 8. Liability Standards Analysis, Dynamic Difference-in-Difference Regression Results: Preventable Delivery Complications, Initially High-Quality States



Cancer Screening Measures—For this liability standards analysis, our primary tables do not include results for the cancer screening measures. For some of these measures—e.g., PSA testing for prostate cancer—data are only available during the 2000s, affording no ability to draw upon relevant standard-of-care reforms. Likewise, with respect to sigmoidoscopy/colonoscopy screening for colon cancer, data are generally unavailable in the pre-reform years for the relevant treatment states to facilitate a difference-in-difference analysis. However, for the remaining cancer screening measures—e.g., those relating to breast and cervical cancer—data are available during a period of time—i.e., the 1990s—in which Indiana, Delaware and Rhode Island can be utilized as treatment states. Our intent, of course, is to separately test for the effect of national-standard adoptions for those treatment states with initially high and initially low cancer screening rates. For the breast-cancer-screening measures, this leaves only one state—Indiana—from which to explore the effect of a liability reform that entails a heightening of standards. In the case of pap smear testing, both Indiana and Rhode Island can be utilized as treatment states in exploring the effect of heightened standards. In either case, with only one or two treatment states, the point estimates from a difference-in-difference analysis are generally thought to be inconsistent (Conley and Taber 2011), leaving us with arguably unreliable estimates (given a higher degree of chance that spurious developments explain the findings). As such, we do not include them alongside the primary results from this analysis, which draw upon much more extensive legal variation. Nonetheless, we present these results in Online Appendix B. Consistent with our other quality measures, our results from cancer screening likewise document an increase in quality attainment (in this case, an increase in cancer screening rates) upon a modification of standard-of-care rules that entail a heightening of expectations.

Specification checks—In addition to those performed above, we note that these liability-standards findings are robust to certain additional specification checks and alternative approaches. First, we note that the results are largely robust to a weighting of the observations using the inverse of the propensity of each treated/untreated observation to be treated/untreated based on the available covariates (Rosenbaum and Rubin, 1983; Imbens and Rubin, 2014). In an effort to better approach randomization in the assignment of treatment and control states and to achieve covariate balance, this approach effectively puts more weight on those observations whose treatment status is difficult to predict based on the observables. For instance, while Table 2 depicts a 35-55 percent reduction (i.e., improvement) in low-discretionary avoidable hospitalization rates in initially-low-quality states upon a national standard adoption, we estimate a statistically-significant 26-35 percent negative response when imposing inverse propensity weights (with similar patterns across the range of additional estimates).[15] Second, in the Online Appendix, we discuss the use of randomization inference in exploring the statistical significance of the effects of national-standard adoptions, a flexible approach that generates unbiased standard error estimates even in the face of a limited number of treatment groups.

[15] The full set of inverse propensity weight results are available from the authors upon request. The results remain nearly unchanged when trimming the sample to those observations with propensity weights in the 0.1-0.9 range (Crump et al, 2009).

Table 3: Relationship between Damage Caps and Non-Obstetric Health Care Quality Metrics

	(1)	*(2)*	*(3)*
Panel A. Dependent Variable: Inpatient Mortality Rate for Selected Conditions (Logged)			
Non-Economic Damage Cap	0.008 (0.030)	-0.015 (0.027)	-0.038 (0.029)
Collateral Source Rule Reform	-	0.007 (0.023)	0.011 (0.040)
Punitive Damage Cap	-	0.003 (0.038)	0.013 (0.045)
Joint and Several Liability Reform	-	0.018 (0.041)	-0.003 (0.038)
95% Confidence Band for Coefficient of Non-Economic Damage Cap Variable	[-0.052, 0.066]	[-0.068, 0.039]	[-0.096, 0.021]
F-Statistic (Malpractice Variables Jointly = 0)	-	0.13	0.54
Prob > F (p value)	-	0.97	0.71
N	1154	1093	1093
Control Variables?	NO	YES	YES
State-Specific Linear Trends?	NO	NO	YES
Panel B. Dependent Variable: Low-Discretionary Avoidable Hospitalization Rate (Logged)			
Non-Economic Damage Cap	-0.005 (0.024)	-0.041* (0.020)	-0.035 (0.023)
Collateral Source Rule Reform	-	-0.006 (0.027)	0.000 (0.037)
Punitive Damage Cap	-	-0.006 (0.028)	-0.037 (0.039)
Joint and Several Liability Reform	-	0.004 (0.042)	0.011 (0.032)
95% Confidence Band for Coefficient of Non-Economic Damage Cap Variable	[-0.052, 0.043]	[-0.081, 0.000]	[-0.082, 0.012]
F-Statistic (Malpractice Variables Jointly = 0)	-	1.19	0.91
Prob > F (p value)	-	0.328	0.464
N	1139	1128	1128
Control Variables?	NO	YES	YES
State-Specific Linear Trends?	NO	NO	YES

Notes: robust standard errors corrected for within-state correlation in the error term are reported in parentheses. All regressions included state and year fixed effects. Regressions in Panel A are weighted by the number of admissions (for the relevant state and year) in the sub-sample of discharges associated with the relevant selected conditions (e.g., acute myocardial infarction). Regressions in Panel B are weighted by the low-variation health index (i.e., the sum of discharges for acute myocardial infarction, stroke, hip fracture or gastrointestinal bleeding) associated with each state-year cell. Mortality rates are risk-adjusted for the incidence (among the sub-sample) of each of the conditions comprising the sub-sample of selected conditions.
Source: 1977 – 2005 National Hospital Discharge Surveys.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

B. Damage-Cap Analysis

Overview—Tables 3 – 5 present estimation results from specifications that explore the relationship between non-economic damage caps and health care quality. In general, we estimate an association between a cap and the relevant indicator that is statistically indistinguishable from zero, though relatively tightly bound around zero. As such, though we cannot rule out that greater malpractice pressure within the existing system of liability standards—as identified through the lack of a non-economic damages cap—induces higher quality health care on average for these quality domains, we can rule out that such forces induce substantially higher levels of quality. That is, while the above analysis suggests that an elevation in liability standards has the potential to deter medical errors and improve patient quality, the existing custom-focused system of liability standards does not appear to be substantially improving quality on the margin.

AHRQ-Inspired Measures and Preventable Delivery Complications—We begin by describing the results for the AHRQ-inspired health care quality indicators and the preventable delivery complications measure (Tables 3-4), considering that these measures all reflect lower levels of quality as the relevant indicator level rises (and vice versa), whereas the cancer screening measures, which we discuss in subsection B(2) below (and Table 5), reflect higher levels of quality as the screening levels rise. We separate the discussions with this difference in mind to ease confusion in exploring the relevant associations.

Upon the adoption of a non-economic damage cap, we estimate mean changes in the inpatient mortality rate for selected conditions, the low-discretionary AH rate, the maternal trauma rate and the preventable delivery complication rate of 0.8, -0.5, -7.8, and -4.9 percent, respectively. This pattern of point estimates does not change meaningfully upon the inclusion of state-year covariates, other tort laws and state-specific linear time trends, as demonstrated by Columns 2 and 3 in each of Tables 3 – 4 (-3.8, -3.5, -4.9, and -0.8 percent, respectively).

Table 4: Relationship between Damage Caps and Obstetric Health Care Quality Metrics among Delivery Sample

	(1)	*(2)*	*(3)*
Panel A. Dependent Variable: Maternal Trauma Rate (logged)			
Non-Economic Damage Cap	-0.078*	-0.095**	-0.049
	(0.044)	(0.045)	(0.045)
Collateral Source Rule Reform	-	-0.017	-0.004
		(0.082)	(0.087)
Punitive Damage Cap	-	-0.116**	-0.169***
		(0.056)	(0.053)
Joint and Several Liability Reform	-	0.169*	0.010
		(0.010)	(0.096)
95% Confidence Band for Coefficient of Non-Economic Damage Cap Variable	[-0.165, 0.009]	[-0.185, -0.006]	[-0.143, 0.045]
F-Statistic (Malpractice Variables Jointly = 0)	-	1.94	2.93
Prob > F (p value)	-	0.11	0.03
N	1105	1053	1053
Control Variables?	NO	YES	YES
State-Specific Linear Trends?	NO	NO	YES
Panel B. Dependent Variable: Preventable Delivery Complication Rate (Logged)			
Non-Economic Damage Cap	-0.049	-0.028	-0.008
	(0.034)	(0.031)	(0.033)
Collateral Source Rule Reform	-	-0.013	0.030
		(0.055)	(0.063)
Punitive Damage Cap	-	-0.020	0.007
		(0.058)	(0.072)
Joint and Several Liability Reform	-	0.027	-0.028
		(0.062)	(0.063)
95% Confidence Band for Coefficient of Non-Economic Damage Cap Variable	[-0.119, 0.020]	[-0.091, 0.035]	[-0.074, 0.057]
F-Statistic (Malpractice Variables Jointly = 0)	-	0.24	0.12
Prob > F (p value)	-	0.92	0.97
N	1142	1083	1083
Control Variables?	NO	YES	YES
State-Specific Linear Trends?	NO	NO	YES

Notes: robust standard errors corrected for within-state correlation in the error term are reported in parentheses. All regressions included state and year fixed effects.
Source: 1977 – 2005 National Hospital Discharge Surveys.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

These estimates are generally not significant at the p=0.05 level of significance. One exception, however, is the maternal trauma results in some of its specifications. Even in this case, however, the estimated coefficient suggests that damage caps improve, rather than harm, maternal trauma outcomes, contrary to a story in which the current system of liability standards is deterring harmful errors on the margin. With respect to the remaining measures, we cannot rule out that positive associations between damage caps and these various quality indicators exist. However, even at the upper end of the 95 percent confidence interval, we find that the adoption of non-economic damage caps is associated with only a 2.1, 1.2, 4.5, and 5.7 percent increase in those same quality measures, respectively, as indicated in Tables 3 and 4 (or a 6.6, 4.3, 0.9, and 3.4 percent increase in the case of the basic difference-in-difference specifications). That is, higher malpractice pressure within our given liability system—captured by the lack of a damage cap—can at most lead to a modest level of improvement in quality (i.e., a modest amount of deterrence).

Table 5: Relationship between Damage Caps and Cancer Screening Rates

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
	MAMMO-GRAM	PHYSICAL BREAST EXAM	PROCTO-SCOPIC EXAM	PSA TESTING	DIGITAL RECTAL EXAM	PAP SMEAR
Non-Economic Damage Cap	-0.003 (0.006)	-0.005 (0.007)	-0.006 (0.005)	0.002 (0.006)	0.014 (0.008)	-0.007 (0.006)
95% Confidence Band for Coefficient of Non-Economic Damage Cap Variable (Percentage Point Impacts)	[-0.015, 0.008]	[-0.019, 0.009]	[-0.016, 0.003]	[-0.009, 0.013]	[-0.001, 0.030]	[-0.019, 0.005]
95% Confidence Band, scaled by mean screening rate (Percentage Impacts)	[-0.021, 0.011]	[-0.030, 0.014]	[-0.040, 0.008]	[-0.017, 0.025]	[-0.002, 0.060]	[-0.032, 0.008]
N	1009965	1155814	843960	252232	340931	1662616

Notes: robust standard errors corrected for within-state correlation in the error term are reported in parentheses. All regressions included state and year fixed effects.
Source: 1987 – 2008 Behavioral Risk Factor Surveillance System Records.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

In Table B3 of Online Appendix B, we present dynamic variants of the difference-in-difference specifications estimated in Tables 3 and 4, which include leads and lags of the damage-cap incidence variable, allowing us to explore how the differential in quality across treatment and control states evolves on a year-to-year basis. While the confidence bounds for each coefficient in this dynamic specification expand slightly with the inclusion of this additional set of policy variables, they continue to bound zero at a relatively tight rate confirming the conclusion of an at-most modest association between damage-cap adoptions and the various quality indicators. Online Appendix B likewise demonstrates the robustness of these findings to various additional specification checks, including, the consideration only of damage-cap adoptions that apply to tort contexts broadly, easing legislative endogeneity concerns—i.e., dropping states that adopted damage caps that apply only in the malpractice context.

In the Online Appendix, we also estimate specifications that take a novel, alternative approach to the codification of the damage-cap incidence variable. While the malpractice literature customarily codifies damage-cap adoptions in a simple binary fashion (0/1), non-economic damage cap provisions, in fact, take on a range of forms across jurisdictions. For instance, California imposes a flat, nominal $250,000 cap on non-economic damages awards, while Wisconsin imposes a $750,000 cap. Inspired by certain policy simulations performed in Hyman et al. (2009) and by the simulated instrument methodology employed in Currie and Gruber (1996), we codify caps by using closed-claims data from Texas during the period of time prior to the imposition of its non-economic damage cap (with information on the breakdown of economic versus non-economic damages associated with the claim) to simulate the degree to which the various damage-cap provisions across the various states reduce liability awards. We then use the results of this simulation exercise as the relevant damage-cap variable within the difference-in-difference specification, as opposed to the simple binary approach. The estimated mean coefficients from those specifications using this alternative codification of damage-cap variables do not differ substantially from those derived from the traditional binary approach.

Finally, we note that the non-economic damage cap results generalize to the other traditional tort reforms included as covariates (joint and several liability reforms, collateral source rule reforms and punitive damages caps), suggesting a generally weak relationship between both inpatient and outpatient health care quality and a broader range of remedy-focused reforms. In the case of the inpatient mortality rate, avoidable hospitalization rate, and preventable delivery complications measures, the results of an F-test of joint significance of all remedy-focused tort measures fail to reject the hypothesis that the coefficients of the various tort reforms are all jointly equal to zero. In the case of the maternal trauma specifications, the estimated coefficient of the punitive damages cap dummy is negative and bounded away from zero, suggesting an improvement in quality in connection with such reforms and thus counter to any expectation that such reforms would relax malpractice pressures to the detriment of patient quality.

The above-estimated specifications include state-year controls for physician concentration rates (and OB/GYN concentration rates in the case of the obstetrics measures). Such controls may absorb any impact of the reforms that occur through changes in the physician population. However, these simple controls may not absorb all supply-related consequences of such reforms. One effect of non-economic damage cap adoptions sometimes hypothesized is that lower-quality physicians may be attracted to the jurisdiction subsequent to the reform (Seabury 2010), a development which could otherwise confound any attempt to isolate the impact of malpractice pressure on the quality provided by any given provider. Of course, to the extent that non-economic damage caps would attract low-quality physicians and lead to a decline in observed quality – e.g., to an increase in the quality indicators explored in Tables 3 and 4 – this omission could only help to explain any positive effects of such reforms on the indicators explored. That is, a correction for this bias would likely push the estimated impacts of the reforms on the observed indicators even lower, only lending further support to the claim that marginal increases in liability pressure within the current liability system do not appear to substantially improve the quality of care being delivered by physicians.

Cancer Screening Measures—As presented in Table 5, the pattern of results from the cancer-screening / damage-cap analysis mirrors that from the AHRQ-inspired quality measures (with even greater precision in the estimates). We estimate mean associations between damage-cap adoptions and the various cancer screening rates that are very nearly zero in magnitude. As above, we cannot rule out some level of reductions in quality—i.e., some reduction in screening rates—in connection with damage cap reforms that are designed to reduce liability pressure. However, the 95-percent confidence bounds for each rate suggest that we can rule out that substantial reductions in screening rates are associated with caps. Lower bounds for these intervals suggest a 2.1, 3.0, 4.0, 1.7, 0.2 and 3.2 percent reduction (and an even lower percentage-point reduction) in mammography, physical breast, sigmoidoscopy/colonoscopy, PSA testing, and digital rectal and pap smear examinations, respectively. To simplify the presentation of these results, we present only the results from the basic difference-in-difference specifications. In Online Appendix B, we demonstrate the robustness of these findings to the addition of a range of control variables, along with alternative constructions of the screening rates.

C. DISCUSSION AND CONCLUSION

An extensive number of empirical malpractice studies have endeavored to test for the existence and scope of so-called "defensive medicine.[16]" While deterrence of medical errors can be viewed as a primary objective of the medical liability system, defensive-medicine is best characterized as a possibly unfortunate side-

[16] For a recent defensive-medicine analysis, see Mello et al. (2010).

effect / cost of this system. Physicians may act defensively when they unnecessarily order costly tests, procedures and visits over fear of malpractice liability (OTA 1994). However, even if one's primary focus is to explore these side effects of liability, rather than to assess whether the law is achieving its stated goal of deterring medical errors, it is critical to bear in mind that labeling a response as "defensive" requires more than a mere understanding of whether liability encourages additional utilization of medical care. Since a defensive response is defined with reference to the necessity (or optimality) of the chosen level of treatment, this assessment requires a determination as to whether or not any malpractice-induced expansion in treatment is accompanied by corresponding improvements in quality or outcomes (Mello et al. 2010).

As such, whether the goal is to make an independent evaluation of the deterrent impact of medical liability—i.e., to simply determine if liability forces are encouraging medical providers to avoid the commission of harmful errors—or to properly diagnose a "defensive" response to liability, it is necessary to estimate the impact of the malpractice system on medical errors and health care quality. To date, however, most studies which assess the impact of malpractice pressure on health care quality focus on coarse measures such as aggregate mortality rather than more direct measures of the quality of care provided by physicians. A major contribution of our analysis is to use clinically validated measures of health care quality to estimate the effect of malpractice pressure on the care provided by physicians. In this process, it is also important to bear in mind the structure of the malpractice system itself, a factor generally overlooked in most empirical discussions of this nature. In estimating the impacts of remedy-focused / non-substantive reforms such as non-economic damage caps, one is effectively teasing out the marginal impacts of malpractice penalties under the current standards of care legally expected of physicians. The confidence bounds presented in our analysis suggest, at most, a modest degree of deterrence stemming from the present liability system. The mean point estimates suggest that under existing liability standards, malpractice penalties generate little to no benefits in health care quality. We caution that these findings should perhaps not be interpreted so as to suggest that medical liability forces are universally incapable of improving quality. Rather, they should be interpreted in light of the largely self-regulatory nature of our present malpractice system—i.e., in light of the fact that the law itself is presently not designed to impose independent expectations regarding quality.

The first half of our empirical analysis provides some hope, however, in the potential for medical liability to influence physician behavior. Drawing upon the one type of standard-of-care reform that states have experimented with to date—i.e., locality rule abdications—we investigate the impact of changing the clinical standards of care imposed upon physicians under the law, both in terms of elevated standards and slackened standards. All told, it appears that the relationship between health care quality and changes in clinical malpractice standards works in an expansionary direction only. That is, once physicians provide a high level of quality, they may maintain such practices even when the law may loosen its expectations at a later date. In contrast, physicians who provide a quality of care that is below what is expected by the law raise their practices to meet the higher expectations set by the law. Malpractice forces that alter the legal clinical standard to which physicians are held may therefore be effective in elevating the quality floor.

If our findings are taken to suggest that structural reforms to the way in which physicians are evaluated may substantially alter health care delivery practices, one may wonder whether subsequent reforms—e.g., caps—that blunt the impact of the now altered liability system may cause practices to revert back to where they were before the structural reforms. Informational models of physician behavior suggest why this may not be so. If physicians, especially newer physicians, form beliefs over proper practices to a large extent through their own past experiences or through the observation of the practices followed by others around them (Phelps and Mooney 1993), then a shift in medical practices that arises in any manner—including that arising from fear over being out of compliance with changed legal expectations—may more gradually come to be assimilated into the belief structures of physicians over time. As such, malpractice-induced changes in practices may come to shape more durable physician norms and customs that may survive subsequent diminishment of liability forces. These considerations may thus help us understand why damage cap adoptions—which primarily arose in states after previous retreats from the locality rule—did not cause physician practices to revert back to their locality-rule-era levels.

Ultimately, empirical malpractice investigations that fail to appreciate the structural considerations underlying tort law may misinterpret the findings derived from our experiences to date with traditional remedy-centric tort reforms. Such findings may suggest only a weak responsiveness to the law despite a potentially meaningful role for the law to play in shaping clinical practices and health care quality. Substantial work remains, of course, to understand the liability structure that will best serve society. Our analysis demonstrates that it would be premature to rule out medical liability from the health care quality discussion based on the limited findings that derive from damage-cap-focused studies.

REFERENCES

Baicker, Katherine, and Amitabh Chandra. 2005. "The Effect of Malpractice Liability on the Delivery of Health Care" in David M. Cutler and Alan M. Garber (Eds.) Frontiers of Health Policy Research Vol 8.

Baicker, Katherine, Elliott S. Fisher, and Amitabh Chandra. 2007. "Malpractice Liability Costs And The Practice Of Medicine In The Medicare Program," Health Affairs 26(3): 841-52.

Centers for Medicare and Medicaid Services. 2011. National Health Expenditure Accounts: Tables. Available at: http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-andReports/NationalHealth ExpendData/ Downloads/tables.pdf (accessed June, 17 2012).

Chandra, Amitabh, and Douglas O. Staiger. 2007. "Productivity Spillovers in Healthcare: Evidence From the Treatment of Heart Attacks." *Journal of Political Economy* 115 (1): 103–40.

Conley, Timothy, and Christopher Taber. 2011. Inference with 'Difference-in-Differences' with a Small Number of Policy Changes. *The Review of Economics and Statistics* 1, 113-25.

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

Dubay, Lisa, Robert Kaestner, and Timothy Waidmann. 1999. The Impact of Malpractice Fears on Cesarean Section Rates. *Journal of Health Economics* 18, 491-522.

Frakes, Michael. 2012. Defensive Medicine and Obstetric Practices. *Journal of Empirical Legal Studies* 9(3): 457-81.

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(1)-257-76.

Hyman, David, Bernard Black, Charles Silver & William Sage. 2009. Estimating The Effect of Damage Caps in Medical Malpractice Cases: Evidence from Texas. *Journal of Legal Analysis* 1, 355-409.

Iizuka, Toshiaki. 2013. Does Higher Malpractice Pressure Deter Medical Errors? 56 *Journal of Law and Economics* 1, 161-88.

Jena, Anupam, Seth Seabury, Darius Lakdawalla, and Amitabh Chandra. 2011. Malpractice Risk according to Physician Specialty. *New England Journal of Medicine* 365, 629-636.

Kessler, Daniel, and Mark McClellan. 1996. Do Doctors Practice Defensive Medicine? *Quarterly Journal of Economics* 111, 353-90.

Klick, Jonathan, and Thomas Stratmann. 2007. Medical Malpractice Reform and Physicians in High-Risk Specialties. *Journal of Legal Studies* 36, S121-S142.

Lakdawalla, Darius, and Seth Seabury. 2009. The Welfare Effects of Medical Malpractice Liability. NBER Working Paper No. 15383.

McClellan, Mark, and Douglas Staiger. 1999. The Quality of Health Care Providers. NBER Working Paper No. 7327.

Mello, Michelle M., Amitabh Chandra, Atul A. Gawande, and David M. Studdert. 2010. "National Costs of the Medical Liability System," *Health Affairs*, 29: 1569-77.

Mello, Michelle M. and Allen Kachalia. 2011. Evaluation of options for medical malpractice system reform: a report to the Medicare Payment Advisory Commission, (MedPAC, 2010), available at http://www.medpac.gov/documents/Apr10_MedicalMalpractice_ CONTRACTOR.pdf.

National Center for Health Statistics. 1977-2005. National Hospital Discharge Survey. Centers for Disease Control and Prevention (last accessed at NCHS Research Data Center on January 31, 2012).

Office of Technology Assessment. 1994. Defensive Medicine and Medical Malpractice. OTA-H--6O2.

Paik, Myungho, Bernard Black and David Hyman. 2013. The Receding Tide of Medical Malpractice Litigation Part 2: Effect of Damage Caps. *Journal of Empirical Legal Studies*, forthcoming.

Phelps, Charles E., and Cathleen Mooney. 1993. "Variations in Medical Practice Use: Causes and Consequences," in Competitive Approaches to Health

Care Reform, edited by Richard J. Arnould, Robert F. Rich, and William D. White, 140–78. Washington, DC: The Urban Institute Press.

Seabury, Seth. 2010. Does Malpractice Liability Reform Attract High Risk Doctors? RAND Working Paper No. WR-674-ICJ.

Sloan, Frank A. 1990. "Experience Rating: Does It Make Sense for Medical Malpractice Insurance?" American Economic Review (Papers and Proceedings), 80(2): 128–133.

Sloan, Frank A., and John H. Shadle. Is there empirical evidence for "Defensive Medicine"? A reassessment. *Journal of Health Economics* 28, 481-491.

Wennberg, John, and Alan Gittelsohn. 1973. "Small Area Variations in Health Care Delivery: A population-based health information system can guide planning and regulatory decision-making." *Science* 182.4117: 1102-1108.

FOR ONLINE PUBLICATION

ONLINE APPENDIX

DOES MEDICAL MALPRACTICE LAW IMPROVE HEALTHCARE QUALITY?

By MICHAEL FRAKES AND ANUPAM B. JENA

Online Appendix A: Data Sources, Quality Measures, and Covariates

National Hospital Discharge Survey

Healthcare quality data is collected from the National Hospital Discharge Survey (NHDS), a nationally-representative sample of inpatient discharge records from short-stay, non-federal hospitals. For approximately 260,000 inpatient records per year, the NHDS contains information on, among other things: (a) primary and secondary diagnosis and procedure codes, (b) certain demographic characteristics of the patient, and (c) certain characteristics of the hospital. We supplement the public NHDS files with geographic identifiers (restricted-use variables) received pursuant to an agreement with the Research Data Center (RDC) at the National Center for Health Statistics (NCHS). All empirical work was performed onsite at the RDC in Hyattsville, Maryland. The resulting sample covers the years 1977 to 2005. Note that some of the more complicated quality indicators focus only on the 1979 to 2005 period given that the 1977 and 1978 NHDS records use the more coarse ICD-8 diagnosis codes (as distinct from the ICD-9-CM codes used thereafter), complicating the ability to form consistent formulations over time of the relevant measures.

Healthcare Quality Measures

For the purposes of this study, we largely look to the AHRQ for guidance in selecting quality metrics. The AHRQ measures are particularly useful for the present study in so far as they are designed for use with administrative inpatient databases such as the NHDS. The AHRQ's quality indicators are essentially classified into 3 modules: (1) Prevention Quality Indicators (PQIs), identifying admissions that could have been avoided through access to high-quality outpatient care, (2) Inpatient Quality Indicators (IQIs), reflecting the quality of care inside hospitals including inpatient mortality for certain medical conditions, and (3) Patient Safety Indicators (PSIs), focusing on potentially avoidable complications during inpatient care.

For the purposes of this analysis, we attempt to construct quality metrics that are meant to cover each of these three domains.

Avoidable hospitalizations. First, we calculate a rate of avoidable hospitalizations (AH) within each state-year cell, a measure inspired by the AHRQ's PQIs. AH rates, generally, and the PQIs, specifically, are measures that are constructed using inpatient data, though meant to reflect the quality of care prevailing in the associated outpatient / ambulatory community. Such measures identify conditions (e.g., asthma, diabetes, malignant hypertension, etc.) with respect to which proper outpatient care would have prevented the need for hospitalization. According to the AHRQ, their PQIs grew out of research in the early 1990s by Joel Weissman and colleagues.[1] The Weissman et al. (1992) AH classification scheme is designed in slightly more general terms than the PQIs and thus arguably lends itself to easier codification using a set of NHDS records that span several decades, considering the complexity associated with tracking variations in ICD classifications over time.[2] For this reason, and in light of the fact that Weismann et al. developed their classification during the middle of the period in which the NCHS sampled physicians to compile the NHDS (unlike the PQIs, which came later), we elect to construct an AH rate for this analysis using the Weissman et al. classification.

To calculate avoidable hospitalization rates for each state and year in the sample, we first count the number of hospitalizations within the NHDS records for that state-year cell in which a diagnosis is indicated for any of the conditions included in the Weissman et al. (1992) classification. We perform such counts under two alternative approaches: one in which the conditions are identified in any one of the indicated diagnosis codes and one in which the conditions are identified in the primary diagnosis code only (the preferred approach that we take). To form the relevant rate, it is of course necessary to normalize these AH counts in some manner. Following Frakes (2013), we elect to use measures internal to the NHDS records to form the relevant denominator for each state-year AH rate, taking several alternative approaches to this normalization.[3] In one approach, for example, we normalize each AH count by the number of

---

[1] See http://www.qualityindicators.ahrq.gov/Downloads/Modules/PQI/PQI%20Summary%20Report.pdf.
[2] Those conditions represented in the Weissman et al. (1992) classification include: ruptured appendix, asthma, cellulitis, congestive heart failure, diabetes, gangrene, hypokalemia, immunizable conditions, malignant hypertension, pneumonia, pyelonephritis, and perforated or bleeding ulcer.
[3] The NHDS weights are not designed to generate representative state-specific estimates. Of course, observing within-state changes over time in the set of records included in the state-year cells nonetheless affords the ability to identify the intended relationships (Dafny and Gruber 2005). In any event, though noisier, the results of this exercise generally persist under alternative approaches that either (1) multiply observations by the NHDS sample weights and form AH rates by dividing weighted AH counts by the total population of that state (yet another normalization approach), or (2) forming dependent variables based on the natural log of the state-year AH counts (i.e., under no normalization at all). The primary approaches taken, however, soften some of the sampling variability that occurs within states over time, while normalizing by a measure that is more directly reflective of the scale of the hospital sampled.

hospitalizations associated with the delivery of a child found in the NHDS records for the relevant state and year. This approach allows for a scaling of the AH count by a measure reflective of the size of the associated state-year sample, while also offering a denominator that is itself not likely to be significantly impacted by the prevailing malpractice environment (allowing for a focus on the influence of malpractice on the AH count comprising the numerator, our margin of interest).

Primarily, however, based on the same premise as the delivery approach and following Frakes (2013), we normalize each state-year AH count by an index of hospitalizations equal to the count of admissions associated with any of the following conditions and events: (1) acute myocardial infarction, (2) stroke, (3) gastro-intestinal bleeding or (4) hip fracture. Such events represent situations characterized by relatively little variation across regions (see, for example, Wennberg 1984 and Wennberg and Cooper 1999), even in the face of environments that impose varying legal and financial incentives (i.e., where such hospitalizations are better seen as proxies for the underlying disease environment, as opposed to reflections of immediate healthcare utilization decisions). As such, this index likewise affords an appropriate scaling of the numerator count with arguably little concern over the malpractice environment impacting the scaling metric.[4] In yet another alternative approach, we simply normalize by the count of acute myocardial infarction discharges (primary diagnosis only) for the relevant state and year.

Low-discretionary avoidable hospitalizations. As a more refined AH rate, we focus on those subset of avoidable hospitalizations over which physicians have less discretion in admitting patients. Use of this alternative measure will ease concerns that fluctuations in the liability regime will capture changes not just in outpatient quality but in inpatient admission decisions. Following Weismann et al. (1991), Wennberg (1988) and Twigger and Jessop (2000) for guidance, we select the following conditions out of the Weissman et al. (1992) conditions as being on the lower end of the discretionary scale: ruptured appendix, pneumonia, and congestive heart failure.

With these latter selection concerns in mind, the main text focuses on low-discretionary avoidable hospitalizations. We note that the findings from this more refined approach are virtually identical to the broader AH rate construction (results available upon request).

Inpatient mortality for selected conditions. Following the AHRQ's IQIs, we next construct a quality measure in which we calculate the composite rate of inpatient mortality among a sub-sample of discharges in which the primary diagnosis code indicates any one of the following conditions: acute myocardial infarction, heart failure, acute stroke, gastrointestinal bleeding, hip fracture or pneumonia. Such events are generally high volume in occurrence, allowing for robust sample sizes. It is worth noting that such conditions, for the most part, also represent low-

[4] See Frakes (2013) for empirical support over the contention that the incidences of these low-variation conditions are not sensitive to medical liability standards. Note that higher quality outpatient care may be effective at reducing some amount of hospitalizations for the above-indicated low-variation conditions, though likely to an extent less than quality care may reduce the incidence of the Weissman et al. (1992) avoidable conditions, in which case the proposed avoidable hospitalization rate nonetheless identifies a relative quality measure.

discretionary hospitalizations, whereby inpatient admissions generally follow upon their occurrence.[5] With this in mind, mortality rates among this sub-sample of admissions can be seen as more likely reflective of the quality of care observed during the inpatient stay itself, rather than as a result of risk selection by providers or patients.

Of course, a concern arises regarding fluctuations in the proportions of the various conditions comprising this selected-conditions sub-sample. That is, a reduction in the composite mortality rate could arise from a relative increase in the rate of hip fracture admissions (where mortality rates are lower for such admissions relative to the other selected conditions), as opposed to reductions in mortalities that would actually be attributable to improvements in quality. We take two approaches to dealing with this concern. First, in some specifications, we include state-year controls for the proportion of this sub-sample made up of each of the respective conditions. In the primary approach, however, we follow the AHRQ and standardize the composite mortality rate for state-year changes in the various incidences of the conditions.

To risk adjust mortality rates, we employ an indirect standardization approach, in which we first predict the mortality rate that a national sample of patients would be expected to experience if they faced the relevant patient characteristics of each state-year cell. We generate such predictions based on the estimated coefficients from national, annual regressions of mortality incidence on the incidence of the relevant set of conditions. We then calculate the standardized mortality rate by (1) taking the ratio between the observed state-year composite mortality rate and this predicted national mortality rate and (2) multiplying this ratio by the observed national mortality rate.

Patient safety incidents and delivery complications. For the reasons set forth in the text, we focus our patient-safety analysis on the delivery-related PSI's inspired by the AHRQ, which represent third and fourth degree lacerations during deliveries (aggregating this analysis across vaginal and cesarean deliveries). Again following Currie and MacLeod (2008), we supplement these PSI delivery measures by forming a measure equal to the incidence of preventable delivery complications: fetal distress, excessive bleeding, precipitous labor, prolonged labor, or dysfunctional labor.

Behavioral Risk Factor Surveillance System

Our data source for the cancer-screening analysis is the Behavioral Risk Factor Surveillance System (BRFSS). The data consists of repeated cross-sections for the years 1987 through 2008, collected via monthly telephone surveys of individuals aged 18 years and older. The BRFSS is a nationally representative survey of the United States and has been conducted by state health departments in coordination with the CDC for the purpose of collecting state-level data pertaining to certain personal health behaviors. Fifteen states took part in the first survey in 1984. By 1994,

[5] For a discussion of the selection of low-discretionary hospitalization categories, see Carter (2003).

all 50 states and the District of Columbia became involved. The survey was administered to an average of 817 individuals per state in 1984, rising to an average of nearly 8000 per state in 2008.

Cancer-Screening Measures

Sigmoidoscopy / Colonoscopy. In our primary specification, we aimed to construct a proctoscopy screening measure in line with recommended screening guidelines. As such, we focused on the age group between 50 and 75 years old and created an indicator variable equal to "1" if the respondent has had a sigmoidoscopy or a colonoscopy within the last 5 years. In alternative specifications we simply indicate whether or not the respondent within this age range has ever had a sigmoidoscopy or a colonoscopy. Proctoscopic examination information within the BRFSS is available from 1988 onwards.

Mammogram. In our primary specification, we construct a mammogram screening measure in line with the recommended screening guidelines in place for most of our sample period. Accordingly, limiting our sample to those female respondents with an age between 40 and 75 year olds, we created an indicator variable reflecting whether or not the respondent received a mammogram within the last 2 years. In alternative specifications, we simply indicate whether or not the respondent within this age range has ever had a mammogram. Mammography information within the BRFSS is available from 1987 onwards.

Physical breast exam. Likewise in line with recommended guidelines, our primary specifications construct physical or clinical breast exam utilization measures by looking at the sample of at least 40 years of age and asking whether or not they have had a breast exam within the last year. In alternative specifications, we simply indicate whether or not they have ever had a physical breast exam. Physical breast exam information within the BRFSS is available from 1990 onwards.

PSA Testing. Consistent with recommendations, at least with respect to those recommendations operating over our sample period, we focus on the sample of males over the age of 50 (and under the age of 75) and construct an indicator regarding whether or not they have received Prostate-Specific Antigen (PSA) testing within the last year. In alternative specifications, we simply indicate whether or not they have ever had PSA testing. PSA testing information within the BRFSS is available from 2001 onwards.

Digital Rectal Exam. Consistent with recommendations, at least with respect to those recommendations operating over our sample period, we focus on the sample of males over the age of 50 (and under the age of 75) and construct an indicator regarding whether or not they have received a Digital Rectal Exam (DRE) within the last year. In alternative specifications, we simply indicate whether or not they have ever had a DRE. DRE information within the BRFSS is available from 1988 onwards, though not at sufficient numbers until 1993 onwards (with several years omitted in the late 1990s).

Pap smear. Consistent with recommendations, at least with respect to those recommendations operating over our sample period, we focus on the sample of females 21 years old and over and construct an indicator regarding whether or not they have received pap testing within the last year. In alternative specifications, we simply indicate whether or not they have ever had a pap smear. Pap testing information within the BRFSS is available from 1987 onwards.

<u>Additional Notes on Non-Economic Damage Caps</u>

Following Frakes (2012), we also classify states as having non-economic damages provisions if they have laws that place caps on total damages awards. Such laws, after all, necessarily cap non-economic damages as well. In light of the imposition of state fixed effects, this classification only has relevance in the context of 1 state (Texas) that adopted a total damages cap at a time when it did not have a specific non-economic damage cap in place. Only 1 additional state – i.e., Colorado – adopted a total damages cap over the sample period (2 years following the adoption of a non-economic damages cap). With this in mind, we do not separately control for the incidence of a cap on total damages. However, we estimate nearly identical results for the remaining coefficients when we do include this additional covariate and treat total and non-economic damage caps separately.

Frakes (2013) documents a relationship between the adoption of laws requiring physicians to follow national (as opposed to local) standards and a resulting convergence in physician practices across regions. In light of the fact that two of the damage-cap treatment states used in the defensive-medicine analysis below (Hawaii and Texas) were dropped from the specifications estimated in Frakes (2013) (due to an inability to classify the full history of their standard-of-care laws), we exclude controls for national-standard laws in the damage-cap specifications estimated below and focus instead on the traditional tort reform measures. However, the results presented below are robust to the inclusion of controls for national-standard laws (not shown).

<u>Other Tort Reforms</u>

A number of specifications include the incidence of additional tort measures as covariates, including reforms of the collateral source rule and the joint and several liability rule and caps on punitive-damages awards. Traditional collateral source rules generally prohibited defendants from introducing evidence of compensatory payments made to plaintiffs from outside sources (e.g., insurers). Thirty-three states currently have laws in place that eliminate this traditional rule, effectively reducing the scope of compensatory damage awards. Much of these reforms likewise occurred during the mid-1980s; however, there are a substantial amount of independent reforms of each type, facilitating identification of their separate impacts.

Punitive damages are awarded on a much rarer basis in malpractice actions than are non-economic damages awards (without a correspondingly large increase in average payouts).[6] Thus, relative to non-economic damages, it is arguable that the threat of liability for punitive damages will have a weaker impact on physician behavior. Nonetheless, despite the infrequent application of such awards, considering that punitive damages are generally not insured by liability carriers, it remains reasonable to believe that physicians may be sensitive to the threat posed by punitive awards (Malani and Reif 2012).

Finally, we look to reforms of the common law joint and several liability rule. Under the common law approach, when there is more than one liable defendant, the plaintiff can seek full recovery against any one defendant, even if that one defendant was only responsible for a small portion of the damages. Reforms to this common law rule generally pushed in the direction of holding defendants responsible for a share of the damages proportionate with their responsibility (specified in various ways).

Other Covariates (by Quality Indicator)

Inpatient mortality rate for selected medical conditions. In the case of the mortality rates specifications, estimated according to equation (1) in the text, $\mathbf{X}_{s,t}$ represents certain demographic characteristics: the percentage of patients in various age-sex categories,[7] race categories (white, black and other), insurance categories (private, government, no insurance and other), along with the percentage of patients visiting hospitals of various bed sizes (0-100, 100-200, 200-300, 300-500 and 500+ beds) and of various ownership types (proprietary, non-profit and government).[8] $\mathbf{Z}_{s,t}$ represents certain other state-year characteristics (HMO penetration rate and its square, physician concentration rate, and median household income).[9]

In alternative specifications, we also control for the average length of stay associated with hospitalizations for such medical conditions. To the extent that medical liability forces also impact lengths of stay for such hospitalizations, any such development could confound the estimation of

[6] For evidence of this claim, see Cohen (2005) and Hyman et al. (2009).

[7] Age-sex categories for the inpatient mortality and AH specifications are as follows: male under 30, female under 30, male 30-45, female 30-45, male 45-55, female 45-55, male 55-65, female 55-65, male 65-75, female 65-75, male over 75 and female over 75. Age-sex categories for the obstetric specifications are as follows: 15-19, 20-24, 25-29, 30-34, 35-39 and 40+ years old.

[8] We form the incidences of the relevant demographic variables using the NHDS sample itself, though the results are entirely robust to alternative state-year controls based off of the Census data. Following Frakes (2013), in the AH rate and mortality rate specifications, we form the relevant incidences using the sample of discharges in which patients present themselves for acute myocardial infarction, stroke, gastro-intestinal bleeding or hip fracture. This subsample consists of patients that will almost universally seek hospitalization upon the occurrence of the event, in which case the sample itself is generally not sensitive to the prevailing legal environment. In any event, the results of this exercise are also robust to the formation of the demographic covariates using the entire sample of state-year NHDS discharges. In the obstetrics specifications, we form all relevant incidences using the subsample of discharges associated with deliveries.

[9] HMO penetration rates are from Interstudy Publications. Household income data is from the decennial Census files and the American Community Surveys. Data on physician population counts are from the American Medical Association (AMA) administrative records and were obtained from the Area Resource File.

liability forces on inpatient mortality rates insofar as longer hospitalizations otherwise increase the probability of an inpatient mortality. The results are virtually unchanged with such controls. Supporting this insensitivity to the inclusion of length-of-stay controls, we also find, in separate specifications (available upon request), no association between the adoption of the various reforms and the length of stay associated with hospitalizations for the selected medical conditions.

Avoidable hospitalization rates. $\mathbf{X}_{s,t}$ and $\mathbf{Z}_{s,t}$ in the AH rate specifications are identical to those of the inpatient mortality rate specifications.

Maternal trauma rates and delivery complication rates. In the obstetrics specifications, **X** includes mother's age (15-19, 20-24, 25-29, 30-34, 35-39 and 40+ years old); mother's race (white, black and other); mother's insurance status (private, government, no insurance and other); hospital bed size (0-100, 100-200, 200-300, 300-500 and 500+ beds); and hospital ownership type (proprietary, non-profit and government). $\mathbf{Z}_{s,t}$ includes the state-year fertility rate, the state-year OB-GYN concentration rate,[10] the HMO penetration rate (and its square), and median household income. Obstetric specifications also include controls for cesarean delivery and episiotomy utilization. The maternal trauma specifications also include a control capturing the risk-status associated with the delivery, specified following Frakes (2013) as the predicted probability of cesarean delivery (PPC). PPC values are calculated using fitted values of a logit model (estimated annually) of the incidence of cesarean delivery on a set of individual risk factors and complications. We include this measure from Frakes (2013) simply as a way to capture all such risk factors and complications in a single measure. The results are robust to including separate indicator variables for all such measures. Note that we exclude this control in the main specification of the delivery complications specification given that the outcome variable in that context is meant to capture certain of those complications itself. In alternative specifications of the delivery complications approach, we also include controls for all non-preventable complications and risk factors. The results are virtually identical under such alternative specifications (available upon request).

Cancer Screening Rates. **X** in the cancer screening specifications includes various individual characteristics provided for in the BRFSS files: marital status (married, widowed, divorced, single), race (white, black, and other), educational attainment category, Hispanic origin, income (and its square), age category (by age deciles), and smoking status. **Z** includes certain characteristics of the prevailing state-year health care market (including physician concentration rate and the average number of hospital beds per capita),[11] along with HMO penetration rates and its square.

---

[10] Fertility rates are calculated according to Gruber and Owings (1996) as the number of births per population and come from the Vital Statistics Natality files (also obtained via the ARF).

[11] Average hospital bed data was likewise obtained from the ARF.

Note on Liability-Standards Specifications

Following Frakes (2013), we exclude from this initially-high versus initially-low analysis the state of Maryland, which modified its standard of care laws over the 1990s to retreat from a previous national-standard adoption, insofar as it is difficult to hypothesize the direction in which practices will evolve subsequent to this retreat.

Online Appendix B: Additional Robustness Checks

<u>Dynamic Difference-in-Difference Results</u>

In the following tables, we extend the analysis in the text to include difference in difference regression results that include a number of leads and lags of the key legal reforms—damages caps and national-standard reforms.

Note that we specify the 4-year lead coefficient as turning from zero to one 4 years prior to the national-standard-law adoption in the relevant state and staying at 1 thereafter. The other lead and lag variables are specified accordingly. With this specification, the coefficient of the 4-year lead coefficient captures the differential in the relevant rate between treatment and control states in the period of time between the three- and four-year period prior to a national-standard adoption and the years prior to that period. The coefficient of the 3-year lead variable then captures the *subsequent* change in this differential as we move into the next year-long period—i.e., the differential quality rate between treatment and control states in the 2-3-year-prior period relative to the 3-to-4-year-prior period. And so on and so forth. To capture the cumulative time trend in the differential quality rate between treatment and control states—with time entailing years prior to and subsequent to national-standard adoptions—one naturally adds up these subsequent coefficient levels. The figures presented in the text take this cumulative approach.

Table B1. Relationship between National Standard Laws and the AHRQ-Inspired Quality Indicators in Initially Low-Quality Areas: Dynamic Difference-in-Difference Regression Results

	(1)	(2)	(3)	(4)
	INPATIENT MORTALITY RATE	LOW-DISCRETION AH RATE	MATERNAL TRAUMA RATE	PREVENTABLE DELIVERY COMPLICATIONS
National-Standard Law Dummy				
4-Year Lead Dummy	-0.069	-0.052	0.304	0.012
	(0.118)	(0.169)	(0.229)	(0.147)
3-Year Lead Dummy	-0.154	0.121	-0.193	0.034
	(0.083)	(0.201)	(0.252)	(0.099)
2-Year Lead Dummy	0.104	0.003	0.132	-0.121*
	(0.095)	(0.086)	(0.261)	(0.069)
1-Year Lead Dummy	0.097	-0.088	0.227**	0.067
	(0.147)	(0.086)	(0.077)	(0.154)
Contemporaneous Dummy	-0.133*	-0.065	-0.481*	-0.490***
	(0.069)	(0.115)	(0.263)	(0.142)
1-Year Lag Dummy	-0.031	-0.265*	0.028	0.332**
	(0.079)	(0.134)	(0.330)	(0.139)
2-Year Lag Dummy	-0.083	-0.012	0.126	-0.180
	(0.123)	(0.066)	(0.163)	(0.106)
3-Year Lag Dummy	0.112	-0.058	-0.129	0.182
	(0.179)	(0.106)	(0.150)	(0.176)
4-Year Lag Dummy	-0.184	-0.314	-0.038	0.002
	(0.134)	(0.193)	(0.166)	(0.159)

Notes: robust standard errors corrected for within-state correlation in the error term are reported in parentheses. Each specification controls for state fixed effects, year fixed effects, various covariates and a set of state specific linear time trends. Specifications are weighted per their counterparts in Table 2 of the text. Dependent variables are logged.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

Table B2. Relationship between National Standard Laws and the AHRQ-Inspired Quality Indicators in Initially High-Quality Areas: Dynamic Difference-in-Difference Regression Results

	(1)	(2)	(3)	(4)
	INPATIENT MORTALITY RATE	LOW-DISCRETION AH RATE	MATERNAL TRAUMA RATE	PREVENTABLE DELIVERY COMPLICATIONS
National Standard Law Dummy				
4-Year Lead Dummy	-0.227**	-0.056	0.370	-0.011
	(0.109)	(0.054)	(0.244)	(0.119)
3-Year Lead Dummy	0.124	-0.156	-0.399	-0.283
	(0.153)	(0.137)	(0.278)	(0.190)
2-Year Lead Dummy	0.297	0.104	-0.070	0.029
	(0.283)	(0.070)	(0.146)	(0.096)
1-Year Lead Dummy	-0.066	-0.083	0.332	0.017
	(0.162)	(0.083)	(0.251)	(0.185)
Contemporaneous Dummy	0.140	-0.015	-0.329**	-0.052
	(0.267)	(0.100)	(0.126)	(0.262)
1-Year Lag Dummy	-0.214	-0.011	0.148	-0.002
	(0.197)	(0.101)	(0.228)	(0.171)
2-Year Lag Dummy	0.172	0.066**	-0.119	0.065
	(0.253)	(0.031)	(0.117)	(0.161)
3-Year Lag Dummy	0.030	-0.094**	0.291	-0.068
	(0.126)	(0.039)	(0.240)	(0.084)
4-Year Lag Dummy	0.011	0.009	0.149	0.109
	(0.112)	(0.044)	(0.249)	(0.136)

Notes: robust standard errors corrected for within-state correlation in the error term are reported in parentheses. Each specification controls for state fixed effects, year fixed effects, various covariates and a set of state specific linear time trends. Specifications are weighted per their counterparts in Table 2 of the text. Dependent variables are logged.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

Table B3. Relationship between Damage Caps and the AHRQ-Inspired Quality Indicators: Dynamic Difference-in-Difference Regression Results

	(1)	(3)	(4)	(5)
	INPATIENT MORTALITY RATE	LOW-DISCRETION AH RATE	MATERNAL TRAUMA RATE	PREVENTABLE DELIVERY COMPLICATIONS
Non-Economic Damage Cap				
4-Year Lead Dummy	-0.083***	0.044	0.074	-0.046
	(0.028)	(0.024)	(0.045)	(0.053)
3-Year Lead Dummy	-0.012	-0.026	-0.104**	0.053
	(0.037)	(0.021)	(0.041)	(0.039)
2-Year Lead Dummy	-0.010	-0.006	0.015	0.022
	(0.057)	(0.031)	(0.065)	(0.042)
1-Year Lead Dummy	0.029	-0.001	0.013	-0.051
	(0.042)	(0.023)	(0.061)	(0.050)
Contemporaneous Dummy	-0.051	-0.011	-0.037	0.013
	(0.046)	(0.028)	(0.067)	(0.042)
1-Year Lag Dummy	0.030	-0.011	-0.054	-0.002
	(0.066)	(0.023)	(0.077)	(0.038)
2-Year Lag Dummy	0.037	-0.055**	0.049	0.000
	(0.068)	(0.027)	(0.092)	(0.076)
3-Year Lag Dummy	0.031	0.030	0.007	-0.010
	(0.053)	(0.029)	(0.091)	(0.054)
4-Year Lag Dummy	-0.041	-0.004	0.013	0.005
	(0.050)	(0.021)	(0.064)	(0.041)

Notes: robust standard errors corrected for within-state correlation in the error term are reported in parentheses. Each specification controls for state fixed effects, year fixed effects, various covariates and a set of state specific linear time trends. Specifications are weighted per their counterparts in Tables 3-4 of the text and otherwise track the specifications in such tables. Dependent variables are logged.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

Table B4. Relationship between Damage Caps and Cancer Screening Rates: Dynamic Difference-in-Difference Regression Results

	(1) MAMMO-GRAM	(2) PHYSICAL BREAST EXAM	(3) PROCTO-SCOPIC EXAM	(4) PSA TESTING	(5) DIGITAL RECTAL EXAM	(6) PAP SMEAR
<u>Non-Economic Damage Cap</u>						
4-Year Lead Dummy	-0.009	-0.006	-0.019***	0.013	0.007	-0.003
	(0.006)	(0.006)	(0.007)	(0.015)	(0.017)	(0.006)
3-Year Lead Dummy	0.007	0.010	-0.004	-0.023	-0.009	0.015*
	(0.004)	(0.008)	(0.007)	(0.016)	(0.015)	(0.007)
2-Year Lead Dummy	0.005	-0.001	-0.002	0.022*	0.018	-0.002
	(0.006)	(0.008)	(0.007)	(0.011)	(0.012)	(0.007)
1-Year Lead Dummy	0.017**	0.032***	-0.001	0.024***	-0.006	0.025***
	(0.007)	(0.007)	(0.011)	(0.009)	(0.014)	(0.006)
Contemporaneous Dummy	-0.023***	-0.019**	-0.022	0.017	0.027	-0.018***
	(0.006)	(0.008)	(0.014)	(0.014)	(0.018)	(0.007)
1-Year Lag Dummy	0.019**	0.021	0.018	-0.008	-0.043**	0.021**
	(0.007)	(0.011)	(0.010)	(0.010)	(0.018)	(0.010)
2-Year Lag Dummy	0.000	0.001	-0.007	0.043**	0.032	-0.004
	(0.006)	(0.009)	(0.009)	(0.019)	(0.018)	(0.010)
3-Year Lag Dummy	0.002	-0.000	0.013	0.003	-0.017	0.014
	(0.008)	(0.009)	(0.009)	(0.017)	(0.015)	(0.012)
4-Year Lag Dummy	0.002	0.001	-0.009	0.022**	-0.009	0.005
	(0.006)	(0.007)	(0.008)	(0.010)	(0.012)	(0.006)

Notes: robust standard errors corrected for within-state correlation in the error term are reported in parentheses. Each specification controls for state fixed effects, year fixed effects, various covariates and a set of state specific linear time trends.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

Table B5. Alternative Interaction Specification: the Relationship between National-Standard Laws and Various Health Care Quality Metrics, Interacting Low-Quality Indicator with National-Standard Rule Indicator

	(1)	(2)	(3)
Panel A. Dependent Variable: Inpatient Mortality Rate for Selected Conditions (Logged)			
National-Standard (NS) Law Dummy	0.056**	0.048**	0.095*
	(0.026)	(0.023)	(0.049)
NS Law * Initially Low Quality State	-0.132***	-0.123**	-0.197**
	(0.045)	(0.051)	(0.096)
Panel B. Dependent Variable: Low-Discretionary Avoidable Hospitalization Rate (Logged)			
National-Standard (NS) Law Dummy	-0.014	-0.019	-0.031
	(0.051)	(0.044)	(0.029)
NS Law * Initially Low Quality State	-0.530***	-0.438***	-0.318***
	(0.091)	(0.085)	(0.067)
N			
Panel C. Dependent Variable: Maternal Trauma Rate (Logged)			
National-Standard (NS) Law Dummy	0.129*	0.079	0.112
	(0.073)	(0.108)	(0.113)
NS Law * Initially Low Quality State	-0.527***	-0.385**	-0.410*
	(0.129)	(0.169)	(0.231)
N			
Panel D. Dependent Variable: Preventable Delivery Complication Rate (Logged)			
National-Standard (NS) Law Dummy	0.074	-0.035	0.069
	(0.048)	(0.051)	(0.076)
NS Law * Initially Low Quality State	-0.479***	-0.386***	-0.491***
	(0.099)	(0.101)	(0.125)
N			
Control Variables?	NO	YES	YES
State-Specific Linear Trends?	NO	NO	YES

Notes: robust standard errors corrected for within-state correlation in the error term are reported in parentheses. All regressions include state and year fixed effects. Regressions in Panel A are weighted by the number of admissions (for the relevant state and year) in the sub-sample of discharges associated with the relevant selected conditions (e.g., acute myocardial infarction). Regressions in Panel B are weighted by the low-variation health index (i.e., the sum of discharges for acute myocardial infarction, stroke, hip fracture or gastrointestinal bleeding) associated with each state-year cell. Regressions in Panels C and D are weighted by the number of deliveries associated with the relevant state-year cell. Inpatient mortality rates are risk-adjusted for the incidence (among the sub-sample) of each of the conditions comprising the sub-sample of selected conditions. Each regression also includes a separate dummy variable indicating whether the state has an initially below-average rate of the relevant quality metric (coefficient omitted).
Source: 1977 – 2005 National Hospital Discharge Surveys.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

Additional Specification Checks

Construction of Avoidable Hospitalization Rates. The results presented in the text are robust to alternative constructions of the AH Rates, including those constructions that (1) flag avoidable hospitalizations using any diagnosis field, not just the primary diagnosis field, (2) normalize avoidable hospitalization counts by the number of deliveries of children in the associated state-year cell (an alternative measure of the size of the cell that is not itself subject to influence by the prevailing liability environment), (3) normalize avoidable hospitalization counts by the number of acute myocardial infarctions in the associated state-year cell (rather than the low-variations health index that likewise includes strokes, hip fractures and gastro-intestinal bleedings), (4) use non-logged AH rates as the dependent variable and (5) focus only on the adult (18-plus) population. These results are available upon request from the authors.

Construction of Inpatient Mortality Rate for Selected Medical Conditions. The results presented in the text are robust to alternative constructions of the inpatient mortality rate for selected medical conditions, including those constructions that (1) use non-logged mortality rates as the dependent variable, (2) specify the outcome variable as the incidence of mortality out of an individual sample of admissions for the selected medical conditions (as distinct from the primary specification whose unit of observation is a given state-year cell), (3) use mortality rates as the dependent variable that are not risk adjusted for fluctuations in the state-year incidence of the underlying medical conditions, but instead include as covariates the incidence of such conditions, and (4) focus the analysis only on the adult population. These results are available upon request from the authors.

Note that the unit of observation in the inpatient mortality rate specification estimated in the text is a given state-year cell. In an alternative approach (not shown), we estimate linear probability models where the unit of observation is an individual discharge within the sample of inpatient admissions associated with the selected conditions (e.g., acute myocardial infarctions, strokes, etc.) and where the dependent variable is an indicator for inpatient mortality (in such models, we include controls for the incidence of the relevant conditions). The results from this alternative approach are (perhaps not surprisingly) nearly identical to those of the state-year specifications estimated in Table 4 in the text. In alternative specifications, we likewise take an individual discharge approach for the obstetrics analysis and derive essentially identical results.

Cancer Screening / Damage-Cap Results. The cancer screening results presented in Table 5 of the text are robust to a number of alternative formulations of the relevant cancer screening measures, including alternative formulations of the age restrictions (e.g., those 40 – 75 years old in the case of proctoscopic examination, instead of 50 – 75) and alternative framing of the frequency of the screening—that is, using all of the frequency formulations provided by the BRFSS (e.g., annual, every 2 years, every 5 years, etc.). In the interests of brevity, we do not present the full extent of these alternative formulations, though they are available upon request from the authors. We do, however, present in the following table results (analogous to those from Table 5 in the text) using the incidence of ever having had the relevant screening test as the operable dependent variable.

Table B6. Relationship between Remedy-Centric Tort Reforms and Cancer Screening Rates. Alternative Formulation: Incidence of Ever Having the Indicated Screening

	(1)	(2)	(3)	(4)	(5)	(6)
	MAMMOGRAM	PHYSICAL BREAST EXAM	PROCTOSCOPIC EXAM	PSA TESTING	DIGITAL RECTAL EXAM	PAP SMEAR
Non-Economic Damage Cap	0.008* (0.005)	-0.000 (0.003)	-0.004 (0.007)	-0.004 (0.005)	0.008 (0.008)	-0.002 (0.002)
95% Confidence Band for Coefficient of Non-Economic Damage Cap Variable	[-0.001, 0.019]	[-0.006, 0.006]	[-0.018, 0.010]	[-0.015, 0.007]	[-0.009, 0.025]	[-0.007, 0.002]
95% Confidence Band, scaled by mean screening rate	[-0.001, 0.026]	[-0.010, 0.010]	[-0.045, 0.025]	[-0.028, 0.013]	[-0.018, 0.050]	[-0.011, 0.004]
N	1010415	1156433	849445	252313	341102	1664055

Notes: robust standard errors corrected for within-state correlation in the error term are reported in parentheses. All regressions included state and year fixed effects.
Source: 1987 – 2008 Behavioral Risk Factor Surveillance System Records.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

Table 5 in the text presents results from the basic difference-in-difference specification without the various control variables included. Table B4 above, which includes a full set of leads and lags of the damage-cap variable, presents results from specifications that include a range of covariates (as set forth in Online Appendix B above) along with a set of state-specific linear time trends.

Randomization Inference. Following Frakes (2013), we also endeavored to take an alternative route towards estimating the standard errors associated with our estimates. Accordingly, using the sample of observations from our control states, we simulate a set of placebo laws that match the distribution of timing of actual reforms. We then estimate the association between the relevant quality indicator and the placebo laws, replicating this process 5,000 times. We then observe where the actual coefficient from our primary specifications falls in the distribution of coefficients generated through these simulations. Due to time limitations on our use with the data at the NCHS's Research Data Center, we have only performed this exercise on the liability standards analysis for the inpatient mortality rate for selected medical conditions. In the text, we demonstrate that such mortality rates fall by 7.6 percent—representing an improvement in quality—upon a national standard adoption in those treatments states that begin the sample with above-average inpatient mortality rates—i.e., in those states with initially lower-than-average quality. While we find that this estimate is statistically significant at the 5 percent level in the main text, we find that this estimate is only significant at the 10 percent level through this randomization inference approach (the estimated -7.6 coefficient falls within the bottom 4.5th percentile of this simulated distribution).

Tort-Law Generally Damage Caps

Damage-cap adoptions in many states applied to tort cases broadly, not simply those pertaining to medical malpractice. Damage-cap adoptions in other states applied only to medical malpractice situations. General tort-law caps are arguably likely to pose fewer legislative endogeneity concerns. As such, in other specifications, we replicate the damage-cap analysis by codifying caps using only those adoptions that apply to tort laws more broadly, dropping those states from the analysis that adopted caps in malpractice-specific contexts. If anything, the results of this alternative analysis suggest an even more modest decrease in health care quality connected with damage cap adoptions. For instance, in the case of avoidable hospitalization rates, the coefficient of this modified damage-cap variable is -0.03, with a 95 percent confidence interval of [-0.08,0.02]. In the case of inpatient mortality rates for selected medical conditions, the coefficient is -0.04, with a 95 percent confidence interval of [-0.11, 0.04]. The full set of results for this alternative approach are available upon request.

Cancer Screening Liability Standards Analysis.

As stated in the text, data is available for cancer screening rates over a period of time in which only 3 states modified their standard of care rules: Delaware, Indiana, and Rhode Island. Moreover, only with respect to mammography and pap testing is data available over the full BRFSS period, facilitating any ability to draw upon the experiences of these three treatment states and to properly test for pre-period trends. A further difficulty comes with the fact that even fewer treatment states are available to test the main hypothesis of interest—i.e., that quality will rise in connection with national standard adoptions among those states that begin the sample period with initially low-levels of quality. With respect to mammography, only Indiana is available as a treatment state by which to test this hypothesis. With respect to pap testing, both Indiana and Rhode Island are available for such purposes. While the results of this exercise are arguably unreliable with such few treatment states, we nonetheless present results estimating the relationship between national standard adoptions and the incidence of mammogram screening and pap testing in those states that began with lower than average screening rates and thus with respect to which national standard adoptions arguably represent a heightening of expectations.[12] In Table B6, we demonstrate how these findings are impacted by (1) the inclusion of the relevant set of covariates discussed in Online Appendix A, (2) the inclusion of state-specific linear time trends and (3) the inclusion of a set of leads and lags of the national standard variable. Note that the analysis below only includes 3 lead periods considering that there are not enough years between the beginning of the sample and Indiana's essential reform to facilitate the estimation of a 4-year lead period.

The findings weakly demonstrate that when liability standards change so as to arguably require a heightening of standards, cancer screening rates increase. In the case of mammography screening, rates generally increase subsequent to the reform, strongest with a long lag. However, mammography screening also spiked strongly with a 2-year lead creating some concerns that the increase in quality may reflect a trend that pre-dated the reform. Of course, the 1-year lead coefficient does not support any such trend. Pap testing likewise suggests an increase in screening rates with a long lag, while also raising a concern of a pre-period trend, with a strong increase in rates occurring in the year prior to the reform. While this may in part be a reflection of an anticipation effect (Malani and Reif 2012), it may also be reflective of some external factor that correlates (perhaps spuriously) with the increase in screening and with the adoption of the liability reform.

[12] We focus here on estimating the impact of heightened liability standards as opposed to diminished standards. Estimation of this latter type of variation in the law is also compromised by such few treatment groups. Nonetheless, results of this alternative exercise are available upon request. If anything, the results actually suggest that screening rates also increase slightly upon national standard adoptions in those 1-2 states that adopt such reforms when they arguably entail a slackening of standards.

TABLE B7 The Relationship between National-Standard Laws and the Incidence of Cancer Screening

	(1)	(2)	(3)	(4)	(5)	(6)
	MAMMOGRAM SCREENING			PAPSMEAR SCREENING		
National Standard Law						
3-Year Lead Dummy	-	0.012**	0.014*	-	0.038	0.010
		(0.005)	(0.007)		(0.029)	(0.012)
2-Year Lead Dummy	-	0.046***	0.048***	-	-0.017	-0.014
		(0.005)	(0.006)		(0.014)	(0.015)
1-Year Lead Dummy	-	-0.020*	-0.000	-	0.029*	0.023*
		(0.011)	(0.009)		(0.015)	(0.012)
Contemporaneous	0.040***	.0140944	-0.001	0.045	-0.007	0.005
Dummy	(0.006)	.0102652	(0.009)	(0.033)	(0.020)	(0.027)
1-Year Lag Dummy	-	-0.003	0.025**	-	-0.003	0.013
		(0.023)	(0.12)		(0.013)	(0.013)
2-Year Lag Dummy	-	0.019	0.006	-	-0.014	-0.009
		(0.018)	(0.014)		(0.011)	(0.012)
3-Year Lag Dummy	-	-0.036*	-0.030	-	0.004	-0.002
		(0.021)	(0.023)		(0.020)	(0.021)
4-Year Lag Dummy	-	0.030***	0.025**	-	0.022*	0.018
		(0.007)	(0.012)		(0.011)	(0.014)
N	631592	520955	520955	1098595	912364	912364
Control Variables?	NO	YES	YES	NO	YES	YES
State-Specific Linear Trends?	NO	NO	YES	NO	NO	YES

Notes: robust standard errors corrected for within-state correlation in the error term are reported in parentheses. All regressions include state and year fixed effects. The regressions also include a separate dummy variable indicating whether the state has an initially below-average cancer screen rate (coefficient omitted). Cancer screening data is from the BRFSS.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

Damage-cap Codification

In Table B8, we estimate specifications that take an alternative approach to the codification of the damage-cap incidence variable. While the malpractice literature customarily codifies damage-cap adoptions in a simple binary fashion (0/1), non-economic damage cap provisions, in fact, take on a range of forms across jurisdictions. For instance, California imposes a flat, nominal $250,000 cap on non-economic damages awards, while Wisconsin imposes a $750,000 cap. One might imagine that California's cap would entail a stronger reduction in liability pressure. Hyman et al. (2009) use closed-claims data from Texas during the period of time prior to the imposition of its non-economic damage cap (with information on the breakdown of economic versus non-economic damages associated with the claim) to simulate the potential impact of the various damage-cap provisions across the various states. More specifically, they simulate the percentage of a mean verdict that is reduced through the imposition of the various caps employed across states.

In the present analysis, we build on these preliminary efforts by Hyman et al. (2009) and use the results of this simulation exercise as the relevant damage-cap variable within the difference-in-difference specification, as opposed to the simple binary approach. In applying these simulated measures to each state-year cell, we appropriately adjust this simulated reduction to account for inflation in the case of those damage-cap provisions that do not tie their cap levels to inflation. Inspired by studies in public finance (Currie and Gruber 1996), this codification scheme provides an empirically-informed way to ensure the comparability of the legal modifications under investigation, effectively reframing the treatment of the law in terms of the common function provided by such laws (i.e., reducing awards), as opposed to some coarse measure of their existence.

The estimated mean coefficients from those specifications using this alternative codification of damage-cap variables do not differ substantially from those derived from the traditional binary approach. In the case of inpatient mortality rates for selected medical conditions, low-discretionary AH rates, maternal trauma rates and preventable delivery complication rates, such estimates suggest a 0.1, -7.0, -13.3, and -5.1 percent change in the respective quality indicator upon an increase from 0 percent to 100% in the simulated extent to which a damage cap reduces a jury verdict. These largely negative point estimates are also inconsistent with the expectation that reducing liability pressure through the imposition of a cap will lead to a decline in quality—i.e., an increase in these respective measures. As above, of course, these results are statistically insignificant and cannot rule out some degree of a positive association between these measures and the reduction in damage awards resulting from caps. The associated upper ends of the confidence intervals for these estimates suggest a 18.3, 2.6, 16.4, and 14.9 percent change respectively. While the upper bounds are larger than those for the traditional codification approach discussed above, bear in mind that these estimates are to be interpreted in terms of a shift in the law that leads to a full 100% reduction in malpractice verdicts.

Table B8: Relationship between Simulated Damage Cap Variable and Various Health Care Quality Metrics

	(1)	*(2)*	*(3)*	*(4)*
	INPATIENT MORTALITY RATE	LOW-DISCRETIONARY AVOIDABLE HOSPITAL-IZATION RATE	MATERNAL TRAUMA RATE	PREVENTABLE DELIVERY COMPLICATION RATE
Damage Cap Strength: Simulated Percentage Decline in Mean Verdict	0.001 (0.090)	-0.070 (0.047)	-0.133 (0.148)	-0.051 (0.099)
Collateral Source Rule Reform	0.002 (0.039)	-0.012 (0.028)	-0.043 (0.083)	-0.019 (0.059)
Punitive Damage Cap	0.056 (0.044)	-0.003 (0.027)	-0.103* (0.060)	-0.017 (0.058)
Joint and Several Liability Reform	-0.008 (0.042)	0.005 (0.043)	0.167 (0.100)	0.027 (0.062)
95% Confidence Band for Coefficient of Non-Economic Damage Cap Variable	[-0.180, [0.183]	[-0.166, 0.026]	[-0.430, 0.164]	[-0.251, 0.149]
F-Statistic (Malpractice Variables Jointly = 0)	0.04	0.66	1.46	0.15
Prob > F (p value)	0.99	0.620	0.23	0.96
N	1141	1177	1053	1083

Notes: robust standard errors corrected for within-state correlation in the error term are reported in parentheses. The regression in Column 1 is weighted by the number of admissions (for the relevant state and year) in the sub-sample of discharges associated with the relevant selected conditions (e.g., acute myocardial infarction). The regression in Column 2 is weighted by the low-variation health index (i.e., the sum of discharges for acute myocardial infarction, stroke, hip fracture or gastrointestinal bleeding) associated with each state-year cell. Mortality rates are risk-adjusted for the incidence (among the sub-sample) of each of the conditions comprising the sub-sample of selected conditions.

All regressions included state and year fixed effects, along with the relevant set of state-year controls and state-specific linear time trends.

Source: 1977 – 2005 National Hospital Discharge Surveys.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

Additional References

Carter, Mary. Variations in Hospitalization Rates among Nursing Home Residents: The Role of Discretionary Hospitalizations. Health Services Research 38, 1177-1206.

Cohen, Thomas. 2005. Punitive Damage Awards in Large Counties, 2001. Bureau of Justice Statistics Selected Findings.

Dafny, Leemore, and Jonathan Gruber. 2005. Public Insurance and Child Hospitalizations: Access and Efficiency Effects. Journal of Public Economics 89, 109-29.

Localio, A. Russell, Ann Lawthers, Troyen Brennan, Nan Laird, Liesi Hebert, Lynn Peterson, Joseph Newhouse, Paul Weiler, and Howard Hiatt. 1991. Relation between malpractice claims and adverse events due to negligence: results of the Harvard Medical Practice Study III. New England Journal of Medicine 325, 245-51.

Malani, Anup, and Julian Reif. Accounting for Anticipation Effects: An Application to Medical Malpractice Tort Reform. NBER Working Paper No. 16593.

Peters, Phillip. 2008. Resuscitating Hospital Enterprise Liability. 73 Missouri Law Review 369, 369-97.

Studdert, David M., Michelle M. Mello, Atul A. Gawande, Tejal K. Ghandi, Allen Kachalia, Catherine Yoon, Ann Louise Puopolo and Troyen A. Brennan. 2006. Claims, Errors, and Compensation Payments in Medical Malpractice Litigation. New England Journal of Medicine 354, 2024-2033.

Weissman, Joel S., Constantine Gatsonis, and Arnold Epstein. 1992. Rates of Avoidable Hospitalization by Insurance Status in Massachusetts and Maryland. JAMA 268, 2388-2394.

Zeiler, Kathryn, Charles Silver, Bernard Black, David Hyman, and William Sage. 2007. Physicians' Insurance Limits and Malpractice Payments: Evidence from Texas Closed Claims, 1990-2003. Journal of Legal Studies 36, S9-S45.