Essays on Provider Behavior in Health Care Markets

Essays on Provider Behavior in Health Care Markets

A dissertation presented

by

Matthew Aaron Fiedler

to

The Department of Economics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Economics

Harvard University

Cambridge, Massachusetts

August 2013

Dissertation Advisor: Professor David Cutler Matthew Aaron Fiedler

Essays on Provider Behavior in Health Care Markets

Abstract

This dissertation consists of three essays that investigate the determinants of hospital and physician decisions and the consequences of those decisions for social welfare.

Chapter 1 examines how physicians combine their own experience with evidence from the clinical literature to form beliefs about the probability of rare complications. Focusing on a particular obstetric context, I present event study evidence that physicians respond to their own idiosyncratic experience. I combine this event study evidence with an analysis of the available clinical literature in order to calibrate a model of physician learning. I find that the calibrated model can account for much of the cross-sectional variation in practice patterns in this setting and may be able to account for the dramatic changes in practice patterns in this setting over the last three decades.

Chapter 2 (which is co-authored with Amitabh Chandra) examines the welfare implications of hospital adoption of percutaneous coronary intervention (PCI) since the early 1990s. We find that PCI adoption over this period was approximately welfare neutral, but would likely have generated positive net benefits if utilization had been limited to high-benefit patients. We also find that hospitals' profits on high-benefit patients are not sufficient to cover the fixed costs of adoption. Achieving the efficient outcome would therefore require policy changes that both discourage utilization among low-benefit patients and allow hospitals to capture more of the surplus generated among high-benefit patients.

Chapter 3 exploits day-to-day fluctuations in patient demand to estimate the marginal return to increasing per-patient hospital resources. Such estimates can shed light on the benefits of additional hospital capacity and on theories about the cross-sectional pattern of productivity in medicine. I find that patients arriving when a hospital has a large number of patients in its intensive care unit are less likely to be admitted to intensive care beds and more likely to face delays in the receipt of

time-sensitive care. The hospital responses are consistent with a model in which hospitals ration care based on expected clinical benefit. There is no evidence that hospital congestion adversely affects health outcomes, but firm conclusions will require more data.

Contents

List of Tables

List of Figures

Acknowledgments

This dissertation would not exist without the advice, insight, and encouragement of my advisers David Cutler, Amitabh Chandra, and Edward Glaeser. All of them gave generously of their time and taught me tremendous amounts. David can take credit for several of the most interesting findings herein, which were uncovered only by virtue of his insistence that I persevere in cases where I was convinced the well had run dry. Amitabh was a wonderful collaborator for a novice researcher. I have learned much from his knack for driving toward what is most interesting and important in a set of results and have been gratified by his willingness to take my own insights and contributions seriously. Ed has not only shaped this work, but also shaped my outlook on our field in ways large and small, all of which will make me a better economist in the years ahead.

Much of the data used in this dissertation was either obtained through or stored at the National Bureau of Economic Research (NBER), and virtually all of the data analysis was done on the NBER's servers. I am deeply grateful to the NBER as a whole and to the NBER's information technology team in particular for creating an excellent research computing environment. Special thanks go to Jean Roth and Mohan Ramanujan, who assisted me time and again with installing software packages, troubleshooting server problems, navigating the NBER's extensive data collection, and making security arrangements for restricted data. I am also indebted to NBER's Janet Stein and Alterra Milone for their assistance in navigating the application processes by which I obtained the data analyzed in the final chapter of this dissertation.

A number of staff at Harvard have helped me during my time as a graduate student. Brenda Piquet, and, previously, Nicole Tateosian helped keep me on the right side of the University's various administrative requirements. Clare Dingwell was particularly helpful in setting up meetings with David. On many occasions, she went the extra mile to find a way to slot me in even when my scheduling constraints gave her little to work with.

Classmates, friends, and family have been endlessly supportive throughout my graduate studies, and I appreciate it more than they know. There is too little space to name everyone here, but several people merit special mention. Greg Leiserson has been a fellow traveler since college as we ventured first to Washington, then to Cambridge, and now back once again to Washington. His insight improved this work greatly, and his humor kept me sane. My parents, David and Nancy

Fiedler, and my sister, Lauren Fiedler, have been the best cheering section I could ask for. Even when they found the research process somewhat mystifying, they did a wonderful job of feigning interest in and enthusiasm for whatever was currently on my plate. Finally, I thank my wife, Aviva Aron-Dine, for the countless insights she contributed to this work and also for her unfailing love and support, which was offered even when I failed to express my gratitude quite the way I should have. Thank you, my love.

There are many others who made important contributions to the individual chapters of this dissertation. For chapter 1, I thank Ariel Dora-Stern, Avi Feller, Nate Hilger, Mark Shephard, participants in the Harvard Labor/Public Finance Lunch, and participants in the Harvard Health Care Policy Health Economics Seminar for helpful comments and discussions. I thank Dmitry Dukhovny, Sarah Little, and Irving Olender for sharing their perspectives as physicians. I gratefully acknowledge support from the National Institute on Aging (grant number T32-AG00186).

For chapter 2, which is co-authored with Amitabh Chandra, Amitabh and I thank David Cutler, Maurice Dalton, Avi Feller, and Jonathan Skinner for helpful discussions. We both gratefully acknowledge funding from the National Institute on Aging: grant numbers P01 AG005842 and P01 AG019783 (Chandra) and T32-AG000186 (Fiedler).

For chapter 3, I thank Marcella Alsan, Raj Chetty, Maurice Dalton, Avi Feller, Seth Freedman, Joshua Gottlieb, Amanda Pallais, Mark Shephard, participants in the Harvard Labor/Public Finance Lunch, and participants in the Harvard Health Care Policy Health Economics Seminar for helpful comments and discussions. I thank Paul Smith of the Massachusetts Center for Health Information and Analysis (formerly the Division of Health Care Finance and Policy), Charlene Zion and Kevin Foster of the Massachusetts Department of Public Health, and Chris Campeau of Chloen Systems for their assistance in obtaining and linking the data used in the analysis. Note, however, that all interpretations of these data that I provide herein are my own and not necessarily those of the Center for Health Information and Analysis. I gratefully acknowledge support from the National Institute on Aging (grant number T32-AG00186) and the Lynde and Harry Bradley Foundation.

Chapter 1

Physician Learning and Physician Practice Patterns: Evidence from Obstetrics

Standard models of physician behavior posit that improving patient well-being is a major objective for physicians, either on its own (e.g. Chandra and Staiger (2007)) or in combination with other, typically financial, objectives (e.g. Ellis and McGuire (1986)). With a small number of notable exceptions (Frank and Zeckhauser, 2007; Dickstein, 2012), however, these models assume that the physician knows the medical production function with certainty. Such an assumption grossly simplifies the reality physicians actually face. In practice, physicians confront a formal clinical literature that is frequently incomplete and sometimes contradictory, and they must integrate this formal evidence with informal evidence from their own experience with patients. In this paper, I show that an explicit focus on physician learning provides important insights into physician behavior that models that abstract from physician learning cannot provide.

I examine physician learning in a particular medical context: the choice between elective repeat cesarean section and "trial of labor after cesarean" (TOLAC) for a pregnant woman with a prior cesarean delivery. This choice requires physicians and patients to trade off the reduced maternal morbidity associated with a successful "vaginal birth after cesarean" (VBAC) against the risk that attempting labor will lead to a rare, but serious complication known as uterine rupture. Rupture

occurs when the uterine wall tears along the scar left by the prior cesarean incision, and it can have serious consequences for both the mother and the baby, including the possibility of perinatal death. Two characteristics of this medical setting make it ideally suited to a study of provider learning. First, the debate among obstetricians over appropriate care for this category of patients focuses on a single, observable, high-salience risk of uncertain magnitude: uterine rupture. As stated in a recent editorial by the editor-in-chief of Obstetrics and Gynecology, a leading journal in the field, "VBAC is essentially a uterine rupture issue" (Scott, 2010). Second, any practicing obstetrician routinely attends both vaginal and cesarean deliveries, so changing practice style in response to new information requires no investment in new skills or equipment.

Meanwhile, this setting exhibits in microcosm two puzzling stylized facts from the general literature on physician practice. First, the share of women undergoing TOLAC varies widely across hospitals. After purging small-sample variation, the hospital at the 75th percentile of the TOLAC rate distribution had a rate at least 50 percent higher than the hospital at the 25th percentile of the distribution over the entirety of the period for which data are available. This mirrors a broad literature inside and outside of economics that documents large cross-sectional variations in physician practice across regions (Fisher et al., 2003), across hospitals and physicians within a region (Epstein and Nicholson, 2009), and even across physicians within a single hospital (Doyle et al., 2010). Second, TOLAC rates have followed a dramatic up-then-down path in recent decades, as shown in Figure 1.1. The VBAC rate in 1979 was virtually zero and had been for decades. By 1995, the VBAC rate had gradually risen to almost 30 percent, after which it reversed and gradually fell back below 10 percent by the mid-2000s. This pattern of slow diffusion, followed by slow "un-diffusion" of TOLAC is a case study of unresolved general questions about what forces drive diffusion of medical technology (Phelps, 2000; Skinner and Staiger, 2009).

In this setting, I develop and calibrate a model in which physician decisions are driven by beliefs about the risk of uterine rupture. In forming these beliefs, physicians rationally combine two types of information: (1) formal evidence on the risk of uterine rupture from the published clinical literature; and (2) informal experience with patients, their own and potentially those of colleagues. In this model, changes in average practice over time can arise either from changes in the published evidence or, to the extent that published evidence on the risk of rupture is inconsistent with the true risk, accumulating physician experience. This model can generate rich cross-sectional variation

in practice, arising both from systematic differences in the experience accumulated by physicians who face different risks of rupture (a mechanism related to that studied by Chandra and Staiger (2007)) and from random differences in experience among physicians facing a common risk.

I analyze this model empirically in three steps. I first use an event study design to estimate the effect of physicians' idiosyncratic experience with uterine rupture on practice patterns. I then examine what a rational physician should believe about the risk of rupture on the basis of the clinical literature; I use this analysis to gain insight into whether the event study results can be reconciled with rational learning behavior. Finally, I combine the event study results and evidence from the clinical literature to calibrate and simulate the learning model.

I start by using an event study design to examine how idiosyncratic experiences with uterine rupture affect the behavior of medical providers. In brief, I identify the effect of a uterine rupture event on subsequent practice by comparing the evolution of TOLAC rates at hospitals at which a patient experiences a uterine rupture event to hospitals in the same state of similar size at which no patient experiences such an event. I implement this approach in a database that contains the the universe of hospital discharges records for several large states over a combined 106 state-years stretching from 1993 to 2010. I estimate that delivering a patient who experiences uterine rupture during an attempt of labor immediately reduces the hospital's TOLAC rate by 0.5 percentage points, growing to 0.7 percentage points by 6 quarters later. After scaling this estimate up to account for measurement error, I estimate that experiencing a uterine rupture event reduces a hospital's trial of labor rate by 1.1 percentage points in the medium run. I then repeat the event study at the physician level. I find that only a small share of the hospital-wide response can be accounted for by the particular physician who experienced the event, which implies that the learning processes at work operate at the hospital level (or above). I also present evidence that the common trends assumption underlying the event study is valid, that the event does not lead to any important changes in hospital case mix, and that my estimate is far out in the tail of a distribution of "placebo" event study estimates, all of which support a causal interpretation of the results.

These event study results are consistent with research from other settings finding that physicians respond to their own experience. Choudhry et al. (2006) show that physicians reduce the rate at which they prescribe the blood-thinning drug warfarin to patients with atrial fibrillation after a patient experiences a possibly warfarin-induced bleeding complication. Dranove and Watanbe

(2009) show that obstetricians who are sued for malpractice modestly increase their cesarean section rates over the ensuing quarters. As discussed below, however, my interpretation of these responses differs sharply from that advanced by Choudhry et al. (2006).

As a byproduct of my event study, I also make a contribution to the econometrics literature on event studies. In particular, I derive an event study estimator that extends the conditional difference-in-differences framework of Abadie (2005) to the event study context. The resulting event study estimator is similar to that used by Hilger (2012) in his study of parental layoffs on child college enrollment, and the identification results in this paper provide a formal justification for his approach. Relative to the distributed lag regression estimators frequently used for event studies (e.g. Jacobson et al. (1993)), this estimator has several advantages, most importantly: the resulting estimates are interpretable as treatment effects on the treated, even in the presence of treatment effect heterogeneity; and there is no need for the various *ad hoc* approaches to handling units that experience multiple events commonly used in this literature (Sandler and Sandler, 2012).

In the next part of the paper, I turn to understanding what beliefs a rational physician should hold about the risk of uterine rupture based on the clinical literature. Via a comprehensive review of this literature, I identify more than 100 studies that report experience with trial of labor and uterine rupture. By analyzing this literature, I obtain an important insight: different studies provide very different estimates of the risk of rupture. After accounting for small sample variation, the cross-study standard deviation in the estimated risk is 70 log points. I then show that, if even one-quarter of this variation arises from true cross-provider heterogeneity in risk rather than study-specific measurement error, a rational physician should be quite uncertain about her own risk of rupture even after reviewing the entirety of the literature. Under that assumption, I calculate that the information encompassed by the entire literature available as of 2010 is equivalent to the information obtained from observing fewer than 1,400 attempts of labor by a single physician's patients, despite the fact that the full clinical literature encompasses more than 400,000 attempts of labor.

This fact has important implications for the interpretation of the event study results. In particular, using this estimate of the information contained in the clinical literature and the structure provided by the learning model, I calculate that a rational physician who has seen 1,000 deliveries by women with a prior cesarean section and experiences a uterine rupture event should update her

beliefs about the risk of rupture by 7.6 percent of the baseline risk of rupture. This change in the estimated risk of rupture is substantial, and it implies that rationalizing the event study estimate that a single uterine rupture event leads to a 1.1 percentage point reduction in TOLAC rates does not require absurd assumptions about the sensitivity of patient preferences to changes in risk. This conclusion is contrary to that of Choudhry et al. (2006), who argue that physician responses to personal experience of this kind are prima facie evidence of irrationality in physician learning.

I close my analysis of the learning model by calibrating it using the event study estimate and estimates of what a rational physician would believe based on the clinical literature. I then simulate the model in order to study its ability to explain the striking cross-sectional and time series variation in TOLAC rates discussed earlier. This analysis has three important findings.

First, I find that my event study estimate of how hospitals respond to idiosyncratic experience with uterine rupture implies the existence of substantial and long-lived cross-sectional variation in practice. Surprisingly, this can be true even if there is no true heterogeneity in the risk of rupture across hospitals. In particular, if physician learning occurs at the "community" level (defined, for these purposes, as among a group of four average-sized hospitals), idiosyncratic differences in experience with uterine rupture can explain at least one-quarter of the cross-hospital variation in TOLAC rates in all years through 2010. Meanwhile, if I assume that providers also differ in the their true risk of rupture, I can approximately match the overall level of cross-hospital variation in TOLAC rates.

Second, I find that the rational learning model cannot account for the time series pattern of TOLAC rates. While it can rationalize the rapid rise in TOLAC rates during the 1980s as being the result of accumulating medical evidence showing that the risk of uterine rupture was lower than previously thought, it cannot explain the late-1990s decline. Its failure in the late 1990s occurs because there is little change in the average message of the clinical literature over this period.

Third, however, I show that a modified version of the model in which physicians exhibit "selective memory" can explain the time series pattern. In particular, I modify the model so that physicians always remember attempts of labor that end in rupture but frequently forget attempts that end without rupture. In light of the severity and rarity of uterine rupture events, such a pattern is plausible if physicians assess their own experience using the "availability heuristic" of Tversky and Kahneman (1973). The fact that most hospitals institutionalize a focus on adverse

events through regular "morbidity and mortality" conferences could also help generate this pattern. Once extended in this fashion, the model can account for virtually all of the decline in TOLAC rates relative to their peak. Intuitively, with this modification, physicians now underweight successful instances of TOLAC and, as a result, gradually come to believe that rupture is a far more common outcome than it actually is, leading TOLAC rates to fall accordingly. While this finding suggests that Choudhry et al. (2006) may indeed have been correct that irrationality is an important feature of physician learning, the pattern of irrationality assumed here is precisely the opposite of that proposed by Choudhry et al. (2006). Choudhry et al. argue that physicians overweight their own experience with rare adverse events. In this model, physicians respond completely rationally to adverse events; the irrationality arises because they *under-react* to run-of-the-mill good outcomes.

In the final section of the paper, I examine an alternative theory for the evolution of TOLAC rates over the last three decades, which emphasizes the role of a small number of high-profile events: the publication of clinical guidelines by the American College of Obstetricians and Gynecologists (ACOG); and the publication of a small number of high-profile articles in major medical journals. The role of events of this kind has been emphasized in the medical and epidemiological literature by, for example, Zinberg (2001), Guise et al. (2010b), and MacDorman et al. (2011). To assess the role of these events, I undertake a regression discontinuity analysis around publication of eight ACOG guidelines and five highly-cited research articles to see what, if any, effect these events had on vaginal birth after cesarean (VBAC) rates. I find evidence that only one of these events, the publication of Lydon-Rochelle et al. (2001), had a significant effect on practice. Even this event, however, reduced the VBAC rate by only 2 percentage points, which is a tiny fraction of the overall change in practice patters over the period studied. I also examine whether any of these events causes a change in the trend (as opposed to the level) of the VBAC rate. While these analyses have more limited power, I find no evidence of such changes in trend. I conclude that high-profile events of this kind cannot account for the observed changes in practice.

The remainder of the paper proceeds as follows. Section 1 briefly introduces my medical setting. Section 2 presents the model of physician learning. Section 3 describes my data. Section 4 presents the event study results on the effect of uterine rupture on subsequent practice decisions. Section 5 examines the evolution of the clinical literature. Section 6 calibrates the learning model. Section 7 presents evidence on the effect of a small number of high-profile events, and the final section

concludes.

1.1 Medical setting

I explore the research questions of this paper in an obstetric context: the choice of an appropriate mode of delivery for pregnant women with a history of cesarean section. Women with a prior cesarean delivery have two options in subsequent pregnancies: delivery by elective repeat cesarean section; or a "trial of labor after cesarean" (TOLAC) with the goal of achieving a "vaginal birth after cesarean" (VBAC).[1] Successful vaginal delivery has the same advantages over cesarean delivery in this group as in women with no history of cesarean section; a cesarean section is major abdominal surgery and so typically entails a longer recovery and a greater risk of complications for the mother. On average, women with a prior cesarean delivery who attempt labor have a vaginal delivery roughly three-quarters of the time. However, for women in this group, labor also carries with it a 1 in 200 risk of a complication known as uterine rupture, which occurs when the uterine wall tears along the scar left by the prior cesarean incision. Women who deliver by elective cesarean section face a negligible risk of rupture, as do women without a history of cesarean delivery (Zwart et al., 2009). When uterine rupture occurs, it is an emergency and can have serious consequences, particularly for the neonate. Uterine rupture brings with it an approximately 1 in 20 risk of perinatal death and is believed to carry a considerably larger risk of serious neurological damage to the neonate, although the precise magnitude of this risk is uncertain.

The decision problem for patients with a prior cesarean section therefore trades off the risk of uterine rupture against the possibility that attempting labor will lead to a successful vaginal delivery. Different patients are likely to evaluate this tradeoff differently, due both to variation in maternal preferences over different birth outcomes (see, for example, Emmett et al. (2010) and Grytten et al. (2013)) and to variation in the relevant probabilities. There is, in particular, substantial cross-patient heterogeneity in the probability that an attempt of labor will lead to a vaginal delivery. Women with a prior vaginal delivery and a non-recurring indication for the prior cesarean section (e.g. breech presentation) are much more likely to have a successful VBAC, while

[1] Unless otherwise indicated, the ensuing discussion of the risk and benefits of these two delivery options is based upon the conference statement produced by a recent National Institutes of Health Consensus Development Conference on this topic (NIH, 2010) and the evidence report on which that statement is based (Guise et al., 2010b).

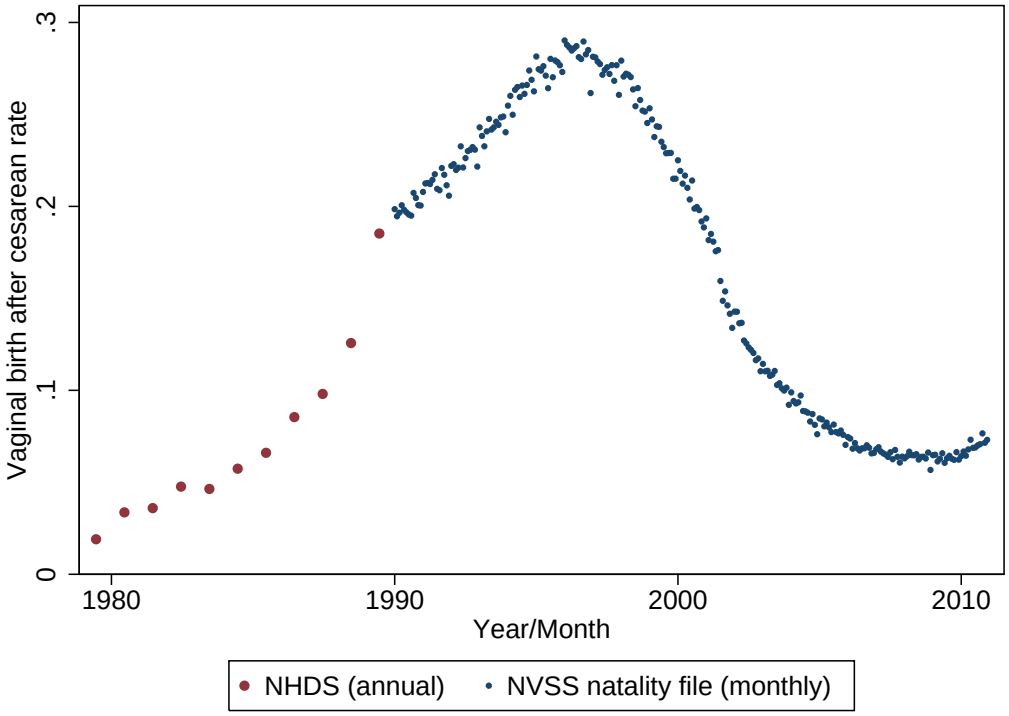
the success rate falls with maternal BMI, fetal size, and maternal age. White race and various indicators of higher socioeconomic status are also predictive of a successful VBAC. By contrast, there are few known predictors of uterine rupture. A short time since the prior cesarean section, multiple prior cesarean sections, and a type of prior uterine incision other than a low transverse incision all raise the risk of rupture, but these risk factors are relevant for only small groups of women. The only aspect of the management of labor known to have an effect on the risk of rupture is inducing labor; induced labor has a risk of rupture roughly twice that of spontaneous labor. It is not known whether the paucity of strong predictors of uterine rupture indicates that the risk is, in fact, relatively homogeneous across the population or instead simply a reflection of the limits of current knowledge.

Recent decades have seen dramatic swings in VBAC rates.[2] In light of the very high risk of uterine rupture associated with the "classical" cesarean incision then in use, obstetric practice in the early 20th century came to be dominated by the dictum "once a cesarean, always a cesarean," which is conventionally attributed to Cragin (1916). Cragin's dictum remained the dominant view until the final decades of the 20th century. As depicted in Figure 1.1, however, practice began to depart from this traditional view in the early 1980s and VBAC rates began to rise. Most accounts attribute this change in practice to the overall rise in cesarean section rates during the 1970s, which spurred interest in reducing overall cesarean delivery rates, while simultaneously focusing attention on appropriate care for the growing share of deliveries in which the mother had a history of cesarean section. Citing these trends in cesarean delivery and a limited body of evidence suggesting that VBAC was safe for many categories of women, the National Institutes of Health Consensus Conference on Cesarean Delivery held in 1980 endorsed wider adoption of VBAC (NIH, 1980). VBAC rates rose steadily over the ensuing 15 years, accompanied by additional evidence confirming that the risk of uterine rupture was modest and several rounds of guideline revisions by the American College of Obstetricians and Gynecologists (ACOG) that supported trial of labor for progressively broader patient populations.

The rise in VBAC rates abruptly halted and reversed itself in the mid-1990s. A variety of hy-

[2]The trends discussed below are discussed in greater detail in Flamm (1997), Zinberg (2001), Korst et al. (2011), and MacDorman et al. (2011). Unless otherwise indicated, the statements in this section are attributable to those sources.

Figure 1.1: Vaginal birth after cesarean rates over time



Notes: This figure plots the vaginal birth rate for deliveries in which the mother had a prior cesarean delivery. NHDS refers to for the National Hospital Discharge Survey. NVSS natality file refers to the National Vital Statistics System Natality Public Use File, which provides information derived from birth certificates for the universe of US births. Both data sources are described in detail in section 1.3. Data for 2003 and later are adjusted for the gradual adoption of the 2003 certificate of live birth as described in Section 1.3. Monthly data from the natality file are plotted such that the month YYYYM1 is plotted at YYYY.00. For visual consistency, annual data from the NHDS are plotted halfway between YYYYM1 and YYYYM12.

potheses for this reversal have been advanced. One hypothesis is that an increase in experience with uterine rupture – a necessary concomitant of rising VBAC rates – placed a damper on enthusiasm for VBAC, perhaps by increasing the perceived risk of rupture and thereby increasing the perceived level of monitoring appropriate for patients undergoing TOLAC (Pitkin, 1991), which would have increased the obstetrician's cost of attending a TOLAC.

Several other factors are also commonly cited as playing a role in the recent decline in VBAC rates. Many sources ascribe an important role to an increased salience of malpractice liability due variously to high profile judgments in cases of uterine rupture, the late-1990's "crisis" in malpractice premiums, and direct pressure from malpractice insurers to reduce VBAC rates. Also commonly cited in this literature are publication of certain high-profile research articles and the ACOG guideline revisions in October 1998 and November 1999, which placed a greater weight on the risk of rupture and recommended that hospitals offering VBACs should have anesthesia capabilities on hand at all times.

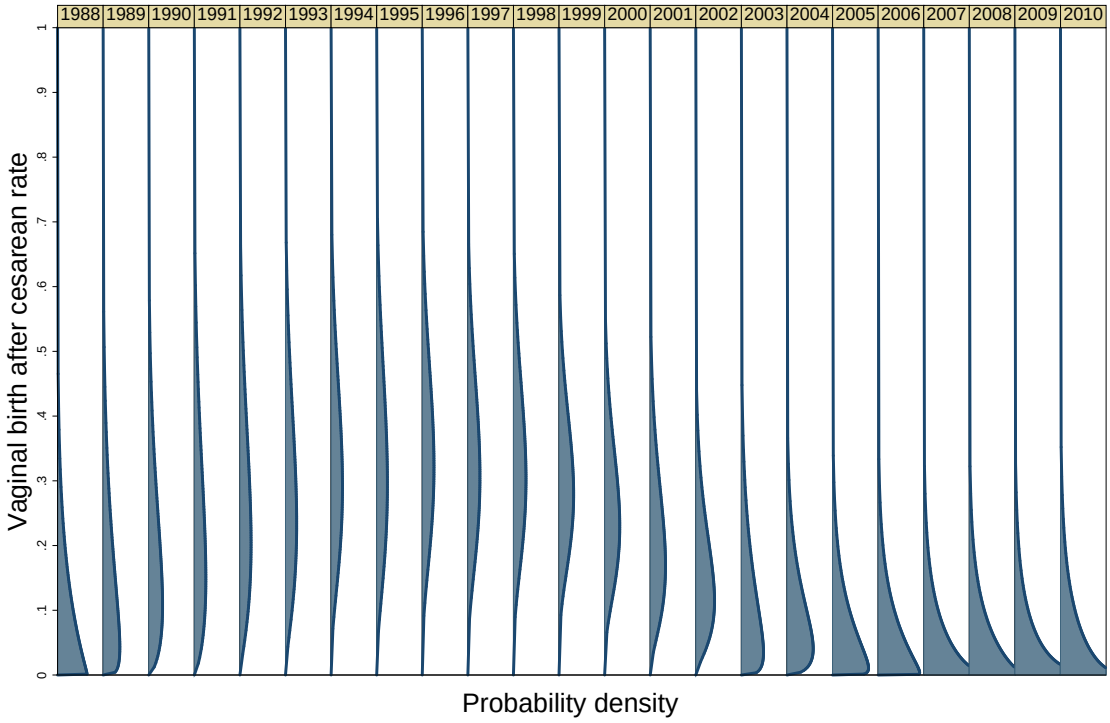
Variation in VBAC rates is not limited to the time series, and there is also considerable cross-sectional variation in VBAC rates. In order to obtain estimates of cross-sectional variation in VBAC rates that are purged of small-sample variation, I estimate a simple model in which each hospital has a "latent" VBAC rate drawn from an underlying beta distribution and its realized VBAC rate is the result of a binomial draw based on that underlying rate. The details of this beta-binomial mixture model are provided in Appendix A.1. I estimate the model using data from the Nationwide Inpatient Sample for 1988-2010 that are described in Section 1.3. Figure 1.2 depicts the estimated cross-hospital distribution of latent VBAC rates. It is evident that the cross-sectional variation is substantial in all years.

1.2 Learning model

I start my analysis by presenting a simple model of how medical providers (e.g. physicians and hospitals) form beliefs about the magnitude of the risk of uterine rupture and how those beliefs affect practice decisions.[3] Providers start with a common prior belief based on the published clinical

[3]Phelps and Mooney (1993) present a learning model that is similar in some respects to the model presented here, notably its use of a beta-binomial learning rule. Their focus is on understanding how new physicians learn the practice norms of their local community, not how physicians learn about the productivity of alternative treatment

Figure 1.2: Distribution of latent hospital VBAC rates by year



Notes: This figure plots the distribution of latent hospital VBAC rates obtained from fitting a beta-binomial mixture model separately for each year in the Nationwide Inpatient Sample (NIS). The NIS is described in Section 1.3, and the beta-binomial model used to estimate these distributions are described in detail in Appendix A.1.

literature and gradually accumulate additional information from their own direct experience with patients or from the experience of colleagues. I suppose that time evolves continuously and that the common prior at time t can be summarized by a beta distribution with parameter vector $(\alpha(t), \beta(t))$ with $\alpha(t), \beta(t) > 0$. The uncertainty in this prior may reflect uncertainty about the population distribution of risks, the provider's position within that population distribution, or both. I denote the number of patients who provider h has observed attempting labor at time t by $L_h(t)$ and the number of uterine ruptures observed by provider h at time t by $R_h(t)$.

Because the beta distribution will play an important role in what follows, it is worth a brief digression to review its properties. The beta distribution is the conjugate prior distribution for the Bernoulli distribution. In particular, starting from a prior belief that the probability of some event is represented by beta distribution with parameters (α, β), the posterior belief after observing a single random realization X of that event is a beta distribution with parameters $(\alpha+X, \beta+1-X)$. Since the uniform distribution is a beta distribution with $\alpha = \beta = 1$, this conjugacy property means that the beta distribution with parameter vector (α, β) can be interpreted as the posterior distribution arising from observing $\alpha - 1$ "hits" and $\beta - 1$ "misses" starting from a state of perfect ignorance. For this reason, I will refer to the sum $\phi = \alpha + \beta$ as the "notional sample size" of the corresponding beta distribution. The beta distribution has mean $\mu = \alpha/(\alpha + \beta)$ and variance $\mu(1-\mu)/(\phi+1)$; note that the variance goes to zero as $\phi \to \infty$. It will frequently be convenient to work in terms of the alternative parametrization (μ, ϕ), and I will switch back and forth freely.

I suppose that each provider faces a true risk of rupture p_h drawn according to some distribution function $G(p)$ and that rupture outcomes are independent across births within a provider (conditional on p_h).[4] Assuming that the provider combines information using Bayes rule, the conjugacy property described above implies that the beliefs of provider h at time t follow a beta distribution with parameter vector $(\alpha(t) + R_h(t), \beta(t) + L_h(t) - R_h(t))$. The posterior mean, which I denote $\tilde{p}_h(t)$, therefore takes the intuitive form

$$\tilde{p}_h(t) = \frac{\alpha(t) + R_h(t)}{\alpha(t) + \beta(t) + L_h(t)}, \tag{1.1}$$

options.

[4]The population distribution of risks need not coincide with the providers' prior beliefs, as the evidence on which those beliefs are based may be flawed or simply include some sampling error.

which combines the prior information with the provider's own experience. Observe that the occurrence of a rupture at any time before t increases the estimated risk of rupture at time t by an amount

$$\tilde{\Delta}_h(t) = \frac{1}{\alpha(t) + \beta(t) + L_h(t)}, \tag{1.2}$$

relative to the counterfactual in which that delivery proceeded without rupture. Equation (1.2) shows that the effect of an earlier event on beliefs at time t is inversely proportional to the sum of the notional sample size of the common prior and accumulated experience. Since this formula applies to events occurring at *any* time before t, it implies that the effect of any given event diminishes only gradually as the provider acquires additional experience or as additional clinical evidence increases the precision of the common prior.

Patients with a prior cesarean delivery arrive according to a Poisson process with continuously-varying rate $\lambda(t)$, and I assume $\lambda(t) \geq \underline{\lambda} > 0$ for all t. I assume that the provider truthfully reports her current estimate of the risk of rupture to each patient. Each patient has a distinct threshold risk p^* such that the patient prefers to labor if $\tilde{p}_h(t) < p^*$ and elects repeat cesarean delivery otherwise.[5] The distribution of thresholds is identical across providers and follows a cumulative distribution function $F(p^*)$ with support on the full interval $[0, 1]$.[6,7] Per the discussion in the last section, variation in the risk threshold could arise from a variety of sources, including differences in maternal preferences over different birth outcomes or differences in the likelihood that an attempt of labor is successful. The trial of labor rate of provider h at time t is therefore given by $y_h(t) = 1 - F(\tilde{p}_h(t))$. In this model, therefore, a provider's current beliefs about the risk of uterine rupture is a sufficient statistic for the provider's practice style.

I turn now to understanding what the model implies about the core questions of interest in

[5]Note that, provided that the patient maximizes expected utility, it does not matter whether the provider reports only the mean of her posterior belief or the full distribution.

[6]The full support assumption ensures that even providers that get an arbitrarily long sequence of bad outcomes will never have their trial of labor rates drop all the way to zero. It is mathematically convenient and unlikely to be particularly restrictive in practice.

[7]The assumptions made here imply that the learning process is inefficient since the provider-patient dyad does not take account of the additional information generated by an attempt of labor when choosing a delivery mode. Optimal learning would involve requiring some patients who actually prefer cesarean delivery to undergo a trial of labor and to do so to a greater degree when uncertainty is greater. Experimentation of this kind is not the main focus on the, but would add substantial mathematical calculation for little gain for understanding the present questions. Frank and Zeckhauser (2007) and Dickstein (2012) explore these incentives for experimentation in detail in the context of anti-depressant prescribing behavior.

this paper: the sources of changes in practice patterns over time and of cross-provider variation in practice patterns at a point in time. Changes in practice patterns over time can come from two basic sources in the model: (1) changes the clinical literature that change the providers' common prior; and (2) experience with the individual patients. Proposition 1 formally characterizes the model's dynamics:

Proposition 1. *Suppose that the prior belief parameters $(\mu(t), \phi(t))$ are everywhere differentiable. Then the rate of change in the expected beliefs of provider h at time t, conditional on the provider's current experience vector and risk type, is*

$$\frac{d}{ds}\mathbb{E}[\tilde{p}_h(s)|R_h(t), L_h(t), p_h]|_{s=t} = \underbrace{\phi(t)\tilde{\Delta}_h(t)\mu'(t)}_{\textit{change in prior mean}} + \underbrace{(\mu(t) - \tilde{p}_h(t))\tilde{\Delta}_h(t)\phi'(t)}_{\textit{change in prior ``sample size''}} + \underbrace{\lambda(t)y_h(t)\tilde{\Delta}_h^{+1}(t)[p_h - \tilde{p}_h(t)]}_{\textit{experiential learning}}, \quad (1.3)$$

where $\tilde{\Delta}_h^{+1}(t) \equiv (\phi(t) + L_h(t) + 1)^{-1}$. The corresponding rate of change in the expected trial of labor rate is given by

$$\begin{aligned}\frac{d}{ds}\mathbb{E}[y_h(s)|R_h(t), L_h(t), p_h]|_{s=t} = &-f(\tilde{p}_h(t))[\phi(t)\tilde{\Delta}_h(t)\mu'(t) + (\mu(t) - \tilde{p}_h(t))\tilde{\Delta}_h(t)\phi'(t)] \\ &+ \lambda(t)y_h(t)[p_h\{1 - F(\tilde{p}_h(t) - \tilde{p}_h(t)\tilde{\Delta}_h^{+1}(t))\} \\ &\quad + (1 - p_h)\{1 - F(\tilde{p}_h(t) - (1 - \tilde{p}_h(t))\tilde{\Delta}_h^{+1}(t))\} - \{1 - F(\tilde{p}_h(t)\}],\end{aligned}$$

where $f(\cdot)$ is the density function associated with the distribution function $F(\cdot)$.

It is instructive to examine each term of equation (1.3) in turn. The first term captures the effect of changes in mean of the prior distribution. As the prior mean increases, the posterior mean increases as well; the factors multiplying $\mu'(t)$ indicate that this effect is more important when the notional sample size of the prior is large relative to total experience. The second term illustrates the effect of increasing the precision of the prior distribution, which is to "shrink" the distribution of beliefs toward the mean of the prior distribution. The third term captures the effect of experiential learning, which, on average, shifts the provider's beliefs toward the true risk; this effect is larger for providers with high trial of labor rates and when the patient arrival rate is larger.

While all three of these effects may drive changes in beliefs for any particular provider at a particular point in time, the second two will frequently average to zero across the population. In particular, the second effect will be zero in the aggregate if $\mathbb{E}[\tilde{\Delta}_h(t) \cdot \{\tilde{p}_h(t) - \mu(t)\}] = 0$, while the third effect will be zero if $\mathbb{E}[y_h(t)\tilde{\Delta}_h^{+1}(t) \cdot \{p_h - \tilde{p}_h(t)\}] = 0$. Neglecting the covariance between the

terms in curly braces and the terms multiplying them, we see that these equalities will fail only when the common prior distribution is "mis-calibrated," that is, if the mean of the common prior differs from the the actual population average risk.

Turning to cross-sectional variation, one driver of variation in beliefs and practice is true dispersion in the risk of rupture. This source of variation in practice is similar to that studied by Chandra and Staiger (2007); here, the source of variation in productivity is underlying heterogeneity in the risk of rupture.[8] Proposition 2 demonstrates that, in the long-run, this is the only source of cross-sectional variation present in the model. Intuitively, providers ultimately learn their true risk levels and, as a result, the steady state distribution of practice patterns exactly follows the population distribution of risks.[9]

Proposition 2. *Suppose that the prior belief parameters* $(\alpha(t), \beta(t))$ *meet the following conditions:*

(i) $\phi(t) > \underline{c}$ *for some constant* $\underline{c} > 0$ *for all* t*;*

(ii) $\mu(t) < \bar{\mu}$ *for some constant* $\bar{\mu} < 1$ *for all* t*; and*

(iii) one of the following holds: (1) $\phi(t)$ *is bounded above; or (2)* $G(p)$ *places mass on a single point* p_0 *and* $\mu(t) \to p_0$ *as* $t \to \infty$*.*

Then $\tilde{p}_h(t) \xrightarrow{a.s.} p_h$ *as* $t \to \infty$*, and the steady state distribution of the trial of labor rate* y_h *has a distribution function* $G(F^{-1}(1-y))$*.*

In the short-run, however, there is a second potentially important source of cross-sectional variation: idiosyncratic differences in experience across providers. Unfortunately, analyzing the cross-sectional distribution of beliefs at any given point in time is complicated by the fact that providers that experience a larger-than-average number of ruptures will respond by making fewer attempts of labor. This behavior introduces correlation of unknown form between $R_h(t)$ and $L_h(t)$ that makes the model very difficult to analyze. To get around this problem, I instead study the cross-sectional distribution of beliefs among providers with the same cumulative number of attempts

[8] The role of productivity variation in this model obviously differs in important ways from Chandra and Staiger (2007). Notably, there are no productivity spillovers in this model. Also, crucially, in this model, providers learn their productivity from experience, so differences in productivity will come to affect behavior only gradually.

[9] The technical conditions on the belief parameters are straightforward. Condition (i) ensures that the prior provides at least a modest amount of information, and condition (ii) ensures that the prior does not imply that the risk of rupture is exactly one. Together with the full support condition on $F(p^*)$, these conditions ensure that providers will not get "stuck" at a trial of labor rate of zero. Condition (iii) ensures that the prior does not converge on a single value unless that value is the true population risk.

of labor, no matter the time at which the provider reaches that experience level.[10] Focusing on this alternative distribution eliminates the problematic dependence between $R_h(t)$ and $L_h(t)$. Nevertheless, this approach still generates important insights about the phenomenon of interest.

Assuming that the prior parameter vector $(\alpha(t), \beta(t))$ is constant as a function of time and letting $\tilde{p}_h(\ell)$ denote provider beliefs after ℓ attempts of labor, I compute that

$$\text{Var}(\tilde{p}_h(\ell)) = \left(\frac{1}{\alpha+\beta+\ell}\right)^2 [\underbrace{(\ell^2-\ell)\text{Var}(p_h)}_{\text{“fundamental” variation}} + \underbrace{\ell\mathbb{E}[p_h](1-\mathbb{E}[p_h])}_{\text{idiosyncratic variation}}]. \tag{1.4}$$

Both sources of cross-sectional variation appear in an intuitive form in equation (1.4); the first term captures variation due to true variation in the underlying risk, while the second term captures the variation that would exist even without any differences in fundamentals across providers. Quite intuitively, the amount of cross-sectional variation in beliefs arising from either of these sources is decreasing in the notional sample size of the common prior.

The two sources of cross-sectional variation exhibit differing dynamics over time. The “fundamental” variation grows monotonically and, consistent with the conclusion of Proposition 2, is equal to the variance of the population distribution of risks in the limit. By contrast, the idiosyncratic variation grows initially as providers gain experience but then begins to shrink once $\ell \geq \alpha + \beta$, that is, once the provider's own experience exceeds the notional sample size of the common prior. In the limit, therefore, this second source of variation vanishes. These long-run patterns may not, however, be an accurate representation of the relative importance of these two sources of variation in the short-run. In the present application $\mathbb{E}[p_h]$ is on the order of 0.01. If the coefficient of variation of the risk is 0.5, a relatively large figure in this context, then $\mathbb{E}[p_h](1-\mathbb{E}[p_h])$ will be approximately 400 times as large as $\text{Var}(p_h)$. This implies that, over the range of experience levels observed in my setting, which at most stretch into the thousands, idiosyncratic variation will be of a similar order of magnitude to variation due to dispersion in fundamentals.

As a final note, observe that equation (1.4) suggests that there will be an important interaction between the the degree of dispersion in fundamentals and the importance of idiosyncratic experiential differences in driving practice variation. In particular, it is reasonable to expect that, when

[10] The arguments given in the proof of Proposition 2 guarantee that every provider will eventually reach all levels of experience $\ell \in \mathbb{N}$, which ensures that this approach is well-defined.

$\text{Var}(p_h)$ is large, the prior information available to providers will reflect this fact, and so providers will have a dispersed prior distribution and $\alpha + \beta$ will be small. As a result, moving from a setting with a small amount of fundamental dispersion to one with a larger amount will often magnify both sources of dispersion, not just directly increase fundamental dispersion.

1.3 Data

The analyses presented in the remainder of this paper are built upon several different data sources. This section provides a brief description of each of these data sources.

1.3.1 State hospital discharge databases

For the event study analysis presented in the next section, I make use of the California Office of Statewide Health Planning and Development (OSHPD) Patient Discharge Databases for 1993-2010 and the Agency for Healthcare Research and Quality's Healthcare Cost and Utilization Project (HCUP) State Inpatient Databases (SID) for the following states and years: Arizona (1995-2010), Colorado (2003-2007), Florida (1997-2010), Maryland (1995-2010), Massachusetts (1999-2010), New Jersey (1995-2009), New York (1995-2009), and Washington state (1999-2009). These databases contain records for the universe of inpatient hospital discharges during the covered state-years. Each discharge record includes the full set of fields typically present on hospital discharge records, including ICD-9-CM diagnosis and procedures codes, quarter and year of discharge (or of admission in California), basic patient characteristics (like age, race, and insurance status) and – crucially for the event study analyses – hospital identifiers.[11] The SID databases for Arizona, Colorado, Florida, Maryland, New Jersey, and New York also report an encrypted identifier for the attending and (if applicable) the operating physician that is suitable for tracking physicians over time.[12]

[11] All SID databases report age in years. For years 1993 and 1994, California reports patient age as a range. In later years, California reports single year of age for approximately half of records, but reports age as a range or not at all on the remaining records for confidentiality reasons. For comparability across datasets, I impute the patient's age in years by using the median age for patients who fall in that age category for whom age in years is reported during that data year. For 1993 and 1994, I impute based on the single-year age data for 1995.

[12] Unfortunately, the physician identifiers are not comparable over the full period that they are reported. A change in the encryption algorithm HCUP used to encrypt the physician identifiers causes a break in comparability for all states between 2002 and 2003. State-level encryption or data collection changes create several more state-specific breaks. In addition, the physician identifiers for Florida in 1997 and Maryland in 2009 appear to have been corrupted during data processing and thereby rendered unusable. As a consequence, at the physician level, each state reporting

I identify deliveries and determine whether the mother had a prior cesarean delivery using the appropriate ICD-9-CM diagnosis codes. The ultimate mode of delivery (vaginal or cesarean) is identified by the presence or absence of procedure codes for a cesarean section.[13] Among women who deliver by cesarean section, I ascertain whether the woman labored using a method proposed by Henry et al. (1995) and Gregory et al. (2002). Their method codes a cesarean delivery as having been preceded by labor if the delivery record reports diagnosis codes for fetal distress, cord prolapse, breech converted to vertex presentation, or certain other labor abnormalities. In the present application, I also code women as having labored if their record reports a diagnosis code for uterine rupture during labor, a code not included by Henry et al. (1995) and Gregory et al. (2002); this code is discussed in greater detail below. The method proposed by Henry et al. (1995) and Gregory et al. (2002) has been validated against data from medical records in a related context (identification of elective primary cesarean section) and found to be highly accurate (Korst et al., 2004). It has been applied previously in the economics literature by Epstein and Nicholson (2009) and Epstein et al. (2010).

Uterine ruptures are identified by the presence of the diagnosis code for uterine rupture during labor (665.1x). While uterine rupture is the main diagnosis falling under this code, Weiss et al. (2000) note that the ICD-9-CM index also directs use of this code for various "incidental" uterine injuries that can occur during a cesarean delivery. Based on an examination of medical charts for Massachusetts deliveries during the years 1990-1997 in which the discharge record reported this code, Weiss et al. (2000) confirm that 665.1 is almost always used for one of these two reasons.

Understanding the relative prevalence of the two uses of this code is important to interpreting the event study estimates. To estimate this relative prevalence, I exploit the dramatic time series variation in trial of labor rates over the period studied.[14] If non-rupture uses of the code are rare,

physician identifiers is broken into 3-5 sub-panels.

[13]Specifically, delivery records are identified by presence of a diagnosis code of the form V27.x in any field. Women with a history cesarean section are identified by the presence of the diagnosis code 654.2 in any field. Cesarean delivery is identified by the presence of a procedure code of the form 74.x in any field.

[14]Weiss et al. (2000) also report an estimate that half of uses of 665.1 corresponded to true ruptures. However, their sample included all deliveries (not just those by women with a history of cesarean section) and examined a period and location with a trial of labor rate differing from the current sample. Since the relative probability of rupture and non-rupture uses of the code is a function of prior cesarean status and delivery mode, this estimate is of limited use for the present study.

then as trial of labor rates fall, the prevalence of code 665.1 should fall approximately proportionally. In contrast, if non-rupture uses are common, then the prevalence of code 665.1 should fall less and potentially even rise. To operationalize this insight, I develop and estimate a statistical model for the use of code 665.1, which I describe in detail in Appendix A.6. The model implies that the "true positive rate" – the share of uses of code 665.1 that correspond to uterine ruptures – is 65 percent. Because the non-rupture uses of the code are unlikely to reflect events that will have a substantial effect on subsequent physician behavior, this estimate implies that the event study estimates presented later in the paper should be scaled up by a factor of 1.5 ($\approx 1/0.65$).

1.3.2 National Vital Statistics System Natality Public Use File

For analyses examining national trends in VBAC rates after 1990, I turn to the National Vital Statistics System (NVSS) Natality Public Use Files for 1990-2010. Each year's file contains de-identified records for the universe of births occurring in the United States during that year. The data are compiled from states' birth registries by the National Center for Health Statistics. Each individual record reports a wide variety of medical and demographic information. Most important for the current analyses, the fields include information on the mother's childbearing history (including whether she has had a prior cesarean delivery), the mode of delivery, and the year and month of the birth. It is, unfortunately, not possible to observe in these data whether a cesarean section was preceded by labor, so analyses using these data are restricted to examining trends in the VBAC rate, rather than both the VBAC rate and the TOLAC rate.

The NVSS data have one important downside for examining trends in VBAC rates in the mid-2000s. Starting in 2003, states began adopting a revised Standard Certificate of Live Birth that changed the wording and format of the questions used to report mode of delivery and the mother's prior cesarean status. These changes appear to have caused slight increases in reported VBAC rates in the NVSS natality file (CDC, 2012). To address this problem, I exploit the fact that adoption of the revised certificate was staggered across states and estimate a simple difference-in-differences model of the effect of the certificate version used on the reported VBAC rate. In estimating this specification, I can use data only from 2004 and earlier because the public use file does not report state of birth after 2004. During this period, eight states plus part of New York adopted the revised certificate; these areas together account for approximately one-fifth of births nationwide

(CDC, 2006). The resulting estimates demonstrate that adoption increased the reported VBAC rate by 2.3 percentage points. I use this estimate to make an appropriate adjustment in analyses that use data from these years.

1.3.3 National Hospital Discharge Survey

Because the birth certificate in use in the United States did not report information on the mode of delivery or the mother's prior cesarean section status prior to 1990, I cannot use the NVSS data for analyses that require information on practice trends prior to that year. I turn instead to the National Hospital Discharge Survey (NHDS), which is available for the years 1979-2010.[15] The NHDS reports a sample of inpatient discharge records from a national stratified random sample of hospitals. The NHDS provides sampling weights to compute nationally-representative estimates.

Each discharge record in the NHDS includes most of the information included on the records in the state discharge databases described above. As such, I identify deliveries and determine delivery mode, maternal prior cesarean status, and whether a mother labored using the same algorithms used for the state discharge databases.

1.3.4 Nationwide Inpatient Sample

To characterize the degree of cross-hospital variation in VBAC and TOLAC rates at the national level, I use the HCUP Nationwide Inpatient Sample (NIS) for 1988-2010. The state inpatient databases are not available in all years or for all states, and the NHDS and NVSS do not report hospital identifiers, making them unsuitable for such analyses. The NIS contains the universe of inpatient discharge records from a sample of hospitals that aims to approximates a 20-percent stratified random sample of all hospitals in the United States. Because not all states participate, the resulting sample cannot be fully nationally representative, and HCUP provides sampling weights intended to account for the differences between the states included in the NIS states and the nation (HCUP, 2012). In addition, HCUP recommends caution in interpreting trends for the years 1988-1993, when the number of states participating was relatively small and growing rapidly (HCUP, 2006). In practice, however, it appears that these concerns about incomplete state coverage are not

[15] For years 1979-2006, the NHDS data files were obtained from the the Inter-university Consortium for Political and Social Research, study number 24281.

relevant to the present analyses. Annual VBAC rates computed using the NIS are very similar to those computed using the NHDS, which, as noted above, is nationally representative in all years.

Each discharge record in the NIS reports the same information that appears on the records in the state discharge databases described above, including a hospital identifier.[16] As such, I identify deliveries and determine delivery mode, maternal prior cesarean status, and whether a mother labored using the same algorithms used for the state discharge databases.

1.3.5 Database of VBAC/TOLAC publications

For analyses aimed at understanding how providers' prior beliefs about the incidence of uterine rupture should have changed over time, I build a comprehensive database of publications on the topic. I started by identifying English-language articles that use the terms "vaginal birth after cesarean," "trial of labor after cesarean," "uterine rupture," or their variants in the Thomson Reuters Web of Science database. The precise set of search terms used is adapted from those used by Guise et al. (2010b) in their comprehensive review of this literature; I report the the full set of search terms in Appendix A.2.

The initial search identified 1,159 articles. I reviewed abstracts for all of these articles to determine which ones potentially included relevant information; 235 articles made it through this initial screen. I then reviewed the full text of each of these articles to determine whether each met the following set of inclusion criteria, again patterned after Guise et al. (2010b): (1) the article reports information on the incidence of uterine rupture from an original case series containing at least 50 trials of labor; (2) the study sample was drawn from the general population of women with a prior cesarean section, rather than a specialized subgroup (e.g. women with multiple prior cesarean sections); (3) the article reports experience from a country in which standards of medical practice are broadly similar to those in the United States, defined herein as Australia, Canada, all of Europe, Israel, Japan, New Zealand, or the United States.[17] A total of 104 such articles were identified.

From each article identified, I extracted the total number of attempts of labor reported as well

[16] The similarity is not just coincidental. The NIS is typically drawn directly from the underlying state databases.

[17] In practice, the vast majority of articles are from the United States or Canada, and most of the rest were from the United Kingdom or Australia.

as the total number of uterine ruptures. In general, this was straightforward, but the definition of "uterine rupture" used merits further comment. As discussed in detail by Guise et al. (2003), different studies use different terminology to categorize separations of the uterine wall. Many studies apply the term "uterine rupture" to separations of the uterine wall that are major, symptomatic, or require surgical intervention and apply the term "uterine dehiscence" to smaller, asymptomatic separations that are discovered incidentally. Other authors, however, describe all separations of the uterine wall as "ruptures" or "dehiscences," but distinguish more important events using an adjective like "symptomatic," "complete," or (in the case of rupture) "true." Unfortunately, none of these terms have standard definitions, and authors are typically do not explicitly define how they are using these terms.

This variation in terminology across studies raises the question of what information should be extracted from each study. The goal of the analyses in this paper is to understand how each piece of clinical evidence should have affected a physician's estimate of the risk of serious complications resulting from separation of the uterine wall during a trial of labor. Guided by this goal, from articles reporting both the number of "ruptures" and the number of "dehiscences," I extracted the number of ruptures. From articles that report only the number of "ruptures" and provide no subcategorization, I also extracted the number of ruptures. From articles that use the term "rupture" or "dehiscence" and identify a subset of the events as serious, I extracted the number of serious events. In cases where each event is described in narrative form, I counted any event that was symptomatic. I excluded a small number of studies that report only the number of incidents of "uterine dehiscence" or the combined number of dehiscences and ruptures.

In several cases, multiple articles were published based on the same underlying case series, with later publications typically including additional cases. In any given analysis, I include only the most complete version of each case series that had been published by the time to which that analysis corresponds.

1.4 Provider learning from experience: event study evidence

I turn first to the question of whether and how physicians and hospitals change their practice styles in response to their own idiosyncratic experiences, focusing specifically on how experiencing a

uterine rupture with a patient with a history of cesearean section affects management of subsequent deliveries. In this section, I first derive the event study estimator I use to investigate this question. I then discuss how I implement the estimator in my context and present my results.

1.4.1 Event study econometric framework

I use an event study approach to estimate the effect of uterine rupture events on subsequent provider decisions. In brief, I identify the effect of an event by comparing the practice style trends of providers who do not experience an event at a given point in time to providers who do experience an event but would have been expected to follow similar trends in the absence of the event.

My approach extends the conditional difference-in-differences framework of Abadie (2005) (which built in turn upon Heckman et al. (1997) and Heckman et al. (1998)) to the event study setting. The key complication of the event study setting is that treatments (i.e. events) occur at multiple points in time and each unit can be treated multiple times (i.e. experience multiple events). In contrast, the Abadie (2005) framework envisions a setting in which all units are treated at most once and all treated units receive treatment at the same time. My approach also explicitly address the hierarchical structure of this setting (in which births are nested within providers). From an applied perspective, the resulting estimator may be viewed as a formalization of the event-study approach taken in Hilger (2012).

The approach taken here has several important advantages relative to dynamic difference-in-differences specifications in the spirit of Jacobson et al. (1993) that regress the outcome of interest on a set of leads and lags of event occurrence, a set of time effects, and a set of unit fixed effects. First, I show that my approach estimates a precisely-defined average treatment effect, even in the presence of arbitrary treatment effect heterogeneity; outside of constant coefficients models, precisely characterizing what dynamic difference-in-differences models estimate is frequently difficult. Second, the approach taken herein provides a computationally simple and conceptually straightforward method of accounting for trend differences across different (observed) categories of units. Finally, the framework prescribes an single, unambiguous method of handling units that experience multiple events, a problem that is typically handled through various *ad hoc* approaches (Sandler and Sandler, 2012).

The approach taken here does have one potential disadvantage relative to standard dynamic

difference-in-differences specifications: it estimates the effect of an event *inclusive* of the effect on the probability of future events, rather than the effect of an event holding the path of future events constant. In the present context, the effect of an event on the probability of future events is modest, so these two quantities will be similar and, in any case, it is not clear that the "partial" effect of an event (i.e. the effect excluding any effect on the probability of future events) is the quantity of interest. There are context, however, where this estimator will be inappropriate. This will be the case, for example, in contexts where all units will eventually experience a single event and the only difference is timing, as in studies of the effects of technology adoption (e.g. McClellan and Newhouse (1997)) or the staggered roll-out of a new program (e.g. Almond et al. (2011)).

In the remainder of this subsection, I first present my econometric framework, then discuss identification and estimation in this framework, and finally describe the application of this framework to my particular question.

Basic setup

Consider a setting with a general hierarchical structure. There exists a population of top-level units indexed by i. Any given top-level unit i may have multiple associated sub-units, which are indexed by j. For the sake of concreteness, I will refer to the top-level units as medical providers (e.g. hospitals or physicians) and the sub-units as births or deliveries. Events occur at the level of providers, but I am interested in estimating (average) causal effects at the level of deliveries. I observe the universe of deliveries j associated with some random sample of providers i. Throughout, expectation and probability operators will be defined with respect to the population birth-level marginal distributions induced by this provider-level sampling scheme.[18]

To formally define the relevant causal estimands and derive an estimator, I develop a potential outcomes framework in the spirit of Rubin (1974), Rubin (1977), Holland (1986). Each delivery occurs at some time T_{ij} and has a set of potential outcome pairs $\{(Y_{ij}^q(0), Y_{ij}^q(1))\}_{q\in\mathbb{Z}}$, where the potential outcome $Y_{ij}^q(1)$ (resp. $Y_{ij}^q(0)$) is the outcome realized by delivery j if provider i did (resp. did not) experience an event at time $T_{ij} - q$. Naturally, I refer to q as "time since event." For

[18]Note that if provider size or volume patterns are informative about patient characteristics, this distribution can differ from the expected marginal distribution for any particular number of clusters (Nevalainen et al., 2013). In general, these differences should disappear quickly as the number of sampled clusters grows. The implications of this issue for estimation in small samples are discussed in detail in Appendix A.5.

convenience, I define the event time $E_{ij}^q = T_{ij} - q$. In practice, I focus on some restricted set of time-since-event values $\mathcal{H} = \{\underline{q}, \underline{q}-1, \ldots, \bar{q}-1, \bar{q}\}$ for some integers $\underline{q} < 0$ and $\bar{q} > 0$.

The causal effects of interest are the differences $Y_{ij}^q(1) - Y_{ij}^q(0)$ for values $q > 0$. For each delivery j, however, only the the realized outcome $Y_{ij} = Y_{ij}^q(D_{ij}^q)$ is observed, where $D_{ij}^q = 1$ if provider i experienced at event at time $T_{ij} - q$ and is zero otherwise.[19] The counterfactual outcome $Y_{ij}^q(1 - D_{ij}^q)$ is not observed. As a consequence, it is not possible to directly compute causal effects for individual deliveries; this is "the fundamental problem of causal inference" (Holland, 1986). The goal, therefore, is to make assumptions that make it possible to impute the missing potential outcome $Y_{ij}^q(1-D_{ij}^q)$ on average for a defined population of deliveries. It is then possible to estimate the average causal effect $Y_{ij}^q(1) - Y_{ij}^q(0)$ for that defined population of deliveries.

To facilitate construction of a suitable comparison group, I suppose that some set of characteristics $\{X_{ij}^q\}_{q \in H}$ are observed for each delivery and time-since-event q. The vector X_{ij}^q could include characteristics of the provider (e.g. hospital volume) or the delivery (e.g. the mother's insurance status), and these characteristics could in principle vary depending on the time horizon. In general, I will focus on some appropriate subset $\mathcal{C}$ of the set of all possible event time and characteristic tuples (E_{ij}^q, X_{ij}^q) such that the identifying assumptions are particularly plausible for tuples in $\mathcal{C}$.

Identification

To identify the effects of interest, I apply a difference-in-differences approach around each event time e conditional on delivery characteristics x. The two substantive conditions that permit identification follow:

Condition NPE (No pre-event effects)**.** For all values $q < 0$, $Y^q(1) = Y^q(0)$.

Condition CT (Common trends)**.** For all event time and characteristic tuples $(e, x) \in \mathcal{C}$ and times-since-event $q, r \in \mathcal{H}$, the following holds:

$$\begin{aligned}\mathbb{E}[Y^q(0) \,|\, E^q = e, X^q = x, D^q = 1] - \mathbb{E}[Y^r(0) \,|\, E^r = e, X^r = x, D^r = 1] \\ = \mathbb{E}[Y^q(0) \,|\, E^q = e, X^q = x, D^q = 0] - \mathbb{E}[Y^r(0) \,|\, E^r = e, X^r = x, D^r = 0].\end{aligned}$$

[19] Consistency, of course, requires that $Y_{ij}^q(D_{ij}^q)$ is the same for all q.

Condition NPE states that events that will occur in the future have no effect on current outcomes.[20] This is a particularly reasonable assumption when events are difficult to predict, as is the case in my context.

Condition CT is the adaptation of the standard (conditional) difference-in-differences common trends assumption to the event study context.[21] The quantity $\mathbb{E}[Y^q(0) \,|\, E^q = e, X^q = x, D^q = 1]$ is the average counterfactual (hence, unobserved) outcome for deliveries by a provider who experienced an event q quarters earlier at time e, restricted to those deliveries with characteristics x. The quantity $\mathbb{E}[Y^q(0)|E^q = e, X^q = x, D^q = 0]$ is the average observed outcome for deliveries by a provider who did not experience an event q quarters earlier at time e, restricted to those deliveries with characteristics x. The assumption states that, for any given event time and set of characteristics, the counterfactual average for event units trends identically to the observed average for non-event units as we vary time-since-event.

To ensure that the relevant conditional expectations are estimable, I also impose the following technical condition:

Condition OO (Overlap and observability)**.** For all event time and characteristic tuples $(e, x) \in \mathcal{C}$ and all $q \in \mathcal{H}$, the following hold:

(i) (Observability) $\mathbb{P}(E^q = e, X^q = x) > 0$; and

(ii) (Overlap) $0 < \mathbb{P}(D^q = 1 \,|\, E^q = e, X^q = x) < 1$.

The first part of this condition ensures that deliveries are observed for all time horizons $q \in \mathcal{H}$ for each event time and characteristic tuple $(e, x) \in \mathcal{C}$. Note that this implies that X^q is discrete. It is conceptually straightforward to accommodate continuous X^q, but since doing so complicates the exposition and my application uses discrete X^q, I eschew those complications here. The second part of this condition is a standard overlap condition. It ensures that for any event time and characteristic tuple $(e, x) \in \mathcal{C}$, we can find suitable non-event (i.e control) observations at any time-since-event $q \in \mathcal{H}$.

[20] In stating an explicit no-anticipation condition, my exposition follows that of Miquel (2003). This assumption is stated implicitly in Abadie (2005), Heckman et al. (1997), and Heckman et al. (1998).

[21] In principle, one could fold E^q_{ij} into X^q_{ij}. In practice, conditioning on event time E^q_{ij} is such a crucial portion of the identification strategy, that is is worth the modestly more burdensome notation to emphasize time's role.

As in the standard difference-in-differences setting, Conditions NPE and CT imply that the average counterfactual outcome for units that do experience events can be written in terms of observed quantities. For $q > 0$ and $r < 0$, these assumptions imply that

$$\begin{aligned}
&\mathbb{E}[Y^q(0) \,|\, E^q = e, X^q = x, D^q = 1] \\
&\qquad = \mathbb{E}[Y^r(0) \,|\, E^r = e, X^r = x, D^r = 1] \\
&\qquad\quad + \{\mathbb{E}[Y^q(0) \,|\, E^q = e, X^q = x, D^q = 0] - \mathbb{E}[Y^r(0) \,|\, E^r = e, X^r = x, D^r = 0]\} \\
&\qquad = \mathbb{E}[Y^r(1) | E^r = e, X^r = x, D^r = 1] \\
&\qquad\quad + \{\mathbb{E}[Y^q(0) | E^q = e, X^q = x, D^q = 0] - \mathbb{E}[Y^r(0) \,|\, E^r = e, X^r = x, D^r = 0]\} \\
&\qquad = \mathbb{E}[Y \,|\, E^r = e, X^r = x, D^r = 1] \\
&\qquad\quad + \{\mathbb{E}[Y \,|\, E^q = e, X^q = x, D^q = 0] - \mathbb{E}[Y \,|\, E^r = e, X^r = x, D^r = 0]\},
\end{aligned}$$

where the first equality follows from Condition CT and the second equality follows from Condition NPE. Since $\mathbb{E}[Y^q(1) \,|\, E^q = e, X^q = x, D^q = 1]$ is also observed, the conditional treatment effect on the treated $\mathbb{E}[Y^q(1) - Y^q(0) \,|\, E^q = e, X^q = x, D^q = 1]$ is identified for all $q > 0$ as

$$\begin{aligned}
&\mathbb{E}[Y^q(1) - Y^q(0) | E^q = e, X^q = x, D^q = 1] \\
&\qquad = \{\mathbb{E}[Y | E^q = e, X^q = x, D^q = 1] - \mathbb{E}[Y | E^q = e, X^q = x, D^q = 0]\} \\
&\qquad\qquad - \{\mathbb{E}[Y | E^r = e, X^r = x, D^r = 1] - \mathbb{E}[Y | E^r = e, X^r = x, D^r = 0]\},
\end{aligned}$$

the usual difference in differences.

Conditions NPE and CT also have a set of testable implications. In particular, it is trivial to show that for any two values $q, r \in H$ with $q, r < 0$, it must hold that

$$\begin{aligned}
&\mathbb{E}[Y \,|\, E^q = e, X^q = x, D^q = 1] - \mathbb{E}[Y \,|\, E^q = e, X^q = x, D^q = 0] \\
&\qquad = \mathbb{E}[Y \,|\, E^r = e, X^r = x, D^r = 0] - \mathbb{E}[Y \,|\, E^r = e, X^r = x, D^r = 0].
\end{aligned}$$

That is, under the identifying assumptions, the difference between event and non-event outcomes is constant in the pre-period. These restrictions permit a formal test of the identifying assumptions, the formal counterpart of the standard "graphical test" of common trends.

In practice, there is rarely enough power to precisely estimate a treatment effect for any par-

ticular tuple $(e,x) \in \mathcal{C}$, and so it will be desirable to average over $(e,x) \in \mathcal{C}$. The discussion of identification given above carries over directly, and, as shown in the following lemma, this approach then identifies a (weighted) average treatment effect on the treated:

Lemma 1. *For any chosen probability measure W on $\mathcal{C}$, define for each $q \in \mathcal{H}$*

$$\Delta^q \equiv \int_{\mathcal{C}} \mathbb{E}[Y \mid E^q, X^q, D^q = 1] - \mathbb{E}[Y \mid E^q, X^q, D^q = 0] dW(E^q, X^q).$$

If Conditions NPE, CT, and OO hold, then

$$\Delta^q - \Delta^r = \int_{\mathcal{C}} \mathbb{E}[Y^q(1) - Y^q(0) \mid E^q, X^q, D^q = 1] dW(E^q, X^q)$$

for all $q, r \in \mathcal{H}$ with $q > 0$ and $r < 0$. In addition, $\Delta^q = \Delta^r$ for all $q, r \in \mathcal{H}$ with $q, r < 0$.

Applying the estimator requires choosing some particular weighting function $W(e,x)$. If the goal is to estimate a treatment effect at a single post-event point in time $q > 0$, then one option with considerable appeal is to use the weighting function

$$W_{\text{TOT}}(e,x) = \mathbb{P}(E^q = e, X^q = x \mid D^q = 1, (E^q, X^q) \in \mathcal{C}).$$

In this case, it is straightforward to see that the resulting weighted average treatment effect is $\mathbb{E}[Y^q(1) - Y^q(0) \mid D^q = 1, (E^q, X^q) \in \mathcal{C}]$, the average effect of treatment on the treated q quarters after the event. Unfortunately, this choice of weight is not well-suited to settings in which treatment effect dynamics are of interest because it prescribes a different weighting function for each $q > 0$ and thus renders the estimated treatment effects for different post-periods non-comparable. The multiplicity of weighting schemes also complicates evaluation of the common trends assumption in the pre-period since there are as many weighting schemes for the pre-period data as there are post-periods. An alternative choice that avoids these problems is to use the weighting function

$$W_{\text{WTOT}}(e,x) = \frac{1}{\bar{q}} \sum_{q=1}^{\bar{q}} \mathbb{P}(E^q = e, X^q = x \mid D^q = 1, (E^q, X^q) \in \mathcal{C}). \tag{1.5}$$

Under this weight, the natural "pooled" estimand, $\frac{1}{\bar{q}} \sum_{q=1}^{\bar{q}} \Delta^q - \frac{1}{\underline{q}} \sum_{q=\underline{q}}^{-1} \Delta^q$, satisfies

$$\frac{1}{\bar{q}} \sum_{q=1}^{\bar{q}} \Delta^q - \frac{1}{\underline{q}} \sum_{q=\underline{q}}^{-1} \Delta^q = \frac{1}{\bar{q}} \sum_{q=1}^{\bar{q}} \mathbb{E}[Y^q(1) - Y^q(0) \mid D^{q_1} = 1, (E^q, X^q) \in \mathcal{C}].$$

The estimand on the right-hand-side is a simple and readily interpretable average of (weighted)

average treatment effects on the treated.

Estimation

To define a suitable estimator, it is helpful to establish some notation. Define $S_d^q(e,x) = \{(i,j) : D_{ij}^q = d, E_{ij}^q = e, \text{ and } X_{ij}^q = x\}$ and $N_d^q(e,x) = \left|S_d^q(e,x)\right|$, and define a random variable $B(e,x)$ that equals one if $N_d^q(e,x) > 0$ for all $q \in \mathcal{H}$ and $d \in \{0,1\}$ and is zero otherwise. When $B(e,x) = 1$, all sample means corresponding to (e,x) will be well-defined. Letting $\hat{W}(e,x)$ be some estimator of the chosen weighting function $W(e,x)$, the natural estimator for each Δ^q can then be written as

$$\hat{\Delta}^q = \sum_{(e,x)\in\mathcal{C}:B(e,x)=1} \hat{W}(e,x) \left[\frac{1}{N_1^q(e,x)} \sum_{(i,j)\in S_1^q(e,x)} Y_{ij} \;-\; \frac{1}{N_0^q(e,x)} \sum_{(i,j)\in S_0^q(e,x)} Y_{ij} \right]. \tag{1.6}$$

Observe that a single delivery may appear in $\hat{\Delta}^q$ for multiple different times-since-event q, serving as a control observation when $D_{ij}^q = 0$ and as a treatment observation when $D_{ij}^q = 1$. In practice, I use the natural empirical counterpart of the choice of weighting function proposed in equation (1.5) for $\hat{W}(e,x)$:

$$\hat{W}_{\text{WTOT}}(e,x) = \frac{1}{\bar{q}} \sum_{q=1}^{\bar{q}} \frac{N_1^q(e,x)}{\sum_{(e',x')\in\mathcal{C}:B(e',x')=1} N_1^q(e',x')}.$$

The following lemma shows that the resulting estimator is consistent. The proof, which is not completely trivial on account of the hierarchical structure, is presented in Appendix A.5.

Lemma 2. *Under standard regularity conditions, if* $\hat{W}(e,x) \xrightarrow{p} W(e,x)$ *for all* $(e,x) \in \mathcal{C}$ *as the sampled numbered of providers* $M \to \infty$*, then* $\hat{\Delta}^q \xrightarrow{p} \Delta^q$ *as* $M \to \infty$*. Furthermore,* $\hat{W}_{WTOT}(e,x)$ *satisfies this condition.*

A concern about using Lemma 2 to justify the use of this estimator is that the the number of births in any particular (e,x) cell may be relatively small, in which case the large-sample properties of the estimator may not be a good guide to its behavior in the present application. One can show, however, that under a modest strengthening of the common trends assumption, the estimators proposed herein are (conditionally) unbiased for a weighted (conditional) average treatment effect on the treated. Furthermore, even in small samples, failures of the identifying assumptions are likely to show up as pre-period differential trends. Because the required arguments are somewhat notationally burdensome but not particularly illuminating, I have confined them to Appendix A.5.

Computationally, it is frequently convenient to compute $\hat{\Delta}^q$ via a regression on an auxiliary dataset that takes account of the multiplicity of roles played by individual deliveries. For each delivery (i,j) and quarter $q \in \mathcal{H}$ such that $(E_{ij}^q, X_{ij}^q) \in \mathcal{C}$ and $B(E_{ij}^q, X_{ij}^q) = 1$, the auxiliary dataset contains a separate record of the form $(\tilde{Y}_m, \tilde{D}_m, \tilde{Q}_m, \tilde{\phi}_m) = (Y_{ij}, D_{ij}^q, q, \phi_{ij}^q)$, where m indexes records in the auxiliary dataset and

$$\phi_{ij}^q = \frac{\hat{W}(E_{ij}^q, X_{ij}^q)}{N_{D_{ij}^q}^q(E_{ij}^q, X_{ij}^q)}.$$

Using this auxiliary dataset, one then runs a regression of the form

$$\tilde{Y}_m = \sum_{q \in H} \sum_{d \in \{0,1\}} \beta_d^q \mathbf{1}\{\tilde{Q}_m = q, \tilde{D}_m = d\} + \epsilon_m, \tag{1.7}$$

weighted by $\tilde{\phi}_m$. It is trivial to see that $\hat{\Delta}^q = \hat{\beta}_1^q - \hat{\beta}_0^q$ for all $q \in H$. This regression specification is very similar to the specification arrived at by Hilger (2012) without a formal justification.[22] Standard errors may be obtained via a block bootstrap at the provider level.

It is tempting to add covariates to equation (1.7). Doing so is difficult to justify in general, however, as the addition of covariates "re-weights" the underlying means, with the result that $\hat{\beta}_1^q - \hat{\beta}_0^q$ may no longer consistently estimate Δ^q.[23]

1.4.2 Estimator implementation and sample construction

In this subsection, I discuss how I apply the event study estimator derived above to the question of interest in this paper: the effect of provider experience with uterine rupture on subsequent treatment decisions for patients with a prior cesarean delivery. In my main analyses, each provider i is an individual hospital, but I also examine specifications in which each provider i is an individual physician. The event indicator D_{ij}^q equals one if provider i experienced uterine rupture with a patient at time $T_{ij} - q$ and is zero otherwise. The unit of time is a calendar quarter, and I examine a window extending 6 quarters before and after each event. The outcomes of interest are whether

[22] As in Hilger (2012), it can greatly speed computation to run the regression in collapsed form, particularly for the purposes of obtaining bootstrapped standard errors. In particular, I compute the mean for each (e, x, q, d) tuple. A regression on this set of means weighted by $\hat{W}(e, x)$ will generate the same point estimates as the full-sample regression.

[23] For a detailed discussion of this point in a related context, see the discussion in Angrist and Pischke (2009) comparing matching estimators and regression estimators under a conditional independence assumption.

a delivery is preceded by an attempt of labor and whether the ultimate mode of delivery is vaginal or cesarean. I present bootstrapped standard errors based on 200 replications.

The most important implementation choice is which variables are included in the conditioning vector X^q_{itj}. In all analyses, X^q_{ij} includes the provider's state, and the set $\mathcal{C}$ of event time and characteristic tuples (E^q_{ij}, X^q_{ij}) is restricted to include only state-quarter tuples for which the data permit a full 6 quarters of follow-up and look-back. I select additional covariates so as to ensure that I am comparing providers that are as similar as possible in their risk of experiencing an event, thereby minimizing the differences between event and non-event units.[24]

For the hospital-level analyses, I also include in X^q_{ij} the following: an indicator for whether the hospital had positive delivery volume at time E^q_{ij}; and the hospital's delivery volume quartile in its state at time E^q_{ij}, where only women with a prior cesarean section are included in delivery volume. I restrict the set $\mathcal{C}$ so as to exclude from consideration hospitals with no deliveries at time E^q_{ij}. Thus, this event study estimator identifies the effect of an event at time e by comparing hospitals that do experience an event at a time e to other hospitals that do not experience an event at time e but are located in the same state and are of similar size at time e.

For the physician-level analyses, on the other hand, I also include in X^q_{ij} an indicator for whether the physician attempted labor with at least one patient at time E^q_{ij}, and the set $\mathcal{C}$ is restricted to include only those deliveries for which this indicator is equal to one. Thus, this event study estimator identifies the effect of an event at time e by comparing physicians who do experience an event at a time e to other physicians who do not experience an event at time e but are located in the same state and had a patient who could have experienced an event at time e.

To implement these estimators, I construct a sample of births from the HCUP State Inpatient Databases and OSHPD Patient Discharge Databases described in Section 1.3. I start by constructing a sample that includes all deliveries by women with a prior cesarean delivery. I exclude a small number of records for which the quarter of delivery is missing, the associated provider is missing, or the provider has very low volume. In my main analyses, in which each provider i is a different hospital, "very low volume" is defined as never exceeding 20 deliveries (by all women, regardless of prior cesarean status) in a single quarter. In supplemental analyses, each provider i is a different

[24]Ideally, of course, the covariates would perfectly capture a provider's risk of an event, in which case the effects of interest are identified even without the common trends assumption.

Table 1.1: Descriptive statistics for the event study samples

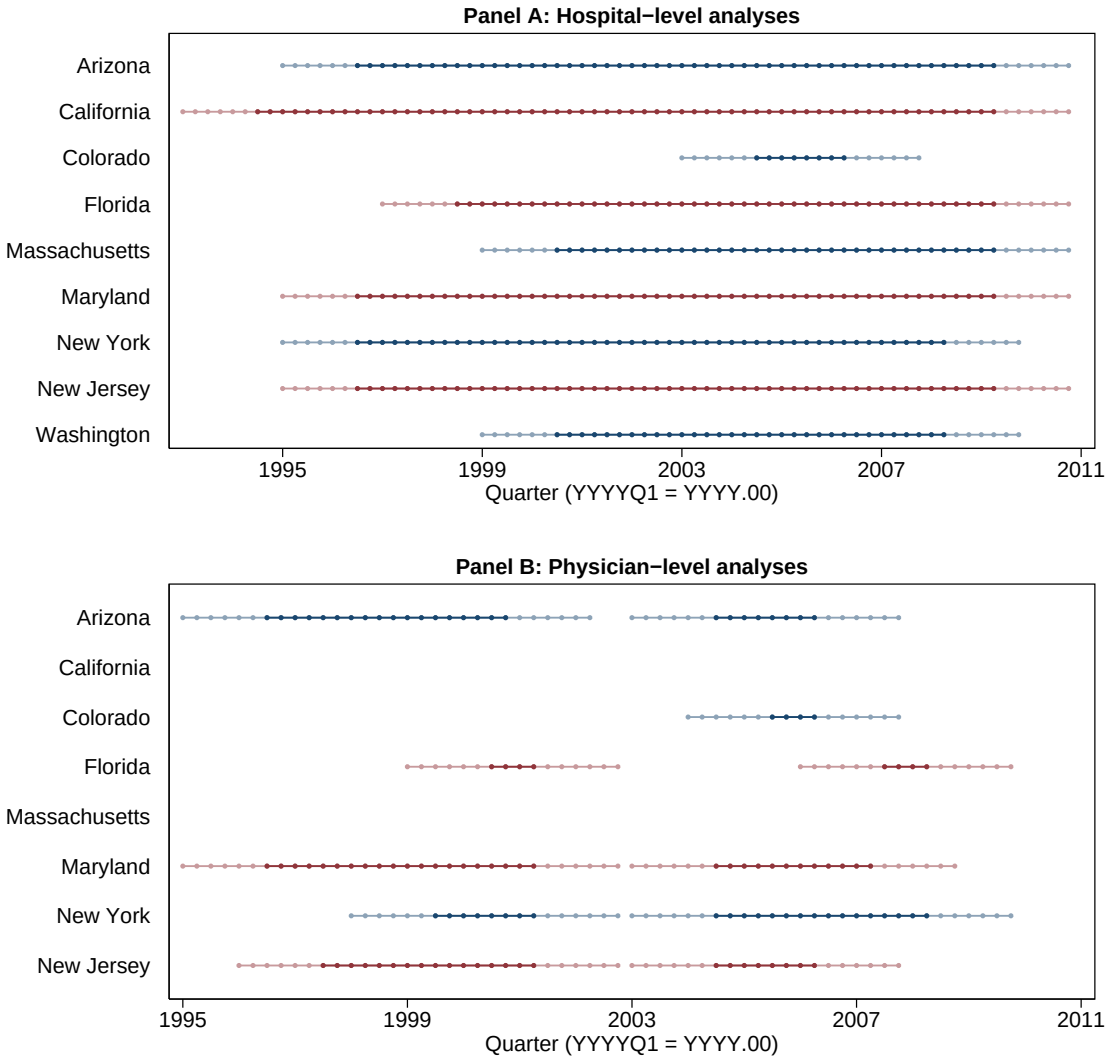
Provider level:	Hospital			Physician		
	Mean	SD	Share missing	Mean	SD	Share missing
Delivery characteristics						
Trial of labor	0.245	0.430	0.00	0.221	0.415	0.00
Vaginal delivery	0.179	0.383	0.00	0.163	0.369	0.00
Uterine rupture	0.003	0.050	0.00	0.003	0.051	0.00
Length of stay	3.05	2.19	0.00	3.16	2.15	0.00
Maternal characteristics						
Age	30.38	5.60	0.02	30.63	5.68	0.00
Black	0.121	0.327	0.12	0.171	0.376	0.03
Hispanic	0.322	0.467	0.12	0.208	0.406	0.03
Medicaid	0.395	0.489	0.00	0.335	0.472	0.00
Counts						
Events	5,107			1,223		
Providers	995			11,828		
Births	2,981,860			1,002,054		

Notes: The event study samples are constructed from the HCUP State Inpatient Databases and the OSHPD Patient Discharge Databases as described in Section 1.4.2. Means and standard deviations are computed using all observations for which that particular field is non-missing. Delivery characteristics other than induced labor are identified using ICD-9-CM codes as described in the text. Most cases of missing age data are attributable to California. Most records with missing race and ethnicity data are from Washington state.

physician, and the threshold is set at 5 deliveries.

The set of state-quarters with useable data after restricting to those permitting 6 quarters of follow-up and look-back as described above is depicted in Figure 1.3. Table 1.1 provides descriptive statistics on the analysis sample of births. As indicated in the table, the number of events available for the hospital-level analyses is considerably larger, owing to the fact that hospital identifiers are available for all states and the hospital identifiers are consistent over the full period, which avoids the loss of state-quarter cells around breaks in those identifiers. While this larger sample size is partially offset by the presence of a far smaller number of clusters at the hospital-level, the hospital-level results will still be considerably more precise.

Figure 1.3: State-quarters included in the event study analyses



Notes: This figure depicts the state-quarters for which the HCUP State Inpatient Database or OSHPD Patient Discharge Data are usable for each set of event study analyses. Connected state-quarters have comparable identifiers. Births in the darkly shaded quarters fall more than 6 quarters from the beginning or end of a segment; these quarters permit the full 6 quarters of required follow-up and look-back and, thus, estimating the effect of events occurring during these quarters is feasible. Births occurring during the lightly shaded quarters do not permit the necessary follow-up or look-back, so estimating the effect of events occurring during these quarters is not feasible; these births are used only for follow-up or look-back when estimating the effect of events occurring during the darkly-shaded quarters.

1.4.3 Event study results

I now present my main results and several auxiliary specifications aimed at understanding the sources of the effects I document. After that, I turn to a variety of specification checks aimed at ruling out plausible violations of the identifying assumptions.

Main results

I start by presenting graphical evidence from estimating equation (1.7) at the hospital level. Figure 1.4 depicts the average delivery mode for births at hospitals that will or will not experience an event at $q = 0$, weighted as prescribed by equation (1.7). The plotted points therefore correspond, respectively, to the coefficient sets $\{\beta_1^q\}$ and $\{\beta_0^q\}$ from equation (1.7). Examining the pre-period, the event time series lies above the non-event series, which is to be expected since providers with higher trial of labor rates are at higher risk of experiencing an event. Both the event and non-event series also exhibit a significant downward trend, a reflection of the fact that these analyses use data for a period when trial of labor and vaginal delivery rates were declining. Reassuringly, the trends affecting the two series appear to be very similar, lending support to the common trends assumption necessary for identification. Turning to the post-period, Figure 1.4 shows that the gap between the event group mean and the non-event group mean narrows in the post-period, indicating that a uterine rupture event causes the hospitals' subsequent patients to be less likely to attempt labor.

In Figure 1.5, I present these results in differences rather than levels. Specifically, the figure plots the difference in means between event and non-event births as a function of the time since the event. The dashed horizontal line depicts the mean pre-period difference, and the error bars correspond to 95 percent confidence intervals for the difference between the mean pre-period difference and the current period difference. Failure of the 95 percent confidence interval to include the dashed line corresponds to the existence of a statistically significant causal effect of a uterine rupture event on practice at that time horizon. Figure 1.5 reinforces both key conclusions drawn from Figure 1.4: mean delivery mode at event and non-event providers trends very similarly prior to the event; and the gap between event and non-event providers narrows in the post-period. The immediate narrowing is roughly half a percentage point, but this grows to roughly three-quarters of a percentage point by the end of the six-quarter follow-up period, indicating that uterine rupture

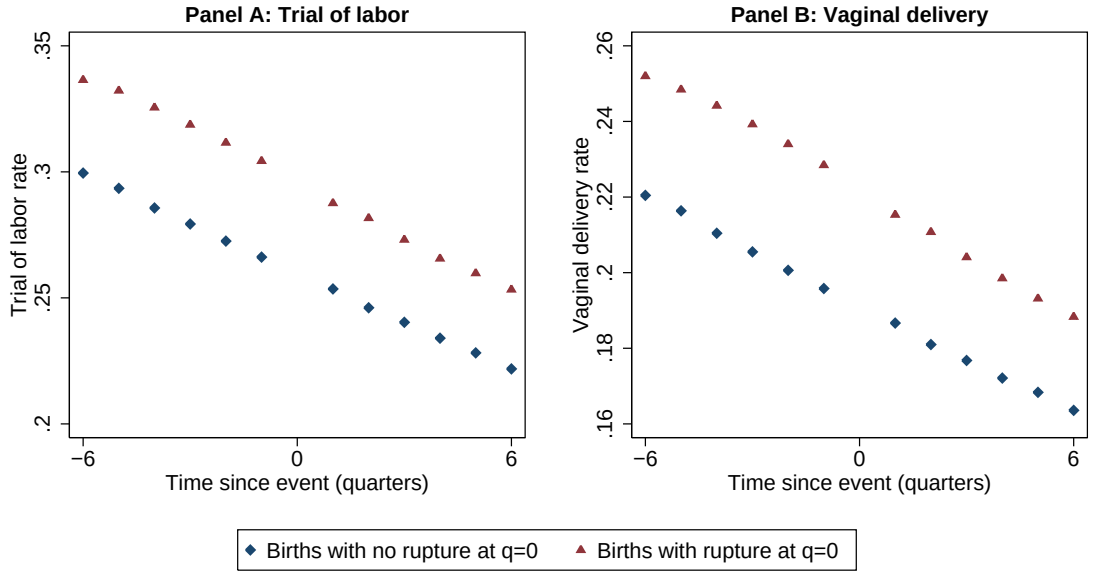
events have a causal effect on hospitals' subsequent practice patterns.

Table 1.2 summarizes these results in tabular form. It reports separate estimated causal effects for each individual post-period, defined (as justified earlier) as $\hat{\Delta}^q - \frac{1}{6}\sum_{r=-6}^{-1}\hat{\Delta}^r$, as well as an estimated "pooled" causal effect that corresponds to the average causal effect for births over the full 6 quarter post-period: $\frac{1}{6}\sum_{q=1}^{6}\hat{\Delta}^q - \frac{1}{6}\sum_{q=-6}^{-1}\hat{\Delta}^q$. The table also reports the results of a formal test of the hypothesis that $\Delta^q = \Delta^r$ for all $q, r \in \mathcal{H}$ with $q, r < 0$, which, as shown in Lemma 1, provides a formal test of the common trends assumption. The tabular results codify the graphical evidence: there is no evidence against the common trends assumption; and a change in practice starts immediately after the event and then grows modestly over the ensuing quarters. Before proceeding, I note that, as discussed in Section 1.3 and demonstrated in detail in Appendix A.6, approximately one-third of events identified using the ICD-9-CM code for uterine rupture appear to be relatively minor non-rupture injuries that occur during cesarean section. The point estimates reported in Table 1.2 and in subsequent tables should therefore be scaled up by a factor of 1.5 to account for this measurement error.

The fact that the causal effects appear to grow over the first few quarters following an event merits further comment. Since all the new information is revealed when the event occurs or soon thereafter, the model presented in Section 1.2 would imply that the full response should occur almost immediately. There are at least three things that may be missing from the model, however. First, for this population of women, whether to attempt labor is frequently discussed in early prenatal visits. To the extent that obstetricians are loathe to revisit "settled" conversations, effects on practice will phase in only gradually as the obstetrician's panel turns over. Dranove and Watanbe (2009) find a similar pattern in obstetricians' responses to being sued for medical malpractice, which they attribute to this mechanism. Second, to the extent that these responses are occurring through hospital-level policy changes, some delay is to be expected since a process of discussion and consultation among the hospital's obstetricians is likely to be required. Finally, some of the response to these events may be mediated through the malpractice claims that they generate. Malpractice suits stemming from the rupture event may take many months to be filed, which could also generate a delayed effect.

The pattern of results also provides some insight into the margins along which affected hospitals are adjusting their behavior. It is clear that the occurrence of an event makes a hospital less likely

Figure 1.4: Trends in mode of delivery for event and non-event births by time since event with events defined at the hospital level

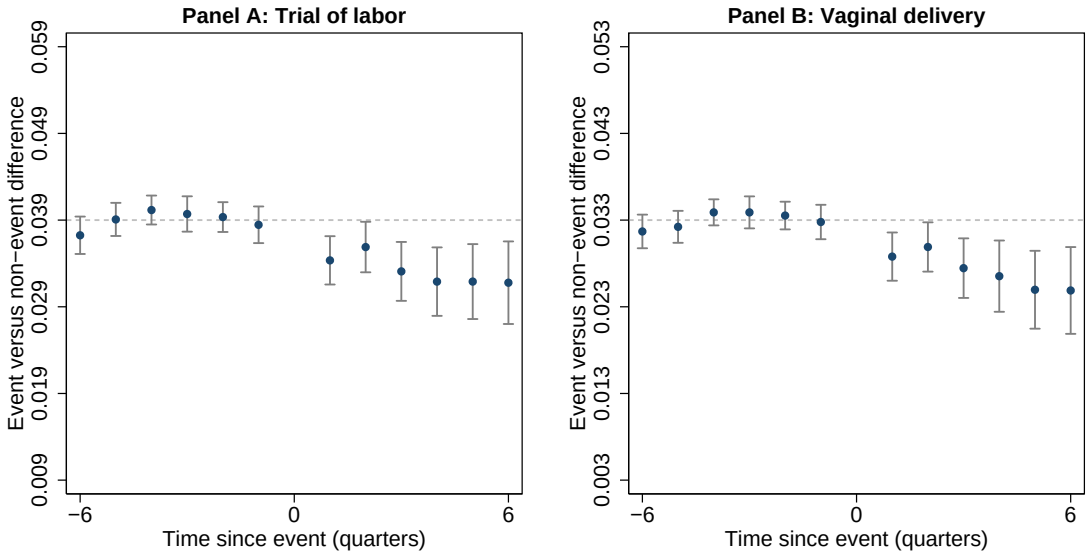


Notes: This figure reports the weighted average delivery mode for women with a prior cesarean section at hospitals that will or will not experience an event at $q = 0$. The reported estimates are obtained by estimating equation (1.7) as described in the text with events defined at the hospital level and the outcome an indicator for trial of labor (panel A) or vaginal delivery (panel B). Estimation uses HCUP State Inpatient Databases for the states and years described in the text and the OSHPD Patient Discharge Data for 1993-2010. The "births with rupture at $q = 0$" time series corresponds to the estimates $\hat{\beta}_1^q$, while the "births with no rupture at $q = 0$" time series corresponds to the estimates $\hat{\beta}_0^q$. The means for $q = 0$ are mechanically distorted due to the presence (or non-presence) of the event itself and therefore omitted.

to make an attempt of labor with a subsequent patient. The results provide suggestive evidence, however, that this is not the only margin of response. On average, only approximately three-quarters of TOLACs lead to vaginal deliveries (see Table 1.1), and it is likely that the marginal TOLAC patient is less likely to be successful than the average TOLAC patient. One would therefore expect the effect of an event on the VBAC rate to be somewhat smaller than the effect on the TOLAC rate if the only margin of response is the probability of initiating labor. In fact, the VBAC point estimate is approximately the same size as the TOLAC point estimate, suggesting that the affected hospitals may also become more likely to abandon an attempt of labor, perhaps because they start using a lower threshold to diagnose complications.

I turn now to understanding the level at which learning is occurring. Specifically, I investigate whether the estimated response reflects diffuse hospital-wide changes in behavior or instead reflects

Figure 1.5: Difference in mode of delivery between event and non-event births by time since event with events defined at the hospital level



Notes: This figure reports the difference in delivery outcomes between births at hospitals that will or will not experience an event at $q = 0$. The reported estimates are obtained by estimating equation (1.7) as described in the text with events defined at the hospital level and the outcome an indicator for trial of labor (panel A) or vaginal delivery (panel B). Estimation uses HCUP State Inpatient Databases for the states and years described in the text and the OSHPD Patient Discharge Data for 1993-2010. The reported point estimate for each time-since-event q is $\hat{\Delta}^q = \hat{\beta}_1^q - \hat{\beta}_0^1$. The difference for $q = 0$ is mechanically distorted due to the presence (or non-presence) of the event itself and therefore omitted. The horizontal dashed line depicts the mean pre-period difference: $\frac{1}{6}\sum_{q=-6}^{-1}\hat{\Delta}^q$. The error bars depict 95 percent confidence intervals corresponding to a two-tailed test of the hypothesis that the current period difference is the same as the mean pre-period difference. Failure of the current-period confidence interval to include the pre-period mean therefore implies existence of a statistically significant causal effect at the 5 percent level. The underlying covariance matrix is obtained via a block bootstrap at the hospital level with 200 replications.

Table 1.2: Effect of a uterine rupture event on management of subsequent patients (with events defined at the hospital level)

Column:	(1)	(2)
	Dependent variable	
Post-event horizon	Trial of labor	Vaginal delivery
+ 1 quarter	-0.0046**	-0.0042**
	(0.0014)	(0.0014)
+ 2 quarters	-0.0031*	-0.0031*
	(0.0015)	(0.0014)
+ 3 quarters	-0.0059***	-0.0055**
	(0.0017)	(0.0018)
+ 4 quarters	-0.0071***	-0.0065**
	(0.0020)	(0.0021)
+ 5 quarters	-0.0071**	-0.0080***
	(0.0022)	(0.0023)
+ 6 quarters	-0.0072**	-0.0081**
	(0.0024)	(0.0026)
Pooled	-0.0058***	-0.0059***
	(0.0017)	(0.0018)
Auxiliary information		
Common trends p-value	0.383	0.688
Unique births	2,955,195	2,955,195
Nominal N	28,763,506	28,763,506

Notes: This table reports event study estimates of the effect of a uterine rupture event on hospitals' subsequent management of deliveries by women with a prior cesarean delivery. The reported estimates are obtained by estimating equation (1.7) as described in the text with events defined at the hospital level. Estimation uses HCUP State Inpatient Databases for the states and years described in the text and the OSHPD Patient Discharge Data for 1993-2010. The point estimate reported in the row labeled "$+q$ quarters" is defined as $\hat{\Delta}^q - \frac{1}{6}\sum_{r=-6}^{-1}\hat{\Delta}^r$; the results in the text show that this may be interpreted as the causal effect of an event q quarters later. The "pooled" estimate is defined as $\frac{1}{6}\sum_{q=1}^{6}\hat{\Delta}^q - \frac{1}{6}\sum_{q=-6}^{-1}\hat{\Delta}^q$; this quantity may be interpreted as the average causal effect over the first six post-event quarters. Standard errors are obtained via a block bootstrap at the physician level using 200 replications and displayed in parentheses. The common trends p-value is obtained from a standard χ^2 test of the of the hypothesis that $\Delta^q = \Delta^r$ for all $q, r < 0$. Statistical significance is denoted as follows: $^{+}$ $p < .1$, * $p < .05$, ** $p < .01$, *** $p < .001$.

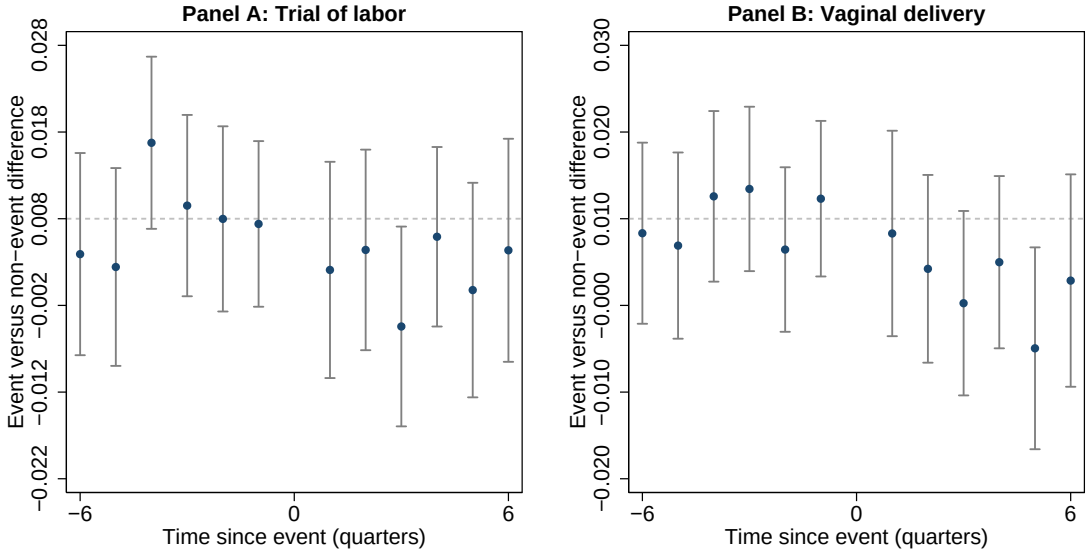
large changes in behavior by just the physician experiencing the event. To address this question, I re-estimate equation (1.7), defining event status at the physician level rather than the hospital level (and altering the matching strategy as described previously). Figure 1.6 and Table 1.3 report the results; to facilitate comparison to the hospital-level results, Figure 1.6 uses the same scale as Figure 1.5. The resulting physician-level estimates are very similar to the hospital-level estimates, albeit considerably less precise, suggesting that the bulk of the response is due to changes in behavior by physicians other than the physician experiencing the event. A simple calculation is instructive. In the year an event occurs, the physician directly experiencing the event accounts for less than 10 percent of the hospital's volume. Even taking the top end of the 95 percent confidence interval for the effect of an event on the VBAC rate at the physician level (1.5 percentage points), this means that the physician-level response can account for less than one-quarter of the hospital-wide response. I conclude, therefore, that the main response mechanisms must be hospital-wide.[25]

The learning model implies that hospitals should respond less to events that occur later in time, for two reasons. First, as hospitals accrue more experience, the effect of any single event on beliefs is smaller. Second, TOLAC rates move away from 50 percent over the time period examined in these analyses, and under most plausible assumptions about the patient risk preference distribution, this implies that the sensitivity of treatment decisions to any particular change in beliefs will fall as well. To test this prediction, I re-estimate equation (1.7) separately for two sets of events: those occurring in 2002 or earlier and those occurring in 2003 or later. Figure 1.7 and Table 1.4 report the results, which show that events have large effects in the earlier period, but little or no effect in the later period, a pattern qualitatively consistent with the model's predictions.[26]

[25]One caveat to this interpretation is that most obstetricians practice in multi-physician groups, as documented in Appendix A.3. In such groups, the physician with whom a woman decides on her planned mode of delivery will often not be the physician who attends the delivery. The physician-level estimates will miss any effects on patients the physician counsels but does not deliver, and, thus, may understate the total effect of an event on that physician's practice. Even so, since a physician is much more likely to deliver the patients she counsels than a randomly selected patient at the hospital, we should still expect the physician-level response to be considerably larger than the hospital-level response if the main response mechanisms operate at the level of the individual physician. This does not appear to be the case.

[26]One additional factor may be at work. As discussed previously, the estimates reported here are attenuated by measurement error that arises because the ICD-9-CM code for uterine rupture during labor can also be used for non-rupture events. As discussed in detail in Section 1.3 and Appendix A.6, the share of identified rupture events that are actual ruptures will tend to fall as the trial of labor rate falls. Calculations based on the model in Appendix A.6 imply that 74 percent of identified ruptures are true ruptures in the earlier period, but this share falls to 50 percent in the later period. This change would be expected to reduce the estimated effect by approximately one-third, even without changes in true responsiveness.

Figure 1.6: Difference in mode of delivery between event and non-event births by time since event with events defined at the physician level



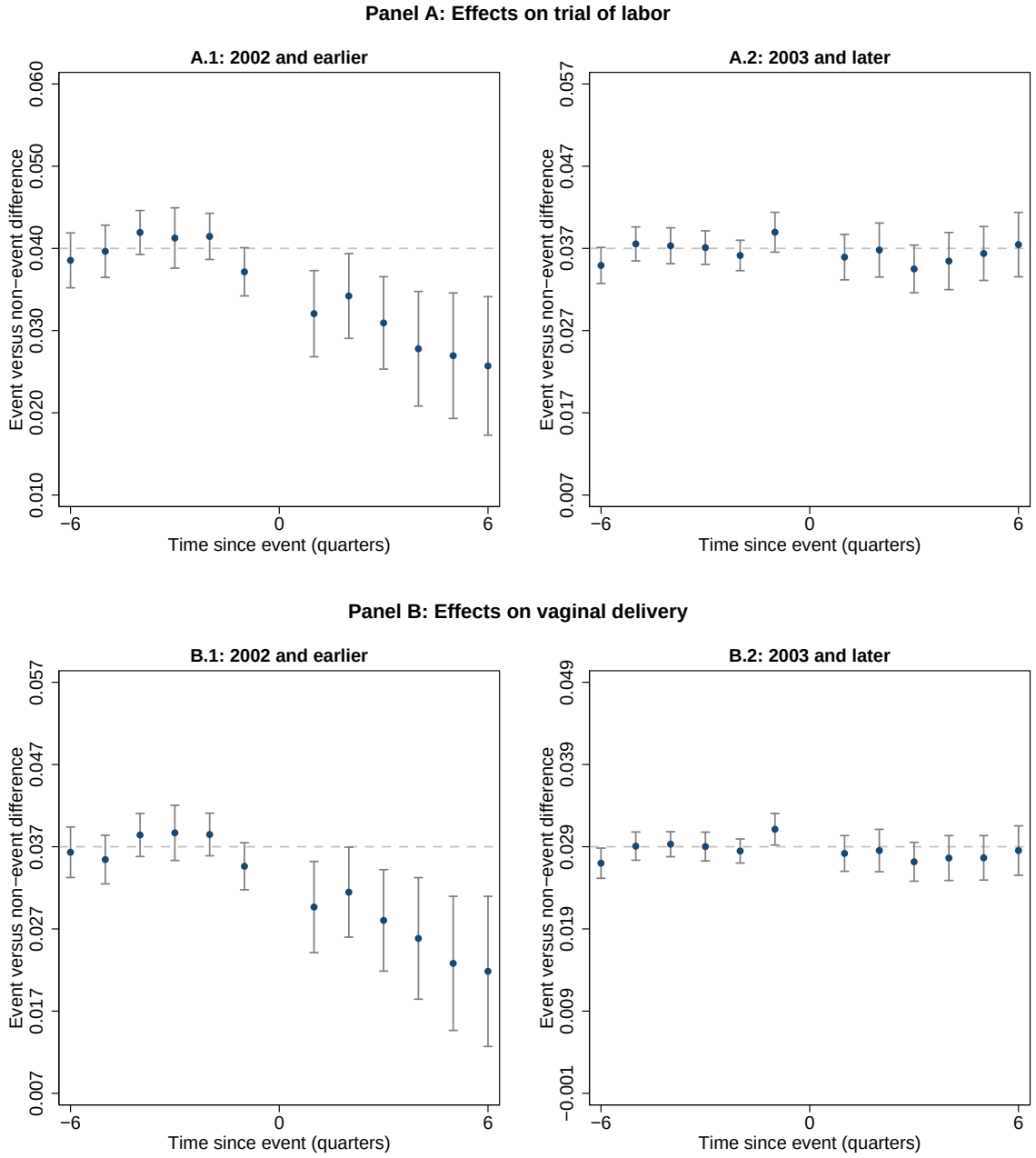
Notes: This figure reports the difference in delivery outcomes between births attended by physicians that will or will not experience an event at $q = 0$. The reported estimates are obtained by estimating equation (1.7) as described in the text with events defined at the physician level and the outcome an indicator for trial of labor (panel A) or vaginal delivery (panel B). Estimation uses HCUP State Inpatient Databases for the states and years described in the text. The reported point estimate for each time-since-event q is $\hat{\Delta}^q = \hat{\beta}_1^q - \hat{\beta}_0^1$. The difference for $q = 0$ is mechanically distorted due to the presence (or non-presence) of the event itself and therefore omitted. The horizontal dashed line depicts the mean pre-period difference: $\frac{1}{6}\sum_{q=-6}^{-1}\hat{\Delta}^q$. The error bars depict 95 percent confidence intervals corresponding to a two-tailed test of the hypothesis that the current period difference is the same as the mean pre-period difference. Failure of the current-period confidence interval to include the pre-period mean therefore implies existence of a statistically significant causal effect at the 5 percent level. The underlying covariance matrix is obtained via a block bootstrap at the hospital level using 200 replications.

Table 1.3: Effect of a uterine rupture event on management of subsequent patients (with events defined at the physician level)

Column:	(1)	(2)
	Dependent variable	
Post-event horizon	Trial of labor	Vaginal delivery
+ 1 quarter	-0.0059	-0.0017
	(0.0064)	(0.0060)
+ 2 quarters	-0.0036	-0.0058
	(0.0059)	(0.0055)
+ 3 quarters	-0.0124*	-0.0097^{+}
	(0.0059)	(0.0054)
+ 4 quarters	-0.0021	-0.0050
	(0.0053)	(0.0051)
+ 5 quarters	-0.0082	-0.0149*
	(0.0063)	(0.0059)
+ 6 quarters	-0.0036	-0.0071
	(0.0066)	(0.0063)
Pooled	-0.0060	-0.0074^{+}
	(0.0040)	(0.0038)
Auxiliary information		
Common trends p-value	0.540	0.879
Unique births	867,811	867,811
Nominal N	3,386,723	3,386,723

Notes: This table reports event study estimates of the effect of a uterine rupture event on physicians' subsequent management of deliveries by women with a prior cesarean delivery. The reported estimates are obtained by estimating equation (1.7) as described in the text with events defined at the physicians level. Estimation uses HCUP State Inpatient Databases for the states and years described in the text. The point estimate reported in the row labeled "$+q$ quarters" is defined as $\hat{\Delta}^q - \frac{1}{6}\sum_{r=-6}^{-1}\hat{\Delta}^r$; the results in the text show that this may be interpreted as the causal effect of an event q quarters later. The "pooled" estimate is defined as $\frac{1}{6}\sum_{q=1}^{6}\hat{\Delta}^q - \frac{1}{6}\sum_{q=-6}^{-1}\hat{\Delta}^q$; this quantity may be interpreted as the average causal effect over the first six post-event quarters. Standard errors are obtained via a block bootstrap at the physician level using 200 replications and displayed in parentheses. The common trends p-value is obtained from a standard χ^2 test of the of the hypothesis that $\Delta^q = \Delta^r$ for all $q, r < 0$. Statistical significance is denoted as follows: $^{+}$ $p < .1$, * $p < .05$, ** $p < .01$, *** $p < .001$.

Figure 1.7: Time heterogeneity in the difference in mode of delivery between event and non-event births by time since event



Notes: This figure reports the difference in delivery outcomes between births for which the associated hospital will or will not experience an event at $q = 0$. The reported estimates are obtained by estimating equation (1.7) separately for events in 2002 or earlier and events in 2003 or later. Events are defined at the hospital level, and the outcome is an indicator for trial of labor (panel A) or vaginal delivery (panel B). The point estimate for each time-since-event q is $\hat{\Delta}^q$. The difference for $q = 0$ is mechanically distorted due to the presence (or non-presence) of the event itself and therefore omitted. The horizontal dashed line depicts the mean pre-period difference: $\frac{1}{6}\sum_{q=-6}^{-1}\hat{\Delta}^q$. The error bars depict 95 percent confidence intervals corresponding to a two-tailed test of the hypothesis that the current period difference is the same as the mean pre-period difference; failure of the current-period confidence interval to include the pre-period mean therefore implies existence of a statistically significant causal effect at the 5 percent level. The underlying covariance matrix is obtained via a block bootstrap at the hospital level using 200 replications.

Table 1.4: Time heterogeneity in the effect of a uterine rupture event on management of subsequent patients

Column:	(1)	(2)	(3)	(4)
Outcome:	Trial of labor		Vaginal delivery	
Time Period/ Post-event horizon	≤ 2002	≥ 2003	≤ 2002	≥ 2003
+ 1 quarter	-0.0079**	-0.0011	-0.0073**	-0.0008
	(0.0027)	(0.0014)	(0.0028)	(0.0011)
+ 2 quarters	-0.0058*	-0.0002	-0.0055*	-0.0005
	(0.0026)	(0.0017)	(0.0028)	(0.0013)
+ 3 quarters	-0.0091**	-0.0025^{+}	-0.0090**	-0.0018
	(0.0029)	(0.0015)	(0.0031)	(0.0012)
+ 4 quarters	-0.0122***	-0.0015	-0.0112**	-0.0014
	(0.0036)	(0.0018)	(0.0038)	(0.0014)
+ 5 quarters	-0.0131***	-0.0006	-0.0142***	-0.0013
	(0.0039)	(0.0017)	(0.0042)	(0.0014)
+ 6 quarters	-0.0143***	0.0005	-0.0152**	-0.0005
	(0.0043)	(0.0020)	(0.0047)	(0.0015)
Pooled	-0.0104***	-0.0009	-0.0104**	-0.0010
	(0.0031)	(0.0014)	(0.0034)	(0.0011)
Auxiliary information				
Common trends p-value	0.244	0.122	0.352	0.089
Unique births	1,552,521	1,923,648	1,552,521	1,923,648
Nominal N	13,243,389	15,520,117	13,243,389	15,520,117

Notes: This table reports event study estimates of the effect of a uterine rupture event on physicians' subsequent management of deliveries by women with a prior cesarean delivery. The reported estimates are obtained by estimating equation (1.7) separately separately for events occurring in 2002 or earlier and for events occurring in 2003 or later with events defined at the hospital level. The point estimate reported in the row labeled "$+q$ quarters" is defined as $\hat{\Delta}^q - \frac{1}{6}\sum_{r=-6}^{-1}\hat{\Delta}^r$; the results in the text show that this may be interpreted as the causal effect of an event q quarters later. The "pooled" estimate is defined as $\frac{1}{6}\sum_{q=1}^{6}\hat{\Delta}^q - \frac{1}{6}\sum_{q=-6}^{-1}\hat{\Delta}^q$; this quantity may be interpreted as the average causal effect over the first six post-event quarters. Standard errors are obtained via a block bootstrap at the physician level using 200 replications and displayed in parentheses. The common trends p-value is obtained from a standard χ^2 test of the of the hypothesis that $\Delta^q = \Delta^r$ for all $q, r < 0$. Statistical significance is denoted as follows: $^{+}$ $p < .1$, * $p < .05$, ** $p < .01$, *** $p < .001$.

Another prediction of the learning model is that, all else equal, larger hospitals should respond less to events since larger hospitals will typically have accrued more experience. To test this prediction, I re-estimate equation (1.7) separately by hospital delivery volume quartile in the quarter of the event, where delivery volume includes only deliveries by women with a prior cesarean delivery. Table 1.5 reports the results. Contrary to the model's prediction, I find no evidence that larger hospitals respond less to events. If anything, it appears that larger hospitals respond slightly more, although the differences across volume quartiles are not statistically significant.

There are two interpretations of this fact. The first is that the model is incorrect and the response to events is being mediated through a non-learning channel. The other possibility, however, is that a hospital's own volume is not the relevant measure of the stock of experience it has to draw on. If hospitals learn from their own experience but also from the experience of other hospitals in their local areas, then the differences in experience across volume groups might be quite modest, leading to a pattern like the one observed.

For completeness, Appendix A.7 reports evidence on whether the size of the response varies by patient characteristics or event severity. I find little evidence of heterogeneous responses across subgroups, subject to the caveat that my power to detect cross-group differences is limited. It is also worth noting, however, that because it is impossible to link maternal records to the accompanying neonatal record (or records), my measures of event severity do not incorporate information on adverse neonatal outcomes, which are typically the most feared outcomes of uterine rupture. For this reason, the event severity results should be taken with a significant grain of salt.

Threats to validity

One major threat to the validity of these results is that events cause a change in the mix of patients seen by the affected providers, either because events cause providers to seek out different types of patients or because events cause some categories of patients to seek out alternative providers. If changes in patient mix of this form occur, then the trend in delivery outcomes experienced by the non-event providers will not provide a valid counterfactual for the trend experienced by event providers. The causal effects would then be biased.

I look for a causal effect of events on provider patient mix by estimating the same event study specifications as before, replacing the delivery outcomes with patients' estimated propensity to ex-

Table 1.5: Heterogeneity by hospital size in the effect of a uterine rupture event on management of subsequent patients

Column:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Outcome:	Trial of labor				Vaginal delivery			
Volume quartile/ Post-event horizon	Bottom	Second	Third	Top	Bottom	Second	Third	Top
+ 1 quarter	-0.0099	-0.0033	-0.0048*	-0.0046**	-0.0175*	-0.0018	-0.0055**	-0.0039*
	(0.0091)	(0.0039)	(0.0022)	(0.0015)	(0.0082)	(0.0034)	(0.0019)	(0.0016)
+ 2 quarters	0.0014	-0.0044	-0.0060*	-0.0024	-0.0066	-0.0035	-0.0064**	-0.0023
	(0.0088)	(0.0037)	(0.0025)	(0.0017)	(0.0075)	(0.0037)	(0.0022)	(0.0017)
+ 3 quarters	-0.0120	-0.0012	-0.0081**	-0.0057**	-0.0067	-0.0034	-0.0072**	-0.0053**
	(0.0110)	(0.0044)	(0.0025)	(0.0019)	(0.0088)	(0.0042)	(0.0023)	(0.0019)
+ 4 quarters	-0.0021	-0.0014	-0.0100***	-0.0069**	0.0006	-0.0034	-0.0080***	-0.0064**
	(0.0106)	(0.0051)	(0.0028)	(0.0023)	(0.0080)	(0.0044)	(0.0024)	(0.0023)
+ 5 quarters	-0.0027	-0.0049	-0.0067**	-0.0074**	-0.0101	-0.0043	-0.0077***	-0.0083**
	(0.0105)	(0.0043)	(0.0026)	(0.0025)	(0.0089)	(0.0040)	(0.0021)	(0.0026)
+ 6 quarters	-0.0050	-0.0008	-0.0070**	-0.0077**	-0.0063	-0.0012	-0.0058*	-0.0091**
	(0.0099)	(0.0046)	(0.0026)	(0.0027)	(0.0082)	(0.0038)	(0.0022)	(0.0028)
Pooled	-0.0050	-0.0027	-0.0071***	-0.0058**	-0.0078	-0.0029	-0.0068***	-0.0059**
	(0.0074)	(0.0033)	(0.0020)	(0.0019)	(0.0060)	(0.0029)	(0.0018)	(0.0020)
Auxiliary information								
Common trends p-value	0.679	0.354	0.462	0.614	0.095	0.364	0.965	0.754
Unique births	279,509	777,024	1,360,160	1,937,389	279,509	777,024	1,360,160	1,937,389
Nominal N	951,042	3,641,999	7,732,987	16,437,478	951,042	3,641,999	7,732,987	16,437,478

Notes: This table reports event study estimates of the effect of a uterine rupture event on hospitals' subsequent management of deliveries by women with a prior cesarean delivery. The reported estimates are obtained by estimating equation (1.7) separately for each hospital delivery volume quartile, where delivery volume includes only women with a prior cesarean delivery. Events are defined at the hospital level. Estimation uses HCUP State Inpatient Databases for the states and years described in the text and the OSHPD Patient Discharge Data for 1993-2010. The point estimate reported in the row labeled "$+q$ quarters" is defined as $\hat{\Delta}^q - \frac{1}{6}\sum_{r=-6}^{-1}\hat{\Delta}^r$; the results in the text show that this may be interpreted as the causal effect of an event q quarters later. The "pooled" estimate is defined as $\frac{1}{6}\sum_{q=1}^{6}\hat{\Delta}^q - \frac{1}{6}\sum_{q=-6}^{-1}\hat{\Delta}^q$; this quantity may be interpreted as the average causal effect over the first six post-event quarters. Standard errors are obtained via a block bootstrap at the hospital level using 200 replications and displayed in parentheses. The common trends p-value is obtained from a standard χ^2 test of the of the hypothesis that $\Delta^q = \Delta^r$ for all $q, r < 0$. Statistical significance is denoted as follows: $^{+}$ $p < .1$, * $p < .05$, ** $p < .01$, *** $p < .001$.

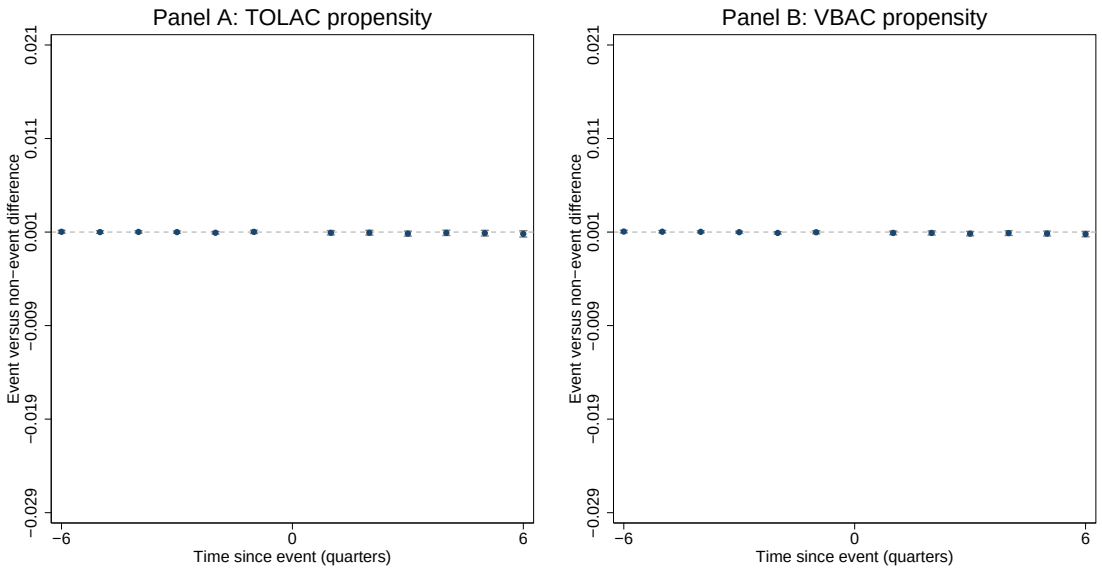
perience those delivery outcomes based on observable characteristics (Baicker et al., 2006; Chandra and Staiger, 2007). I obtain these propensities as the predicted values from ordinary least squares regressions of delivery mode on a full set of indicators for single year of age, several categories of patient race, several categories of insurance status, and dummies for each calendar quarter. I estimate these specifications separately for each state since coding of age, race, and insurance status can differ across states. In general, age and calendar quarter are the strongest (available) predictors of delivery mode.

Figure 1.8 depicts the results of these analyses; to facilitate comparison to the main results, I plot these on a scale of the same magnitude. The figure indicates that events do not cause shifts in providers' patient mixes, at least not shifts that are of substantial concern for the present analysis. Table 1.6 reports the point estimates in tabular form, as well as corresponding estimates for several underlying patient characteristics: maternal age, black race, Hispanic ethnicity, and Medicaid enrollment.[27] The results in Table 1.6 generally support the conclusion that events do not lead to changes in provider case mix. Experiencing an event does appear to reduce a hospital's Medicaid share, but this effect is only marginally significant, and the absolute magnitude of the effect is small.

A second threat to the validity of these results is that events are more likely to occur at hospitals with high trial of labor rates. Indeed, referring back to Figure 1.5, TOLAC rates at hospitals that will experience an event next quarter are approximately 4 percentage points higher than those at hospitals that will not experience an event. For a variety of reasons, one might expect these high-TOLAC hospitals to experience a relative fall in TOLAC rates over the ensuing quarters, perhaps due to simple statistical mean reversion or perhaps because all hospitals were gradually transitioning to a lower-TOLAC practice style and only the high-TOLAC hospitals had so far failed to do so. To address this concern, I randomly generate 2000 sets of "placebo" uterine rupture events. In order to match the process actually generating those events as precisely as possible, I generate each event according to the model for the use of the underlying ICD-9-CM code that was described in Section 1.3 and that is discussed in detail in Appendix A.6. I then re-estimate the event study for

[27] For the purposes of the individual characteristic analyses only, I code missing ages as 99, so these analyses will also be sensitive to sharp changes in the share of deliveries in which maternal age is coded as missing. For the race, ethnicity, and Medicaid enrollment indicators, I code missing data as zero.

Figure 1.8: Difference in patient characteristics between event and non-event births by time since event



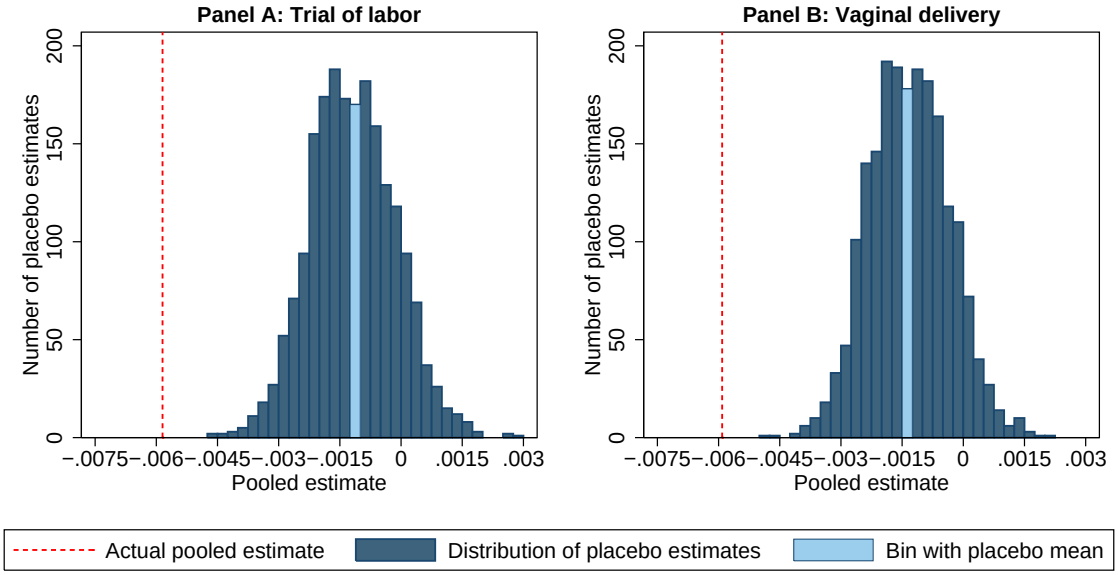
Notes: This figure reports the difference in patients' estimated propensity for different delivery outcomes between births for which the associated hospital will or will not experience an event at $q = 0$. The reported estimates are obtained by estimating equation (1.7) as described in the text with events defined at the hospital level and the outcome the patient's estimated propensity for for trial of labor (panel A) or vaginal delivery (panel B). The point estimate for each time-since-event q is $\hat{\Delta}^q$. The difference for $q = 0$ is mechanically distorted due to the presence (or non-presence) of the event itself and therefore omitted. The horizontal dashed line depicts the mean pre-period difference: $\frac{1}{6}\sum_{q=-6}^{-1}\hat{\Delta}^q$. The error bars depict 95 percent confidence intervals corresponding to a two-tailed test of the hypothesis that the current period difference is the same as the mean pre-period difference; failure of the current-period confidence interval to include the pre-period mean therefore implies existence of a statistically significant causal effect at the 5 percent level. The underlying covariance matrix is obtained via a block bootstrap at the hospital level using 200 replications.

Table 1.6: Effect of a uterine rupture event on subsequent patient mix

Column:	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable					
Post-event horizon	Age	Black	Hispanic	Medicaid	TOLAC propensity	VBAC propensity
+ 1 quarter	-0.0139	-0.0013^{+}	-0.0008	-0.0042*	-0.0001	-0.0001
	(0.0174)	(0.0007)	(0.0019)	(0.0019)	(0.0001)	(0.0001)
+ 2 quarters	-0.0041	-0.0005	0.0000	-0.0028	-0.0001	-0.0001
	(0.0176)	(0.0008)	(0.0022)	(0.0022)	(0.0001)	(0.0001)
+ 3 quarters	0.0005	0.0004	-0.0013	-0.0036	-0.0002	-0.0002
	(0.0191)	(0.0008)	(0.0023)	(0.0023)	(0.0001)	(0.0001)
+ 4 quarters	-0.0031	0.0006	0.0001	-0.0038	-0.0001	-0.0001
	(0.0201)	(0.0008)	(0.0024)	(0.0027)	(0.0001)	(0.0001)
+ 5 quarters	0.0200	0.0000	-0.0012	-0.0066*	-0.0001	-0.0002
	(0.0189)	(0.0009)	(0.0026)	(0.0029)	(0.0002)	(0.0001)
+ 6 quarters	0.0170	0.0008	0.0007	-0.0063*	-0.0002	-0.0002
	(0.0197)	(0.0010)	(0.0028)	(0.0030)	(0.0002)	(0.0001)
Pooled	0.0027	0.0000	-0.0004	-0.0046*	-0.0001	-0.0001
	(0.0125)	(0.0007)	(0.0022)	(0.0023)	(0.0001)	(0.0001)
Auxiliary information						
Common trends p-value	0.552	0.292	0.740	0.163	0.880	0.818
Unique births	2,955,195	2,955,195	2,955,195	2,955,195	2,955,195	2,955,195
Nominal N	28,763,492	28,763,506	28,763,506	28,763,506	28,763,506	28,763,506

Notes: This table reports event study estimates of the effect of a uterine rupture event on hospitals' subsequent patient mix. The reported estimates are obtained by estimating equation (1.7) with events defined at the hospital level and the listed patient characteristic as the outcome. The point estimate reported in the row labeled "$+q$ quarters" is defined as $\hat{\Delta}^q - \frac{1}{6}\sum_{r=-6}^{-1}\hat{\Delta}^r$; the results in the text show that this may be interpreted as the causal effect of an event q quarters later. The "pooled" estimate is defined as $\frac{1}{6}\sum_{q=1}^{6}\hat{\Delta}^q - \frac{1}{6}\sum_{q=-6}^{-1}\hat{\Delta}^q$; this quantity may be interpreted as the average causal effect over the first six post-event quarters. Standard errors are obtained either by the analytic method described in the text or via a block bootstrap at the hospital level using 200 replications; the resulting standard errors are displayed in parentheses. The common trends p-value is obtained from a standard χ^2 test of the of the hypothesis that $\Delta^q = \Delta^r$ for all $q, r < 0$. Statistical significance is denoted as follows: $^{+}$ $p < .1$, * $p < .05$, ** $p < .01$, *** $p < .001$.

Figure 1.9: Distribution of placebo estimates of the effect of a uterine rupture event relative to the actual estimate



Notes: This figure reports the distribution of 2000 placebo estimates obtained by randomly generating "placebo" events, estimating equation (1.7) on the placebo dataset, and computing the "pooled" estimate $\frac{1}{6}\sum_{q=1}^{6}\hat{\Delta}^q - \frac{1}{6}\sum_{q=-6}^{-1}\hat{\Delta}^q$ from the placebo results. The navy blue bars depict the distribution of the placebo estimates, and the highlighted bar depicts the bin containing the mean of the placebo estimates. The dotted red line depicts the estimates obtained from the actual data and were reported in Table 1.2.

each of these placebo datasets and compute the corresponding pooled causal effect estimate.

Figure 1.9 plots the distribution of the placebo estimates, the mean placebo estimate, and the actual estimate. The results demonstrate that the actual estimate is far out in the left tail of the placebo estimates for both outcomes. The mean placebo estimate is slightly negative, but this should not be alarming, as providers that actually experienced events will naturally be somewhat overrepresented in the placebo event group relative to the placebo non-event group; thus, if events have a true causal effect, this is precisely what we should expect to see. I conclude that my main results cannot be accounted for by the fact that events are concentrated in high-TOLAC hospitals.

1.5 Common prior beliefs based on the clinical literature

I turn next to understanding providers' common prior beliefs about the risk of rupture, which as described in Section 1.2, I assume to be based on the clinical literature. Both the mean and variance

of this prior distribution are of great interest. The evolution of the mean of this distribution, together with the provider's experience to date, determines how the provider's trial of labor rate evolves over time. The variance of the prior – which reflects both sampling uncertainty and, potentially much more important, the provider's uncertainty about where she falls in the population distribution of risks – determines how much weight the provider places on her own experience when forming her beliefs about the risk of rupture.

To characterize the common prior, I assume that each provider rationally assesses the full set of available research articles (whether directly, indirectly through published meta-analyses, or informally through opinion leaders) and forms beliefs accordingly. The remainder of this section proceeds in three steps. I first present the statistical framework I use to compute the common prior from the reported results in the clinical literature. I next examine the estimated cross-study dispersion in the risk of rupture and examine how different assumptions about this dispersion affect the precision of providers' common prior, and I consider the implications for the event study results. I then undertake a rolling meta-analysis of the clinical literature over the full period 1980 to 2010, which provides estimates of the common prior at monthly frequency that I use to calibrate the learning model in the next section.

1.5.1 Framework for interpreting the clinical literature

This subsection presents a statistical model of how the true cross-provider distribution of the risk of rupture relates to the studies in the clinical literature. I then assume that providers have a correct understanding of this model, form beliefs about the parameters of this model using Bayes rule, and compute their common prior on the risk of rupture on the basis of those beliefs.

To start, suppose that the risk of uterine rupture p_h varies across providers h according to

$$p_h = \text{logit}^{-1}(X_h\beta + v_h),$$

where X_h represents observable features of the provider's patient mix and practice patterns that affect of the risk of rupture, β is a vector of coefficients representing the effects of those observable features, and $v_h \sim N(0, \sigma_v^2)$ captures unobserved (to the provider) features of the provider's practice that affect the probability of rupture. As discussed in Section 1.1, the currently-known correlates of the risk of uterine rupture are either quite weak or relevant to a small share of patients. This

implies that $\text{Var}(X_h\beta) \approx 0$, so I assume $X_h\beta = \gamma$ for some constant γ. The distribution of p_h therefore depends upon γ and σ_v^2.

I assume that providers obtain information on γ and σ_v^2 from the clinical literature. Toward this end, I assume that the reported probability of rupture in any given study s takes the form

$$q_s = \text{logit}^{-1}(\gamma + w_s + \epsilon_s),$$

where w_s captures unobserved characteristics of the study site that affect the probability of rupture and $\epsilon_s \sim N(0, \sigma_\epsilon^2)$ reflects measurement error that is unique to each study (e.g. design flaws, tabulation errors, etc.).[28] To make the link between outcomes in individual studies and the population distribution of risks, I assume that study sites are drawn randomly from the population of providers, so that $w_s \sim N(0, \sigma_v^2)$, just like v_h.[29]

Under these assumptions, the number of ruptures R_s in a study s that includes L_s attempts of labor follows the simple random effects logit structure

$$\begin{aligned} u_s &\sim N(\gamma, \sigma_u^2) \\ q_s &\sim \text{logit}^{-1}(u_s) \\ R_s &\sim \text{binomial}(L_s, q_s), \end{aligned} \tag{1.8}$$

where u_s is distributed $N(0, \sigma_u^2)$ with $\sigma_u^2 \equiv \sigma_v^2 + \sigma_\epsilon^2$. As described below, this model is easily estimated using standard tools. Note, however, that while γ is identified in this model, σ_v^2 is not identified separately from σ_ϵ^2. As a consequence, I will need to make an assumption about the share of the cross-study variance σ_u^2 that providers believe arises from true heterogeneity in risk rather than measurement error, which I will refer to as δ. In practice, I report results for a range of possible assumptions about δ.

I take a Bayesian approach to estimating equation (1.8) since I wish to assign a probabilistic interpretation to uncertainty in the estimated parameters (γ, σ_u^2). To do so, I use a specialized Gibbs sampler algorithm that is tailored to models with this structure. This algorithm is implemented

[28] The assumption that ϵ_s has mean zero is not innocuous. It reflects an assumption that the clinical literature is correct "on average."

[29] It seems unlikely that study sites are literally drawn randomly from the population of providers. In particular, it seems plausible that the variance of the risk across study sites is smaller than the cross-provider variance. If this is the case, then the estimates of σ_v^2 obtained below will understate the actual degree of cross-provider dispersion.

in the MCMCglmm package for R and is described in detail in Hadfield (2010).[30] Like all Markov Chain Monte Carlo techniques, the algorithm generates a sequence of draws from the posterior distribution for the model parameters that can then be used to evaluate any moments of those parameters that may be of interest. I generate 25,000 such draws, and, following standard practice, I discard the first 5,000 draws so as to eliminate any draws that depend in an important way on the starting point of the Markov chain. Thereafter, I retain only every tenth draw so as to keep the number of draws manageable while minimizing the correlation between subsequent draws, yielding 2,000 total draws. I use different priors in the next two subsections, and I explain each as it arises.

It will be convenient to translate the posterior distribution for p_h that is implied by the posterior distribution of (γ, σ_u^2) and the chosen value of δ into the terms of the beta distribution that appears in the learning model. I do so by first drawing 100 instances of p_h for each of the 2,000 posterior draws of (γ, σ_u^2) using whatever value has been selected for δ. I then find the "best fit" beta distribution by maximum likelihood.[31]

1.5.2 Meta-analysis of all available studies

I first compute the common prior that would exist after observing the full set of studies published over the period 1980 through 2010. The process by which I identified these studies was described in Section 1.3. As described there, I use only the most recent publication using a given study cohort in order to avoid "double counting" evidence.

For this subsection only, I estimate equation (1.8) using a diffuse prior, so the reported posterior means and standard deviations are very similar to the point estimates and standard errors one would obtain by maximum likelihood.[32] Table 1.7 reports the results. The most salient feature of these results is that the posterior mean for σ_u is extremely large, indicating that different studies are providing meaningfully different estimates of the risk of rupture. Concretely, the estimate of σ_u

[30] The random effects logit model considered here is part of a broader class of models that statisticians refer to as "generalized linear mixed models," which is the origin of the "glmm" in the package name.

[31] It would, of course, have been preferable to select a form for the model in equation (1.8) such that the posterior distribution of interest literally took a beta form. Unfortunately, I am unaware of any specification that would ensure that this is the case. In any case, the practical differences between the two distributions are small.

[32] Specifically, I use the default prior in MCMCglmm, which consists of a normal prior for γ with mean zero and variance 10^{10} and an improper flat prior for σ_u^2.

Table 1.7: Bayesian meta-analysis of the risk of uterine rupture based on studies published 1980-2010

Parameter/quantity	Posterior mean	Posterior SD
Logit intercept (γ)	-5.239	0.098
Logit standard deviation (σ_u)	0.706	0.082
Population mean ($\mathbb{E}[p_h]$)	0.0068	0.0007
Population median ($\text{logit}^{-1}(\gamma)$)	0.0053	0.0005
Observation counts		
Number of posterior samples	2,000	
Number of studies	88	
Number of trials of labor	400,513	

Notes: This table reports features of the posterior distribution of (μ, σ_u^2) obtained from Bayesian estimation of the model specified in equation (1.8). The sample of published studies used for estimation is described in Section 1.3. Estimation uses the specialized Gibbs sampler algorithm implemented in the MCMCglmm package for R. The posterior sample of 2,000 draws was obtained by taking every tenth draw from draws 5,000 through 25,000 produced by the Gibbs sampler. A diffuse prior was used for γ and σ_u^2 and is described in the text. The posterior population mean was obtained by taking 1000 draws of p_h for each posterior draw of (γ, σ_u^2), computing the mean for each posterior draw, and then computing the mean and standard deviation of the resulting means.

implies that a study one standard deviation above the median has a probability of rupture that is approximately 70 log points higher than the median study (after purging small-sample variation).[33]

How this estimated level of cross-study dispersion affects the providers' common prior depends on the share of the cross-study variance that is attributable to true heterogeneity in risk as opposed to simple measurement error; this is the parameter δ defined in the last section. If δ is large, then the population distribution of risks will be very spread out and the common prior will be more diffuse, while if δ is small, the population distribution will be tightly-packed and the common prior will be correspondingly more peaked.

To provide insight on the implications of different values of δ, Table 1.8 reports the results of fitting a beta distribution to the posterior distribution for p_h using the method described in the last section for each of five different values of δ. The results indicate that, under most assumptions, the

[33]Technically, it is the *odds ratio* that is 70 log points larger one standard deviation above the median. For probabilities close to zero, however, the odds ratio is essentially equal to the raw probability, so a 70 log point change in the odds ratio is essentially identical to a 70 log point change in the underlying probability.

common prior will be quite diffuse. Even when examining a relatively small value of $\delta = 0.25$, the common prior has a notional sample size of less than 1,400, meaning that the entirety of the clinical literature has information content equivalent to less than 1,400 of a physician's own cases. This is despite the fact that the literature as a whole encompasses more than 400,000 cases. Only when all of the cross-study variation is attributed to measurement error (i.e. $\delta = 0$) does the common prior become highly-peaked.[34]

The fact that the common prior is this diffuse has important implications for the interpretation of the event study estimates presented in the last section. I use equation (1.2) from the learning model to calculate what effect a single rupture event should have on a rational provider's beliefs about the risk of rupture for each value of the notional sample size reported in Table 1.8. Focusing again on the case where $\delta = 0.25$, I calculate that for a provider with no experience, a single uterine rupture increases the provider's estimate of the risk of rupture by 13.2 percent of the baseline risk. Even for a provider with substantial accumulated experience, the effect of an event on beliefs remains large: 7.6 percent of the baseline risk. When beliefs are this sensitive to events, rationalizing the event study estimates does not require patient preferences to be particularly sensitive to the risk of rupture and is thus relatively easy. This conclusion is inconsistent with the conclusions of Choudhry et al. (2006), who interpret physician responses to idiosyncratic experience as prima facie evidence that physicians overweight their own personal experience.

1.5.3 Rolling Bayesian meta-analysis

I next examine how the common prior evolved over time, which is a needed input into the calibration exercise in the next section. To do so, I repeat the basic Bayesian analysis from above, but on a "rolling" basis. That is, I fit the model separately for each month from 1980 to 2010 in which a new study is published, making use only of the studies that had been published by that point in time. Between publication months, I simply interpolate the results. I then fit a beta distribution to the resulting posterior distributions for p_h at each point in time in the same way as above.

[34] Observe that even if the share of σ_u^2 that reflects true heterogeneity in risk is zero, the notional sample size of the common prior is still well below the total number of cases observed in the clinical literature. The reason for this is essentially the same as the reason that non-clustered standard errors overstate precision in the context of linear regression (Moulton, 1986). In essence, the number of observations is most appropriately thought of as being the number of distinct studies, rather than the number of distinct individuals observed.

Table 1.8: Notional sample size of common prior obtained from meta-analysis

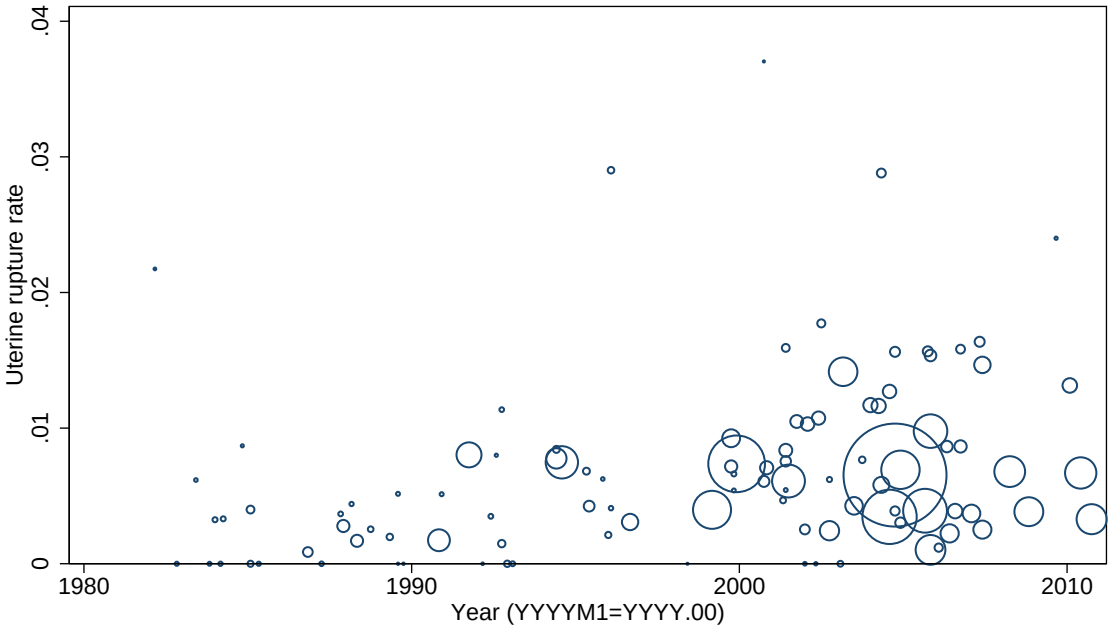
Share of variance due to risk heterogeneity (δ)	Common prior notional sample size ($\alpha + \beta$)	Effect of event on beliefs as share of mean risk	
		No experience	1,000 deliveries
0 percent	20,034	0.009	0.009
25 percent	1,341	0.132	0.076
50 percent	665	0.250	0.100
75 percent	432	0.363	0.109
100 percent	311	0.473	0.112

Notes: Each row of the table reports the results of estimating a "best fit" beta distribution to the posterior distribution of p_h when δ takes the listed value. The "best fit" beta distribution is estimated by maximum likelihood on a sample of draws of p_h constructed by generating 100 instances of p_h for each of the 2,000 posterior draws of (γ, σ_u^2) using the assumed value of δ. The effects of events on beliefs reported in the last two columns are computed according to equation (1.2), and the mean risk is computed using the vector (α, β) from the best fit beta distribution. The underlying draws of (γ, σ_u^2) are obtained from Bayesian estimation of the model specified in equation (1.8). The sample of published studies used for estimation is described in Section 1.3. Estimation uses the specialized Gibbs sampler algorithm implemented in the MCMCglmm package for R. The posterior sample of 2,000 draws was obtained by taking every tenth draw from draws 5,000 through 25,000 produced by the Gibbs sampler. A diffuse prior was used for γ and σ_u^2 and is described in the text.

The analyses in this section differ in one important respect from those in the last section: they require specifying providers' pre-1980 beliefs over (γ, σ_u^2). This choice was unimportant to the results obtained in the last section since those analyses included all studies published through 2010, which ensured that any plausible prior would be dominated by the data. As a result, for those analyses, I could simply specify a vague prior. As depicted in Figure 1.10, however, the available clinical literature during the 1980s is relatively sparse. As a consequence, a provider's pre-1980 beliefs play an important role for much of this period. Specifying such beliefs is necessarily somewhat speculative, but various historical reports make it possible to choose reasonable values. I explore the sensitivity of the estimated common prior to alternative specifications of the pre-1980 beliefs later in this section.

To calibrate the pre-1980 beliefs, I turn to evidence on the risk of rupture for women with a prior classical cesarean section. As noted earlier in the paper, the practice of routine repeat cesarean section originated in the early 20th century when the "classical" approach to cesarean section predominated. Since experience with trial of labor between the early 20th century and 1980 was minimal, it seems likely that beliefs about the risk of rupture that were forged in that era still predominated. Consistent with this view, a major NIH conference report on this topic published

Figure 1.10: Published studies on the risk of uterine rupture, 1980-2010



Notes: This figure plots the risk of uterine rupture reported by published research articles on this topic over the period 1980-2010. The area of each marker is proportional to the number of attempts of labor observed in the study. The method in which studies were identified and processed is described in Section 1.3. The same labor attempt may be reflected in multiple points since multiple studies frequently make use of the same underlying patient cohort. In the formal analyses presented in the text, only the most recently published version of each patient cohort is used.

in 1980 explicitly linked the then-prevailing beliefs regarding the risk of uterine rupture to the era in which the classical procedure predominated (NIH, 1980).[35]

Two meta-analyses from the early 1980s, Lavin et al. (1982) and O'Sullivan et al. (1981), both citing evidence reported by Dewhurst (1957), quote a rupture rate for women with a classical cesarean scar of 4.7 percent. Similarly, Case et al. (1971) reports that, among British physicians, 4 percent was an accepted figure for the risk of rupture during labor for women with a classical cesarean scar. Also consistent with the idea that physicians of the era were reasonably certain that the risk was quite high, Lavin et al. (1982) reports survey results from 1968 that 80 percent of physicians would not consider trial of labor for their patients even after being "shown...overwhelming evidence of the safety of permitting labor following cesarean section." Roughly following this evidence, I parametrize the pre-1980 beliefs on γ as following a normal distribution with a mean $\gamma_0 = \text{logit}^{-1}(0.04)$. I set the standard deviation of this normal distribution to $\sigma_\gamma = 0.50$, which reflects substantial certainty that the risk was high, but limited certainty regarding its precise magnitude.

Specifying sensible pre-1980 beliefs for the cross-study variance σ_u^2 is more difficult, as there is little available information on physicians' pre-1980 views about the degree of cross-study dispersion in the risk of rupture. As a straightforward, if admittedly somewhat unattractive shortcut, I center the prior on the estimate of σ_u^2 that was obtained in the full sample using a diffuse prior. This may be interpreted, roughly speaking, as a sort of "perfect foresight" assumption. Formally, I parametrize the pre-1980 beliefs for σ_u^2 using a scaled inverse χ^2 distribution, which are commonly used to specify prior beliefs for variance parameters. The scaled inverse χ^2 distribution has a location parameter τ^2 and a "degrees of freedom" parameter ν, and it arises naturally as the posterior distribution for the unknown variance of a normal distribution after observing a sample of size ν with sample variance τ^2, starting from a specific minimally-informative prior (Gelman et al., 2004). I set $\tau^2 = 0.49 = 0.7^2$, and I set $\nu = 20$, reflecting only a moderate belief that the variance is close to this value.

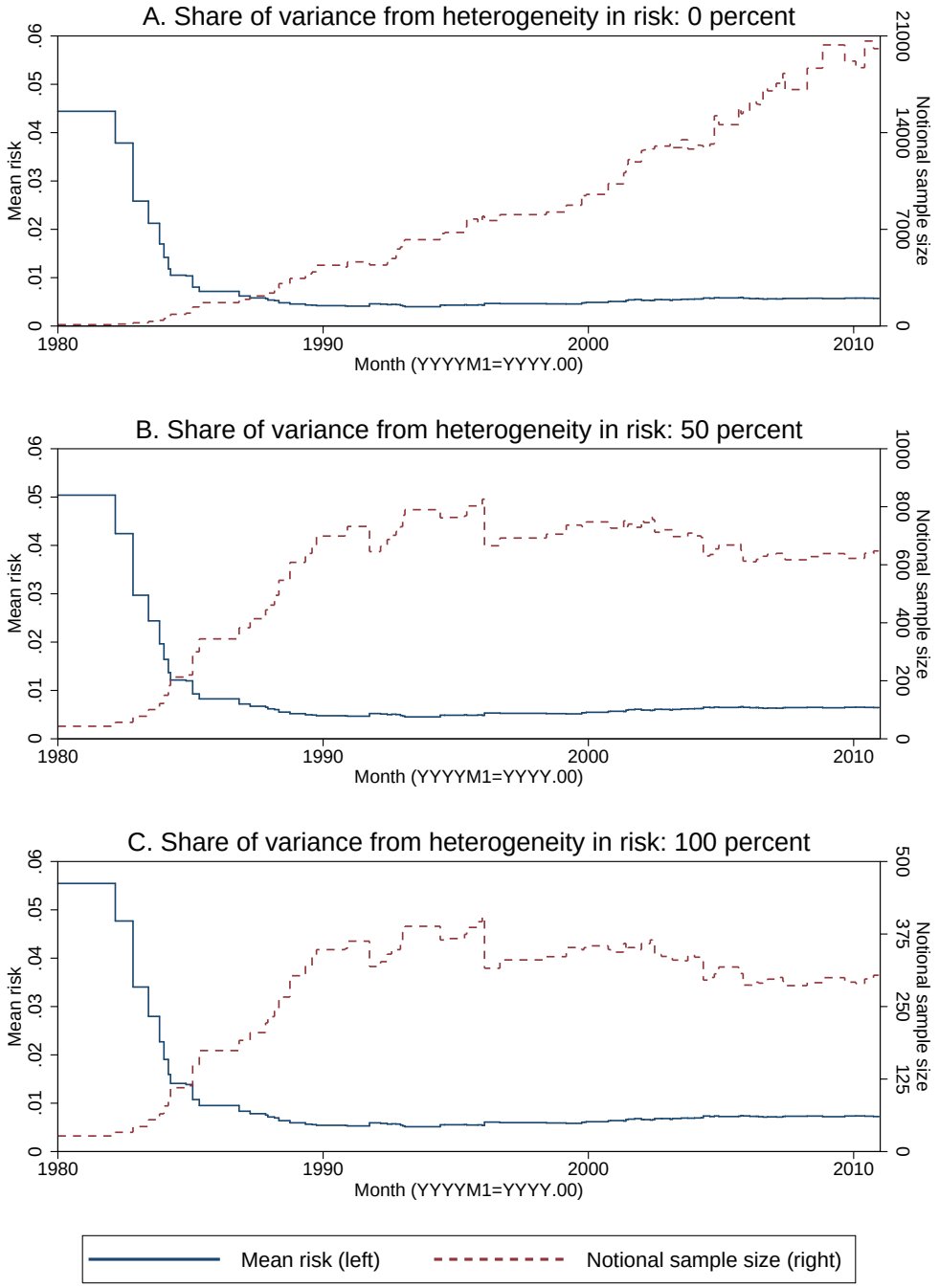
[35] Some evidence on the risk of uterine rupture for women whose prior cesarean had used the modern low-transverse incision rather than a classical incision was available at this time, largely from the United Kingdom. Lavin et al. (1982) surveys this evidence and arrives at an overall estimate of the risk of rupture of 0.7 percent, similar to the estimates obtained using modern evidence. I am assuming that this evidence was not well known in the United States and did not have an important influence on the pre-1980 beliefs of US physicians.

Using these pre-1980 beliefs as the prior for estimation, I once again estimate the model in equation (1.8). The resulting estimates of the common prior are reported in Figure 1.11. I report estimates separately under three values for δ, the share of σ_u^2 that represents true heterogeneity in risk: 0, 0.5, and 1. The evolution of the mean of the common prior appears to be basically insensitive to the assumed value of δ; it falls steadily and rapidly through the 1980s under all three assumptions and then rises very slightly thereafter. This time pattern is intriguing in light of the time series pattern of TOLAC and VBAC rates documented at the outset of the paper, and I return to the ability of these changes in mean beliefs to explain changes in practice in the next section.

While the evolution of mean beliefs is relatively insensitive to assumptions about heterogeneity, the notional sample sizes evolve quite differently depending on the value of δ chosen, which is consistent with the results of the last subsection. With larger values of δ, the notional sample sizes are relatively small and flatten out by 1990. When $\delta = 0$, however, the notional sample size is much larger and continues rising over the full period. Intuitively, this difference arises because, when the risk heterogeneity is substantial, the dispersion in the prior comes to be dominated by the risk heterogeneity rather than sampling error. As a result, the accumulation of additional evidence after 1990 has little to no effect on the notional sample size. Without such heterogeneity, the dispersion in the common prior consists solely of sampling error, so the accumulation of additional evidence drives a steady increase in the precision of the common prior.

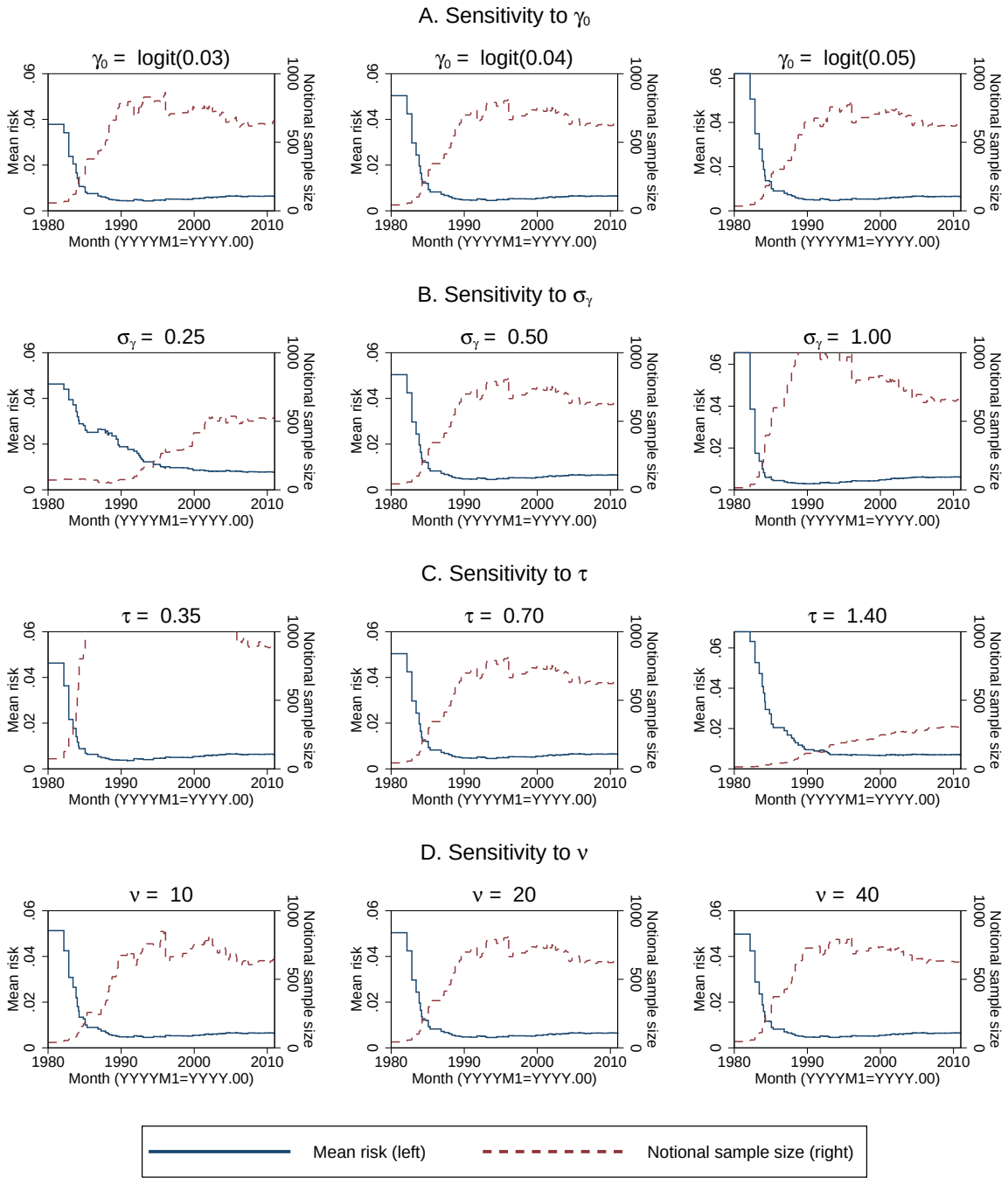
Before concluding this section, I examine these results' sensitivity to alternative parametrizations of the pre-1980 beliefs. To streamline the presentation, however, I focus solely on the case in which $\delta = 0.5$. The results of these sensitivity analyses are reported in Figure 1.12. In each row, the middle plot depicts the base parametrization of the pre-1980 beliefs, and the plots to each side depict the effect of varying the listed parameter. While none of these alternative parametrizations of the pre-1980 beliefs change the qualitative conclusions of the analysis, varying some of the parameters does have interesting effects. When providers are made meaningfully more certain about the pre-1980 median risk of rupture (i.e. when σ_γ is smaller), the mean risk declines much more slowly; the notional sample size also rises more slowly since the model "fits" the data by concluding that σ_u^2 is very large. The prior belief about the magnitude of σ_u^2 also has a meaningful effect on the notional sample size, and the pattern of effects is quite intuitive.

Figure 1.11: Estimated common prior for 1980-2010 obtained from a rolling meta-analysis by the share of cross-study variance due to heterogeneity in risk



Notes: Each panel of the figure reports the results of estimating a "best fit" beta distribution to the posterior distribution of p_h for each month and for three values of δ. The "best fit" beta distribution is estimated by maximum likelihood on a sample consisting of 100 random draws of p_h for each of the 2,000 posterior draws of (γ, σ_u^2) using the listed value of δ. The underlying draws of (γ, σ_u^2) are obtained from Bayesian estimation of equation (1.8) in each month in which a new study is published. The sample of published studies used for estimation is described in Section 1.3. The informative prior used for γ and σ_u^2 is described in the text. See text for additional estimation details.

Figure 1.12: Sensitivity of the common prior to alternative pre-1980 beliefs when 50 percent of the cross-study variance reflects heterogeneity in risk



Notes: Each panel of the figure reports the results of estimating a "best fit" beta distribution to the posterior distribution of p_h for each month using the listed parametrization of pre-1980 beliefs. The "best fit" beta distribution is estimated by maximum likelihood on a sample consisting of 100 random draws of p_h for each of the 2,000 posterior draws of (γ, σ_u^2) using $\delta = 0.5$. The underlying draws of (γ, σ_u^2) are obtained from Bayesian estimation of equation (1.8) in each month in which a new study is published. The sample of published studies used for estimation is described in Section 1.3. The informative prior used for γ and σ_u^2 is described in the text. See text for additional estimation details.

1.6 Learning model calibration

In this section, I use the event study estimates obtained in Section 1.4 together with the estimates of the common prior obtained in the last section to calibrate the model developed in Section 1.2. The first part of this section describes the approach I take in calibrating the model. In the second part of the section, I simulate the model for the period 1980-2010, and I examine whether it can quantitatively account for the dramatic variation in TOLAC rates in the cross-section and the rise and fall of VBAC and TOLAC rates over recent decades.

1.6.1 Calibration and simulation

I calibrate and simulate a discretized version of the model presented in Section 1.2 in which time evolves at a monthly frequency. All simulated results reflect a population of 1000 ex ante identical providers over the period 1980-2010. Simulating the model requires calibrating the following objects: the common prior parameters $\alpha(t)$ and $\beta(t)$, the true cross-provider risk distribution $G(p)$, the patient arrival rate $\lambda(t)$, and the distribution $F(p^*)$ of maternal risk thresholds above which cesarean section is preferred to TOALC and below which TOLAC is preferred to cesarean section.

I obtain the common prior parameters $\alpha(t)$ and $\beta(t)$ directly from the last section. I select the version of the common prior estimates in which providers believe that 50 percent of the cross-study variance in reported risks arises from true provider-to-provider variation in risk (i.e. $\delta = 0.5$).[36]

For the cross-provider risk distribution $G(p)$, I present two alternative calibrations. The first calibration assumes that the common prior at the end of 2010 accurately reflects the cross-hospital risk distribution. Because the common prior is non-degenerate, this calibration implies that providers differ substantially in their true risk of rupture. In the alternative calibration, I assume that each provider has the same risk of rupture and that this common risk is equal to the mean of the common prior at the end of the period. Comparing these alternative calibrations provides insight into how much of the cross-sectional variation in TOLAC rates generated by the model is due to true heterogeneity in risk and how much is due to differences in providers' idiosyncratic experiences.

Calibrating the patient arrival rate $\lambda(t)$ requires making a decision about whether a "provider"

[36]The assumption that $\delta = 0.5$ corresponds approximately to the case in which providers are perfectly ignorant about δ and thus have a uniform $[0, 1]$ prior on this parameter.

in the terms of the model is a physician or a whole hospital and whether each provider learns only from its own cases or also learns from cases in some broader community. As discussed in detail in Section 1.4, occurrence of a uterine rupture event seems to affect all physicians at a hospital, not just the physician who directly experiences the event, so it is reasonable to assume that learning is occurring at the hospital level or above. I thus assume in my initial calibrations that each provider is a single hospital that learns only from its own experience. In other calibrations, I assume that learning occurs at the "community" level, which I suppose (somewhat arbitrarily) consists of four hospitals. These four hospitals could literally be interpreted as four distinct hospitals that pool information or, possibly more realistically, as representing each provider's broad network of contacts.

I obtain values of the patient arrival rate $\lambda(t)$ that are appropriate for a single hospital in two steps. I first estimate the annual delivery volume of the hospital that handles the median delivery (which I define by ranking deliveries by the volume of the hospital at which they occur). Using the Nationwide Inpatient Sample for 1988-2010, I obtain an estimate of 2,144. I then use the National Hospital Discharge Survey to estimate the share of deliveries in which the mother had a prior cesarean section for each year 1980-2010. Using these estimates, I compute $\lambda(t)$ for each month t as the product of the delivery volume from the NIS (divided by 12) and the prior cesarean section share appropriate for that year. For calibrations using a four-hospital community, I simply scale the result by a factor of four.

I turn now to the most important step in the calibration: calibrating the distribution of maternal risk thresholds $F(p^*)$. I assume that $F(p^*)$ has the form of a beta distribution with parameters (μ^*, ϕ^*), where μ^* is the distribution mean and ϕ^* is the notional sample size. The mean μ^* controls the overall risk tolerance of the population and, thus, the overall level of the TOLAC rate. The notional sample size ϕ^* controls the sensitivity of the TOLAC rate to changes in beliefs about the risk of rupture. If ϕ^* is very large, then the distribution of risk thresholds will be very peaked and small changes in beliefs will generate large changes in TOLAC rates. In contrast, if ϕ^* is small, then the distribution of thresholds will be very diffuse and TOLAC rates will be relatively insensitive to changes in beliefs.

I choose the parameters (μ^*, ϕ^*) so that the simulated data match two empirical moments: (1) the average TOLAC rate over the period 1980-2010; and (2) the event study estimate of the

effect of a uterine rupture event on the hospital-level TOLAC rate. Roughly speaking, the average TOLAC rate over the period determines μ^* and the average effect of an event on the TOLAC rate determines ϕ^*.

I calculate the empirical version of the first of these moments from the NHDS and find that it is equal to 24 percent. I compute its simulated counterpart in the natural way. Turning to the second of these moments, I obtain the empirical version of the empirical event study estimate from Table 1.2. I neglect the time path of the response and thus select the "long-run" effect 6 quarters after the event, which is -0.72 percent. I then scale this estimate up by a factor $1/0.65 \approx 1.5$ to account for the fact, discussed in Section 1.3 and Appendix A.6, that some coded uterine rupture events are not true events, which generates a final event study estimate of -1.11 percentage points. As demonstrated carefully in Section 1.4, the event study estimates are "treatment on the treated" estimates that correspond to the particular period covered by the event study sample. To compute a corresponding moment in the simulated data, I identify, in each month t, the full set of women who deliver at providers that experienced at least one rupture in month $t-1$. Among those women, I then calculate the share that would have made an attempt of labor but for the uterine rupture events that occurred in month $t-1$; this is the simulated average treatment effect on the treated in month t. I then weight these treatment effects across months t according to the actual number of deliveries included in the event study analysis in that month.[37]

To identify the particular values of μ^* and ϕ^* that reproduce these moments, I use the Nelder-Mead simplex algorithm to minimize the following objective function:

$$f(\mu^*, \phi^*) = \sum_{j \in \{1,2\}} \frac{(m_j^S(\mu^*, \phi^*) - m_j)^2}{m_j^2},$$

where $m_j^S(\mu^*, \phi^*)$ is the simulated version of moment j and m_j is its empirical counterpart.

Following standard practice (e.g. Davidson and MacKinnon (2004)), I use the same random draws when simulating the model for each new set of parameters in order to ensure that $m_j^S(\mu^*, \phi^*)$ varies in a predictable fashion as the parameter vector (μ^*, ϕ^*) varies. Retaining the provider rupture risk and the number of new patient arrivals in each period is straightforward since the distributions of these quantities are not functions of the parameter vector (μ^*, ϕ^*). The joint

[37] In fact, time is observed at quarterly frequency in event study sample, so the weight applied to each month is the number of deliveries in the corresponding quarter divided by 3.

distribution of the number of attempts of labor and the number of uterine ruptures in each period is, however, a function of (μ^*, ϕ^*). As a consequence, these draws cannot be retained directly. Instead, I retain an underlying uniform random variable on $[0, 1]$ for each of these quantities and obtain the draws for each period by applying the inverse cumulative distribution functions appropriate for the current vector (μ^*, ϕ^*) to these underlying variables.

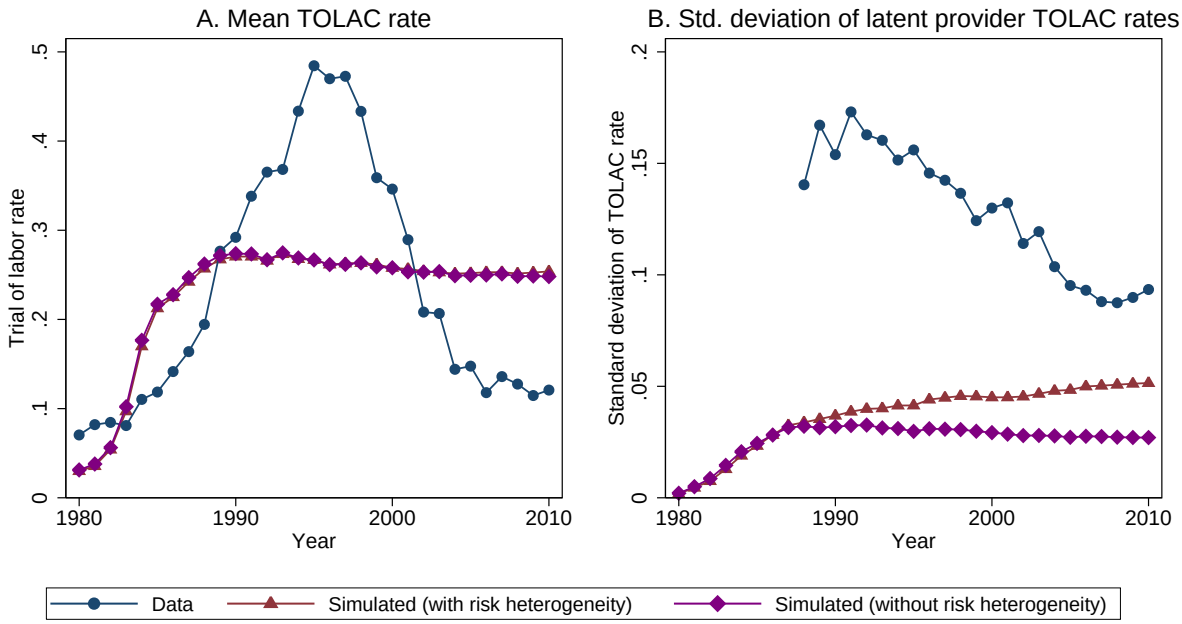
Simulated in this fashion and with appropriate selection of starting values, the optimization algorithm converges to an exact solution (i.e. values such that $f(\mu^*, \phi^*) \approx 0$) after several minutes of runtime.

1.6.2 Simulation results

I now present the results of calibrating and simulating the learning model under several different sets of assumptions. To evaluate the results, I compare the simulated mean TOLAC rates and cross-provider dispersion in TOLAC rates to their actual counterparts. Specifically, for mean TOLAC rates, I compare the simulated mean TOLAC rate to its actual counterpart as computed using the National Hospital Discharge Survey. For cross-hospital dispersion, I focus on the standard deviation of the cross-provider distribution of "latent" TOLAC rates – the probability that a randomly chosen woman delivering with that provider would undergo TOLAC – in order to abstract from small-sample variation. I compute the simulated standard deviation in the obvious way. I compute the actual standard deviation by estimating the beta-binomial mixture model described in Appendix A.1 using the Nationwide Inpatient Sample for 1980-1987. Because the Nationwide Inpatient Sample is not available before 1988, I am not able to document the true level of cross-provider dispersion in 1987 or earlier.

In my first set of results, I simulate the model under the assumption that physicians learn only from deliveries at their own hospital. The results are plotted in Figure 1.13. The simulated and actual TOLAC rates agree well through the 1980s, but the simulated rate then flattens out at a level well below the actual peak and completely misses the late-1990s decline in TOLAC rates. The initial rise in the TOLAC rate results from the fact that the mean of providers' common prior for the risk of rupture falls as new clinical evidence becomes available through the early 1980s; this evolution in the common prior was depicted in the last section in Figure 1.11. Once the changes in the common prior cease, however, there is no force in the model to drive additional changes

Figure 1.13: Level of and cross-provider variation in the TOLAC rate: Actual vs. simulated (within-hospital learning only)



Notes: The simulated data are obtained by simulating the learning model as described in the text when physicians learn from cases at their hospital only. In the simulated series labeled "with risk heterogeneity," each hospital's risk of rupture is drawn according to the common prior distribution for the last month of the data. In the simulated series labeled "without risk heterogeneity," each hospital's risk of rupture is identical and set equal to the mean of the common prior distribution for the last month of the data. The actual mean TOLAC rate is computed using the National Hospital Discharge Survey for the appropriate year. The actual standard deviation of the latent TOLAC rate is computed from the Nationwide Inpatient Sample by estimating the beta-binomial mixture model described in Appendix A.1. The plotted standard deviation is the standard deviation of the estimated beta distribution from this model.

in practice. In particular, because the mean true risk of rupture coincides with the mean of the common prior at the end of the period, the experience providers accrue after 1990 will (on average) match their pre-existing beliefs.

Turning to the results for cross-hospital dispersion, Figure 1.13 demonstrates that cross-hospital dispersion on the order of one quarter of the actual level of cross-hospital dispersion appears quickly once providers start gaining experience with TOLAC. Perhaps surprisingly, most of this variation exists even in simulations in which all hospitals face the same risk of rupture. Furthermore, the figure shows that variation from this source does not dissipate quickly over time.

In Figure 1.14, I examine how these results change if instead of learning only from the cases at a single hospital, physicians learn from all cases in a four-hospital community. This second set of results exhibits a much greater amount of cross-hospital dispersion. When all providers face the

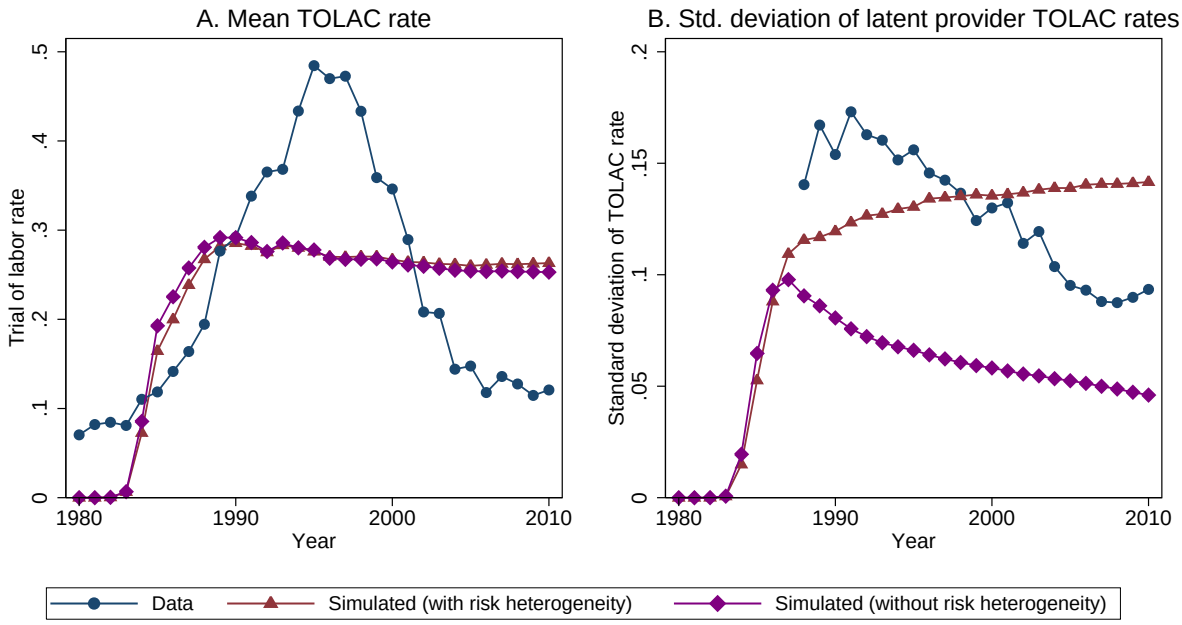
same risk of rupture, the model can now explain roughly half of the actual cross-hospital standard deviation of TOLAC rates, and it remains the case that this source of cross-sectional variation in TOLAC rates is quite durable. Adding true cross-hospital variation in the risk of rupture increases the simulated cross-hospital variation in the TOLAC rate further still, to the extent that the model can approximately match the amount of cross-sectional variation observed in the data.

Embedding physicians in a broader community changes the simulated amount of cross-sectional variation for two reasons. First, as shown in columns (1) and (2) of Table 1.9, the calibrated preference parameters change so as to increase the sensitivity of TOLAC rates to changes in beliefs, which magnifies any given degree of cross-sectional variation in beliefs. Intuitively, when embedded in a broader community, physicians accrue experience at a greater rate, which causes any single event to have a smaller effect on beliefs, and the only way for the model to continue to match the event study estimate is for the sensitivity of the TOLAC rate to beliefs to rise. Second, as also shown in Table 1.9, increasing the rate at which physicians gain experience changes the amount of cross-sectional variation in physician beliefs, although the direction of this effect depends on whether there is true heterogeneity in risk. The intuition for this comes from equation (1.4), which shows that variation in beliefs attributable to differences in risk rises with experience, while variation in beliefs attributable to differences in idiosyncratic experience initially rises with experience, then falls. In the simulations with true heterogeneity in risk, the former effect dominates, while in the simulations without true heterogeneity, only the latter effect exists.

While embedding physicians in a broader community better matches the amount of cross-sectional variation in TOLAC rates, the model still fails to match the decline in the TOLAC rate that starts in the mid-1990s, for essentially the same reasons as before. In an attempt to improve the model's ability to explain this feature of the data, I suppose that physicians exhibit a form of "selective memory" in which they fully remember incidents of rupture, but sometimes "forget" TOLACs that do not lead to rupture. Concretely, I modify the beta-binomial learning rule in equation (1.1) so that it takes the following form:

$$\tilde{p}_h(t) = \frac{\alpha(t) + R_h(t)}{\alpha(t) + \beta(t) + R_h(t) + (1-\theta)[L_h(t) - R_h(t)]}, \tag{1.9}$$

Figure 1.14: Level of and cross-provider variation in the TOLAC rate: Actual vs. simulated (with learning in a four-hospital community)



Notes: The simulated data are obtained by simulating the learning model as described in the text when physicians learn from hospitals in a four-hospital community. In the simulated series labeled "with risk heterogeneity," each hospital's risk of rupture is drawn according to the common prior distribution for the last month of the data. In the simulated series labeled "without risk heterogeneity," each hospital's risk of rupture is identical and set equal to the mean of the common prior distribution for the last month of the data. The actual mean TOLAC rate is computed using the National Hospital Discharge Survey for the appropriate year. The actual standard deviation of the latent TOLAC rate is computed from the Nationwide Inpatient Sample by estimating the beta-binomial mixture model described in Appendix A.1. The plotted standard deviation is the standard deviation of the estimated beta distribution from this model.

Table 1.9: Characteristics of several calibration scenarios

Calibration scenario:	(1)	(2)	(3)	(4)	(5)
A. With heterogeneity in the risk of rupture					
Parameters of the risk threshold distribution					
Mean (μ^*)	0.007	0.004	0.016	0.013	0.019
Notional sample size (ϕ^*)	27	186	78	134	58
Implied sensitivity to 1 in 1,000 change in risk	0.019	0.050	0.017	0.026	0.013
Simulated provider beliefs, Jan. 1995					
Mean	0.0055	0.0059	0.0214	0.0155	0.0262
Standard deviation	0.0019	0.0024	0.0061	0.0030	0.0094
B. Without heterogeneity in the risk of rupture					
Parameters of the risk threshold distribution					
Mean (μ^*)	0.007	0.005	0.020	0.014	0.025
Notional sample size (ϕ^*)	26	202	96	146	76
Implied sensitivity to 1 in 1,000 change in risk	0.018	0.050	0.017	0.026	0.014
Simulated provider beliefs, Jan. 1995					
Mean	0.0056	0.0062	0.0248	0.0166	0.0317
Standard deviation	0.0014	0.0012	0.0027	0.0017	0.0037
Scenario characteristics					
Community size (# of hospitals)	1	4	4	4	4
Selective memory parameter (θ)	0	0	0.9	0.9	0.9
Risk share of cross-study variance (δ)	0.50	0.50	0.50	0.25	0.75

Notes: This table reports the calibrated parameter vector (μ^*, ϕ^*) of the distribution of patient risk thresholds above which a woman elects repeat cesarean and below which she elects TOLAC under several scenarios. The parameters of these scenarios are described in the footer of the table. Panel A reports a version of each scenario in which each hospital's risk of rupture is drawn according to the common prior distribution for the last month of the data. Panel B reports a version of each scenario in which each hospital's risk of rupture is identical and set equal to the mean of the common prior distribution for the last month of the data. The implied sensitivity to a 1 in 1,000 change in practice is computed as one thousandth of the density of the risk threshold distribution, evaluated at the risk level that generates the period mean TOLAC rate of 24 percent. It can be interpreted as the increase in the TOLAC rate that would arise from a 1 in 1,000 reduction in the provider's estimated risk of rupture when starting from TOLAC rates near the full period average.

where θ is the share of non-ruptures that the provider forgets.[38]

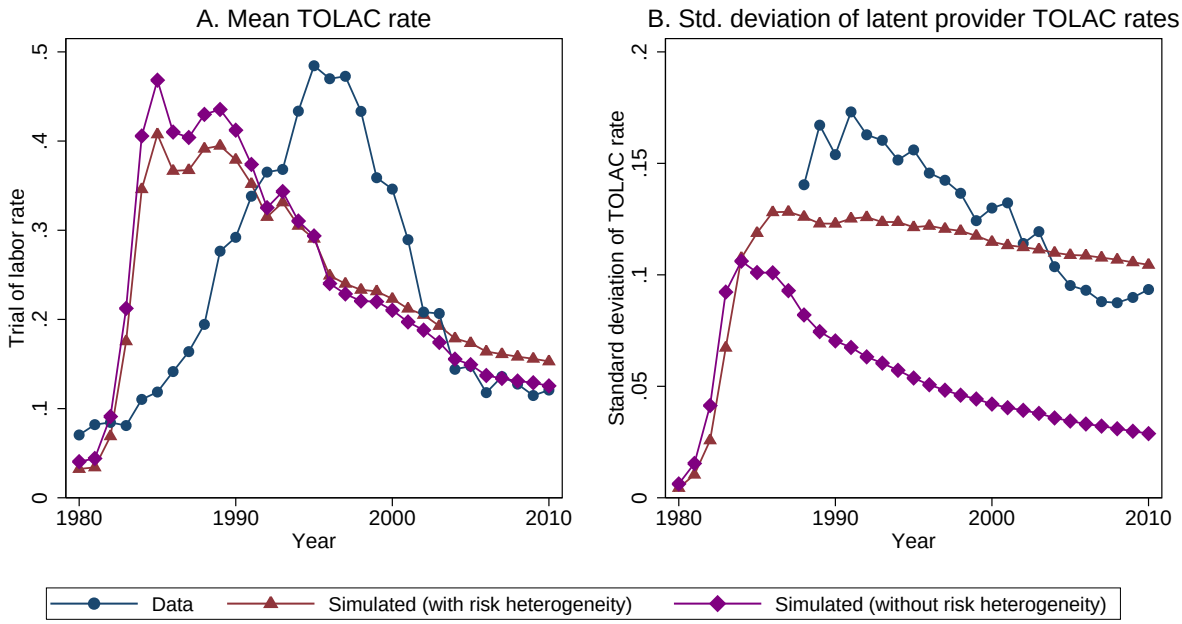
There are three reasons one might expect the learning process to exhibit selective memory of this form. The first is that when drawing upon directly-observed experience, physicians may apply an "availability heuristic," as proposed by Tversky and Kahneman (1973). By virtue of its rarity and severity, an incident of uterine rupture is likely to be a high-salience event. By contrast, most TOLACs that proceed without rupture will be little different from many other TOLACs and, for that matter, little different from most deliveries by women without a prior cesarean delivery. As a consequence, "availability heuristic" behavior would imply that incidents of rupture will be remembered well, while non-rupture TOLACs may not be. Second, for the simulations in which learning occurs at the "community level," the most natural interpretation is that much of this experience is accruing through social networks. For similar reasons, it seems likely that physicians will more frequently share stories about cases of rupture than about labors in which nothing unusual occurs, which will generate a skewed information flow. Finally, most hospitals institutionalize a focus on cases involving significant complications through regular "morbidity and mortality" conferences in which such cases are discussed (Gawande, 2002; Pierluissi et al., 2003; Orlander et al., 2002). This suggests that cases of rupture are more likely to be publicized and remembered at the hospital level.

As discussed previously, Choudhry et al. (2006) also appeal to an availability heuristic in explaining their finding that cardiologists become less likely to prescribe warfarin for atrial fibrillation after experiencing a warfarin-induced bleeding complication. It is important to note, however, that the type of imperfect memory being proposed here is quite different. Whereas Choudhry et al. propose that physicians overweight bad outcomes, I am proposing that physicians underweight the much more common good outcomes. Notably, my approach does not imply that observed responses to adverse events will be substantially larger than under a rational model.

Figure 1.15 reports the results of calibrating and simulating the model with a selective memory parameter $\theta = 0.9$, corresponding to the case where physicians remember only 10 percent of attempts of labor that do not lead to rupture. Addition of selective memory dramatically improves

[38]This approach to incorporating selective memory is, admittedly, ad hoc. A possibly more aesthetically appealing approach would be to assume that, immediately after each trial of labor that does not end in rupture, the memory of that attempt is added to the provider's stock of experience with some probability θ. In practice, this gives extremely similar results, and because the approach taken here modestly reduces the computational burden, I prefer it.

Figure 1.15: Level of and cross-provider variation in the TOLAC rate: Actual vs. simulated (with a four-hospital community and "selective memory")



Notes: The simulated data are obtained by simulating the learning model as described in the text when physicians learn from hospitals in a four-hospital community and have selective memory parameter $\theta = 0.9$. In the simulated series labeled "with risk heterogeneity," each hospital's risk of rupture is drawn from the common prior distribution for the last month of the data. In the simulated series labeled "without risk heterogeneity," each hospital's risk of rupture is identical and set equal to the mean of the common prior distribution for the last month of the data. The actual mean TOLAC rate is computed using the National Hospital Discharge Survey for the appropriate year. The actual standard deviation of the latent TOLAC rate is computed from the Nationwide Inpatient Sample by estimating the beta-binomial mixture model described in Appendix A.1. The plotted standard deviation is the standard deviation of the estimated beta distribution from this model.

the ability of the model to fit the time series. After the rise through the 1980s, TOLAC rates flatten out, then decline steadily through the remainder of the period, reaching levels similar to the actual levels by 2010. The improvement in fit comes because the experience physicians accrue is skewed to include more cases of rupture than actually occur. As a result, beliefs about the risk of rupture drift upward over time, reaching $p_h/(1-\theta)$ in the limit for a provider with true risk p_h. This upward drift in beliefs drives the downward movement in TOLAC rates. Of course, the time series fit is still not perfect, as the rise during the 1980s is still too rapid. It appears that, while the model interprets new clinical publications as being absorbed into the common prior immediately, this process takes some time in reality.

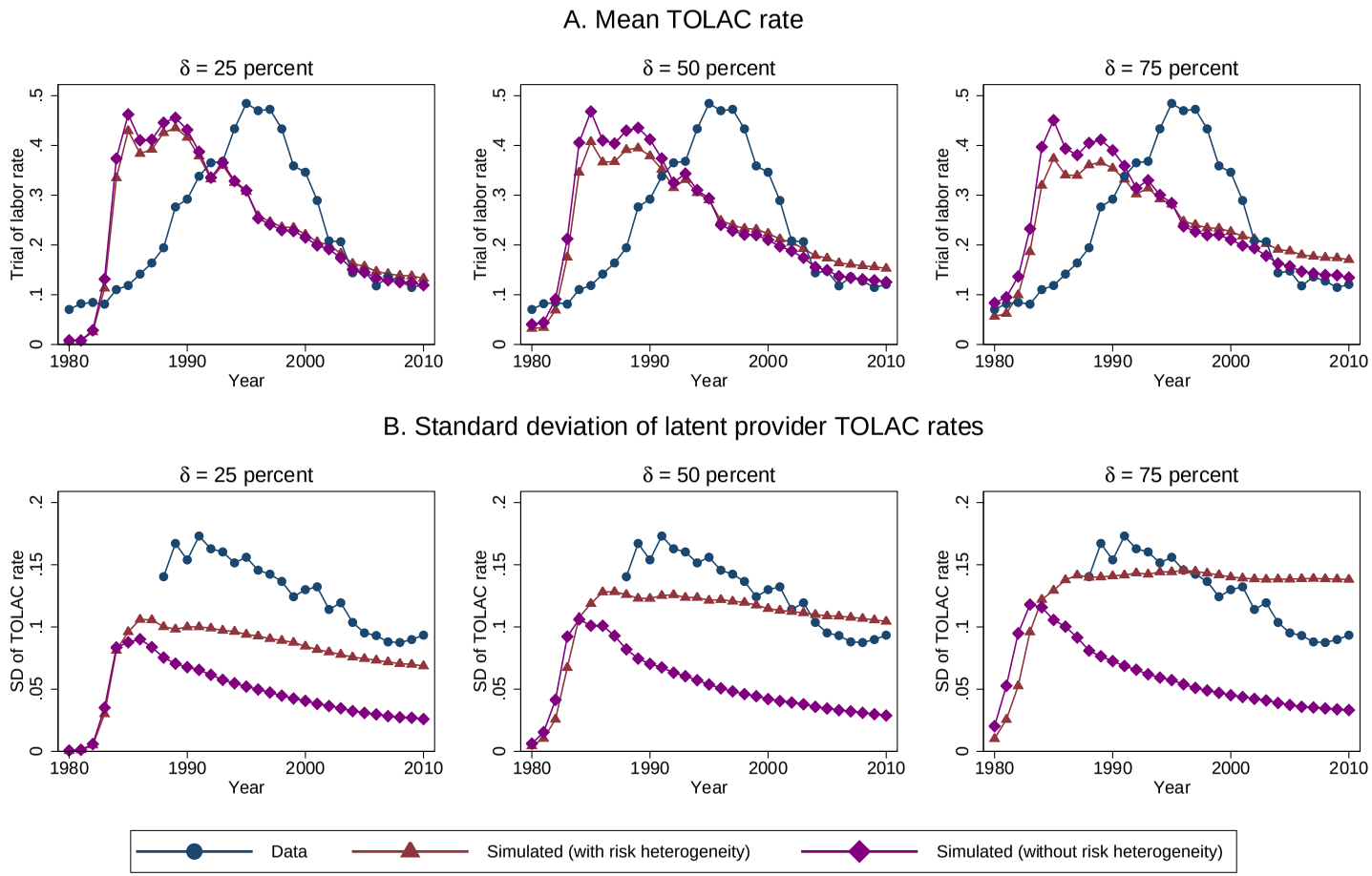
In addition to matching the time series, the version of the model with selective memory still

generates substantial cross-provider variation in TOLAC rates, including in simulations in which all providers share a common risk of rupture. Furthermore, when there is heterogeneity in risk, the model continues to approximately match the actual degree of cross-sectional variation. The mechanism by which the model with selective memory generates this level of cross-sectional variation differs somewhat from the version of the model without selective memory, however. Comparing column (2) to column (3) in Table 1.9 demonstrates that the sensitivity of the TOLAC rate to a given change in beliefs is now considerably smaller, while the cross-sectional variance in beliefs is considerably larger. This reduction in the sensitivity of practice to beliefs arises because physicians' "effective" experience is now given by $R_h(t)+(1-\theta)[L_h(t)-R_h(T)]$ rather than $L_h(t)$, so experience accrues at a considerably slower rate. Since beliefs change more in response to any particular event when experience is lower, matching the event study estimates now requires preferences to be less sensitive to beliefs. Turning to the increase in the cross-provider variance in beliefs, there are two effects to consider. The presence of selective memory slows the rate at which the physician accrues experience and approximately proportionally increases the share of experience that consists of cases of rupture. Equation (1.4) indicates that that the increase in the effective probability of rupture will increase the variance of beliefs, and this will either be reinforced or only partially offset by the reduction in the rate at which experience accrues, so the variance in beliefs rises.[39]

Before closing this section, I examine how the conclusions above would change if the model was calibrated using a version of the common prior that ascribed a smaller or larger share of the cross-study variance in the estimated risk of uterine rupture to true heterogeneity in risk. Thus, Figure 1.16 explores the sensitivity of the last calibration exercise (i.e. that with a four-hospital community and selective memory) to setting $\delta = 0.25$ or $\delta = 0.75$ instead of the value $\delta = 0.5$ used in the main results. Changing δ has almost no effect on the time series results. Changing δ does, unsurprisingly, change the amount of cross-sectional variation in simulations that include true heterogeneity in risk, but for all values of δ, the simulations continue to approximately match the amount of cross-sectional variation observed in the data.

[39]Strictly speaking, equation (1.4) does not apply to the model with selective memory. However, under the version of selective memory described in footnote 38, which gives results very similar to the version of selective memory actually simulated, a counterpart to equation (1.4) can be derived that provides the same intuition.

Figure 1.16: Mean and cross-hospital standard deviation of the TOLAC rate: Actual vs. simulated (with a four-hospital community and "selective memory" and different assumptions about the share δ of the cross-study variance in the risk of rupture that reflects true heterogeneity in risk)



Notes: The simulated data are obtained by simulating the learning model as described in the text when physicians learn from hospitals in a four-hospital community and have selective memory parameter $\theta = 0.9$ under different assumptions about the share δ of the cross-study variance in risk estimates that reflects true heterogeneity in risk. In the simulated series labeled "with risk heterogeneity," each hospital's risk of rupture is drawn from the common prior distribution for the last month of the data. In the simulated series labeled "without risk heterogeneity," each hospital's risk of rupture is identical and set equal to the mean of the common prior distribution for the last month of the data. The actual mean TOLAC rate is computed using the National Hospital Discharge Survey for the appropriate year. The actual standard deviation of the latent TOLAC rate is computed from the Nationwide Inpatient Sample by estimating the beta-binomial mixture model defined in equation (A.1). The plotted standard deviation is the standard deviation of the estimated beta distribution from this model.

1.7 Effects of high-profile events

The last section established that a model based on gradual learning from the clinical literature and individual experience can account for much or all of the cross-sectional variation in TOLAC rates as well as for the most salient features of the pattern of TOLAC rates over time. In this section, I examine an alternative theory of the evolution in TOLAC rates: that it was driven by a small number of high-profile events, namely, the publication of several high profile papers and the release of new practice guidelines by the American College of Obstetricians and Gynecologists (ACOG). Guise et al. (2010b), MacDorman et al. (2011), Zinberg (2001), and Santerre (1996) all argue that events of this kind have had large effects on practice patterns over the period studied in this paper. One such event (the publication of Lydon-Rochelle et al. (2001)) has also been previously studied in the economics literature by Price and Simon (2009).

I study two specific sets of events: (1) publication of the five most highly-cited research articles on the appropriate management of labor and delivery for women with prior cesarean delivery appearing during the period 1980-2010; and (2) all releases of revised practice guidelines by ACOG. I identify the highly-cited studies using the comprehensive database of clinical articles on this topic that I described in Section 1.3.[40] In ranking articles, I use citation counts as reported in the Thomson Reuters Web of Science citation index. The five articles selected by this strategy are Lavin et al. (1982), Flamm et al. (1994), McMahon et al. (1996), Lydon-Rochelle et al. (2001), and Landon et al. (2004). This set includes the two articles (McMahon et al. (1996) and Lydon-Rochelle et al. (2001)) that are frequently identified as having had an important effect on physician attitudes (see, for example, Guise et al. (2010b) and MacDorman et al. (2011)). A brief description of each of these articles is provided in Table 1.10.

The ACOG guidelines I examine were published in 1982, 1984, 1988, 1994, 1995, 1998, 1999, and 2010 (ACOG, 1982; 1984; 1988; 1994; 1995; 1998; 1999; 2010b). In general, the early guidelines progressively broadened the class of women for whom a trial of labor could be considered an appropriate option.[41] This process culminated in the 1994 and 1995 guidelines, which state that

[40] Note that, for this exercise, I include articles that do not report original case series; in particular, the article by Lavin et al. (1982) is a meta-analysis of earlier case series.

[41] As of this writing, I have been unable to obtain copies of ACOG guidelines published before 1994. This description is based upon the description provided in Zinberg (2001).

Table 1.10: Summaries of the five most-cited VBAC/TOLAC publications, 1980-2010

Article	Month	Journal	Summary
Lavin et al. (1982)	February 1982	Obstetrics and Gynecology	Meta-analysis of the English language literature on VBAC/TOLAC from 1950-1980, mostly from outside US and 1970 or earlier. Concluded that TOLAC is "relatively safe."
Flamm et al. (1994)	June 1994	Obstetrics and Gynecology	Observational analysis of a large case series from a group of Kaiser Foundation hospitals. Estimated risk of rupture was 0.8%. Concluded that TOLAC is an appropriate management option.
McMahon et al. (1996)	September 1996	New England Journal of Medicine	Observational analysis of a large case series from Nova Scotia. Estimated that "major maternal complications are twice as likely" among women undergoing TOLAC.
Lydon-Rochelle et al. (2001)	July 2001	New England Journal of Medicine	Observational analysis of all births to women with a prior cesarean section in Washington state for 1986-1997. Estimated risk of uterine rupture was 0.7%, higher when labor is induced. Found increased risk of neonatal death with TOLAC. Accompanied by an editorial by Greene (2001) arguing that repeat cesarean section is "unequivocally" safer for infant.
Landon et al. (2004)	December 2004	New England Journal of Medicine	Observational analysis of all births to women with a prior cesarean section at 19 academic medical centers. Estimated risk of uterine rupture was 0.7%. Concluded that TOLAC carries increased neonatal risk, but that absolute risks are small.

TOLAC should be "encouraged" for a relatively broad class of women. The 1998 and 1999 guidelines represent a break in this trend of increasing ACOG support for TOLAC and state that women should be "offered" a trial of labor, but also that TOLAC is only appropriate in settings that can provide a prompt cesarean section, a standard many smaller hospitals are thought to have difficulty achieving. Accounts of trends in VBAC rates during this period frequently state that these guideline revisions – particularly the requirements on the availability of immediate cesarean section – played an important causal role in the late-1990s decline in VBAC rates (Guise et al., 2010b; MacDorman et al., 2011). The 2010 guideline is generally viewed as a slight reversal in course. While substantively similar to the 1999 guideline, it added language emphasizing patients' prerogative to accept any risks associated with VBAC. It also was accompanied by press materials from ACOG describing the guideline as "less restrictive" (ACOG, 2010a). Each individual guideline is briefly summarized in Table 1.11.

In the remainder of this section, I first describe the empirical approach I use to evaluate the effect of these events and then present my results.

1.7.1 Empirical specification

I estimate the effect of each event on practice by examining whether the event date is associated with a discontinuity in the level or the trend of the VBAC rate. Under the usual assumptions, the discontinuity in level can be interpreted as the short-run effect of the event. As shown below, the discontinuity in trend can be interpreted as the rate at which the effect of the guideline or article is growing or shrinking over time. For events occurring after 1990, I estimate these effects using the NVSS birth certificate data. For events before 1990, I use the NHDS. Because, as discussed previously, the NVSS only reports whether a mother had a vaginal delivery and not whether she attempted labor, I focus on effects on VBAC for this set of results, rather than effects on TOLAC.

To estimate the immediate effect of each article or guideline publication on practice style, I use a standard local linear regression discontinuity specification with a rectangular kernel as suggested by Imbens and Lemieux (2008). That is, I run the following specification:

$$y_{tj} = \alpha + d_t \cdot \mathbf{1}\{d_t \leq 0\}\beta_0 + d_t \cdot \mathbf{1}\{d_t > 0\}\beta_1 + \mathbf{1}\{d_t > 0\}\tau + \epsilon_{tj}, \tag{1.10}$$

where t indexes time and j patients, y_{tj} is an indicator for whether the woman delivered vaginally,

Table 1.11: Summaries of ACOG VBAC/TOLAC guidelines, 1980-2010

Publication year	Summary
1982	The full text of the guideline was not available at the time of writing. Zinberg (2001) describes the guideline as saying that "with careful selection of patients... vaginal delivery appeared to be an acceptable option."
1984	The full text of the guideline was not available at the time of writing. Zinberg (2001) states that the guideline broadened the group of patients considered appropriate for TOLAC.
1988	The full text of the guideline was not available at the time of writing. Zinberg (2001) states that the guideline broadened the group of patients considered appropriate for TOLAC.
1994	Recommended that "in the absence of a contraindication, a woman with one previous cesarean delivery... should be counseled and encouraged to undergo a trial of labor in her current pregnancy" (emphasis added).
1995	Reaffirmed the 1994 guideline and provided a slightly expanded evidence review.
1998	Recommended that "most women with one previous cesarean delivery... are candidates for VBAC and should be counseled about VBAC and offered a trial of labor" (emphasis added). Also recommended that "VBAC should be attempted in institutions equipped to respond to emergencies with physicians readily available to provide emergency care" (emphasis added).
1999	Reaffirmed the 1998 guideline, but strengthened the recommendation that emergency care be "readily available" to a recommendation that it be "immediately available."
2010	Made recommendations similar to those in 1999 guideline, but stated that patients "should be allowed" to undergo TOLAC even when emergency care is not "immediately available" if "clearly informed of such potential increase in risk." Accompanied by press release titled "Ob Gyns Issue Less Restrictive VBAC Guideline."

and d_t is the time relative to the publication date (measured in years).[42] The coefficient of interest is τ. The sample for estimating this regression is limited to women with a prior cesarean delivery for whom $|d_t| \leq h$ for some bandwidth h. I exclude observations with $d_t = 0$ since it is unknown whether births in this category occurred before or after the event. Regressions using the NHDS are weighted using the NHDS sampling weights. I follow the advice of Lee and Card (2008) for settings with discrete running variables and cluster my standard errors at the (monthly) resolution of the running variable. In practice, this adjustment has only a modest effect on the standard errors.

As depicted in Figure 1.17 in the next subsection, the time series of VBAC rates exhibits substantial curvature. While curvature does not alter the (first-order) asymptotic properties of local linear regression discontinuity estimators, it can generate substantial small sample bias if the total change in slope over the months falling inside the bandwidth is large. To evaluate whether small sample bias of this kind is a substantial concern in the present context, I explore the robustness of all results to an analogous local quadratic specification:

$$y_{tj} = \alpha + d_t \cdot \mathbf{1}\{d_t \leq 0\}\beta_0 + d_t \cdot \mathbf{1}\{d_t > 0\}\beta_1 + d_t^2 \cdot \mathbf{1}\{d_t \leq 0\}\gamma_0 + d_t^2 \cdot \mathbf{1}\{d_t > 0\}\gamma_1 + \mathbf{1}\{d_t > 0\}\tau + \epsilon_{tj}. \quad (1.11)$$

Once again, the coefficient of interest is τ. Implementation of this specification is otherwise identical to the local linear specification discussed above.

Due to the large differences in sample size, I use different bandwidths for events studied in the NHDS and those studied in the NVSS. For the NHDS regressions, I use a base bandwidth $h = 24$ months. For the NVSS regressions, I use a base bandwidth $h = 12$ months. I explore the sensitivity of all results to bandwidths twice as large and bandwidths half as large. Rather than using the same fixed bandwidth for all events, I considered computing the "optimal" bandwidth suggested by Imbens and Kalyanaraman (2012) for each event and then examining sensitivity relative to those bandwidths. The approach I use facilitates presentation of the results and comparison across specifications, so I prefer it. Regardless, the optimal bandwidths generally fall within the range of the bandwidths considered.

[42] Per the discussion of the NVSS data in section 1.3, starting in 2003, states begin adopting a revised birth certificate that appears to increase reported VBAC rates by 2.3 percentage points (see Section 1.3 for details). To adjust for this, I subtract 0.023 from y_{tj} for all births reported using the revised certificate.

A downside of the regression discontinuity estimates discussed above is that they estimate only the immediate effect of each event on practice; this immediate effect could either understate or overstate the long-run effect of each informational shock. For example, if some providers respond to the new information with a delay, then the effect of the event will grow over time and the level discontinuity estimate will understate the long-run effects on practice. On the other hand, an informational shock could simply accelerate changes that would have occurred over the ensuing months and years even in the absence of the shock, in which case the effect of the event will shrink over time and the level discontinuity estimate will overstate the long-run effect of publication. To shed light on this question, I estimate the discontinuity in the *slope* of the regression function at the time of the event. Under a plausible smoothness assumption, the discontinuity in the slope is equal to the rate at which the effect of the event is growing or shrinking over time (in a neighborhood of the event date).[43]

To formalize this approach, consider the following potential outcomes framework: for each unit, $Y(0)$ denotes the outcome in the counterfactual world where the event of interest never occurs; and $Y(1)$ denotes the observed outcome in the world where the event does occur. Each unit is observed at some time T. The event occurs at time $t = 0$, and I assume there is no anticipation of the event, so that $Y(0) = Y(1)$ for all units with $T < 0$. The identifying assumption is that, in the absence of the event, average outcomes would have been smooth at time t. That is, the counterfactual expectation $\mathbb{E}[Y(0) \,|\, T = t]$ is continuously differentiable in t. Likewise, I assume that its observed counterpart $\mathbb{E}[Y(1) \,|\, T = t]$ is continuously differentiable in t except possibly at $t = 0$.[44]

I state the key result emerging from this framework in the form of a lemma:

[43] The quantity estimated in this design – the instantaneous change in slope – is the same as in the "regression kink design" of Card et al. (2012), a research design that exploits a discrete jump in the rate of change of treatment intensity as a function of the running variable. While the appropriate estimators are similar between this case and the "regression kink" case, the underlying estimands of interest are distinct. Card et al. are interested in the effect of a marginal change in treatment on the outcome. In their case, therefore, a change in the level of the outcome at the kink point calls into question the validity of the design since any change in slope may reflect a change in who is "on the margin" on the two sides of the kink. By contrast, I am interested in the *total* effect on the outcome including any effect of initial exposure and cumulative effects as the duration of exposure increases. As such (and as shown in what follows), level changes at the time of the event cause no problems for interpreting the change in slope.

[44] These smoothness assumptions are stronger than the assumptions required for a regression discontinuity analysis of the immediate effect of the event. Validity of the regression discontinuity estimates requires only continuity of the expected potential outcomes (Hahn et al., 2001; Imbens and Lemieux, 2008).

Lemma 3. *Define the effect of the event at time t as*

$$\tau(t) = \mathbb{E}[Y(1) - Y(0) \,|\, T = t].$$

Under the smoothness conditions stated above,

$$\lim_{t \to 0^+} \frac{d\tau}{dt} = \lim_{t \to 0^+} \frac{d}{dt}\mathbb{E}[Y(1) \,|\, T = t] - \lim_{t \to 0^-} \frac{d}{dt}\mathbb{E}[Y(1) \,|\, T = t],$$

the change in the observed slope at $t = 0$.

Proof. The proof is brief and instructive, so I state it here. To start, differentiate $\tau(t)$ with respect to t for $t > 0$, which yields

$$\frac{d\tau}{dt} = \frac{d}{dt}\mathbb{E}[Y(1) \,|\, T = t] - \frac{d}{dt}\mathbb{E}[Y(0) \,|\, T = t].$$

To complete the proof, we take limits as $t \to 0$ from above. We obtain

$$\begin{aligned}
\lim_{t \to 0^+} \frac{d\tau}{dt} &= \lim_{t \to 0^+} \frac{d}{dt}\mathbb{E}[Y(1) \,|\, T = t] - \lim_{t \to 0^+} \frac{d}{dt}\mathbb{E}[Y(0) \,|\, T = t] \\
&= \lim_{t \to 0^+} \frac{d}{dt}\mathbb{E}[Y(1) \,|\, T = t] - \lim_{t \to 0^-} \frac{d}{dt}\mathbb{E}[Y(0) \,|\, T = t] \\
&= \lim_{t \to 0^+} \frac{d}{dt}\mathbb{E}[Y(1) \,|\, T = t] - \lim_{t \to 0^-} \frac{d}{dt}\mathbb{E}[Y(1) \,|\, T = t]
\end{aligned}$$

where the second equality uses the fact that $\mathbb{E}[Y(0) \,|\, T = t]$ is everywhere continuously differentiable, and the third equality uses the fact that $Y(0) = Y(1)$ whenever $T < 0$. □

Card et al. (2012) establish that, under reasonable conditions, both the local linear specification in equation (1.10) and the local quadratic specification in equation (1.11) deliver consistent estimates of discontinuities in slope, and the usual standard errors permit valid asymptotic inference. As they discuss, the local quadratic specification has the theoretical advantage that the asymptotic bias of the estimated change in slope will be of lower order than that obtained from a local linear specification if the regression function exhibits curvature around the point of interest.[45] This bias advantage comes at a substantial cost, however, as shifting to a locally quadratic specification increases the asymptotic variance of the estimated slope discontinuity by a factor of 16. Nevertheless,

[45] This result is directly analogous to the results presented by Hahn et al. (2001), Porter (2003), and Imbens and Lemieux (2008) that local linear regression will exhibit lower bias than "locally constant" non-parametric regression for estimating a discontinuity in level when the regression function exhibits substantial *slope* on either side of the point of interest.

because substantial curvature appears to be present in my application, I only present estimated slope discontinuities from local quadratic specifications. For precision reasons, I only present such results using the largest bandwidth used in the level discontinuity regressions (24 months for the NVSS data and 48 months for the NHDS data).

1.7.2 Results

To start, Figure 1.17 plots the basic trends in VBAC rates and overlays lines indicating the month in which each of the events under study occurred. The general visual impression is that the VBAC rate changes continuously and smoothly over time, suggesting that none of these events had a substantial effect on the VBAC rate. The only evident exception is Lydon-Rochelle et al. (2001), for which there is an apparent drop in the VBAC rate after its publication in July 2001.

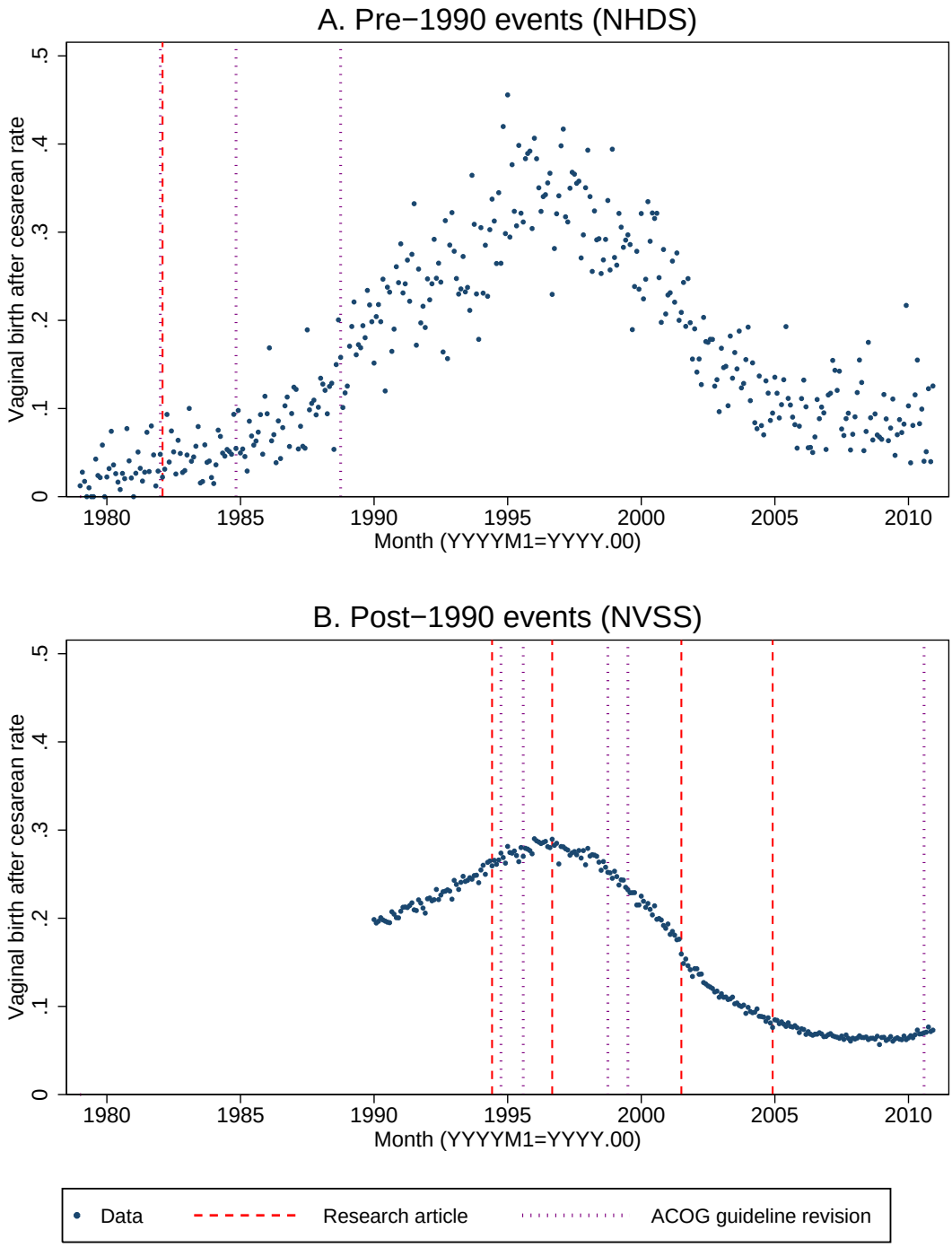
To formalize these visual impressions, Figure 1.18 reports estimates of the change in the level of the VBAC rate around each of these events obtained from estimating the local linear specification in equation (1.10). For completeness, these estimates are reported in tabular form in Table 1.12, and the fitted local linear regressions for the base bandwidth are reported in Figure 1.19. These results confirm that there is a robust and highly significant reduction in the VBAC rate of slightly less than 2 percentage points after the publication of Lydon-Rochelle et al. (2001) in July 2001. Beyond this event, however, there is essentially no evidence that any other events considered had any immediate effect on practice.[46] Indeed, at least for the events analyzed using NVSS data, I can reject effects of much more than one percentage point in either direction. These conclusions are robust to consideration of results from the local quadratic specifications reported in Figure 1.20.

I turn now to the question of whether there is any evidence that these events had a gradual effect on practice, which I ascertain by looking for a discontinuity in the slope of the time series at the time of these events. Results obtained by estimating the local quadratic specification in equation (1.11)

[46] One possible exception is the very small estimated increase in the VBAC rate around the publication of Landon et al. (2004) in December 2004, which is statistically significant under the larger and smaller bandwidths. This result should likely be taken with a grain of salt, however. Several states (including Texas) adopted the revised birth certificate in January 2005. While, as described previously, I adjust for the birth certificate change-over, unless this adjustment is perfect, a small discontinuity would remain even if there is no effect of the publication.

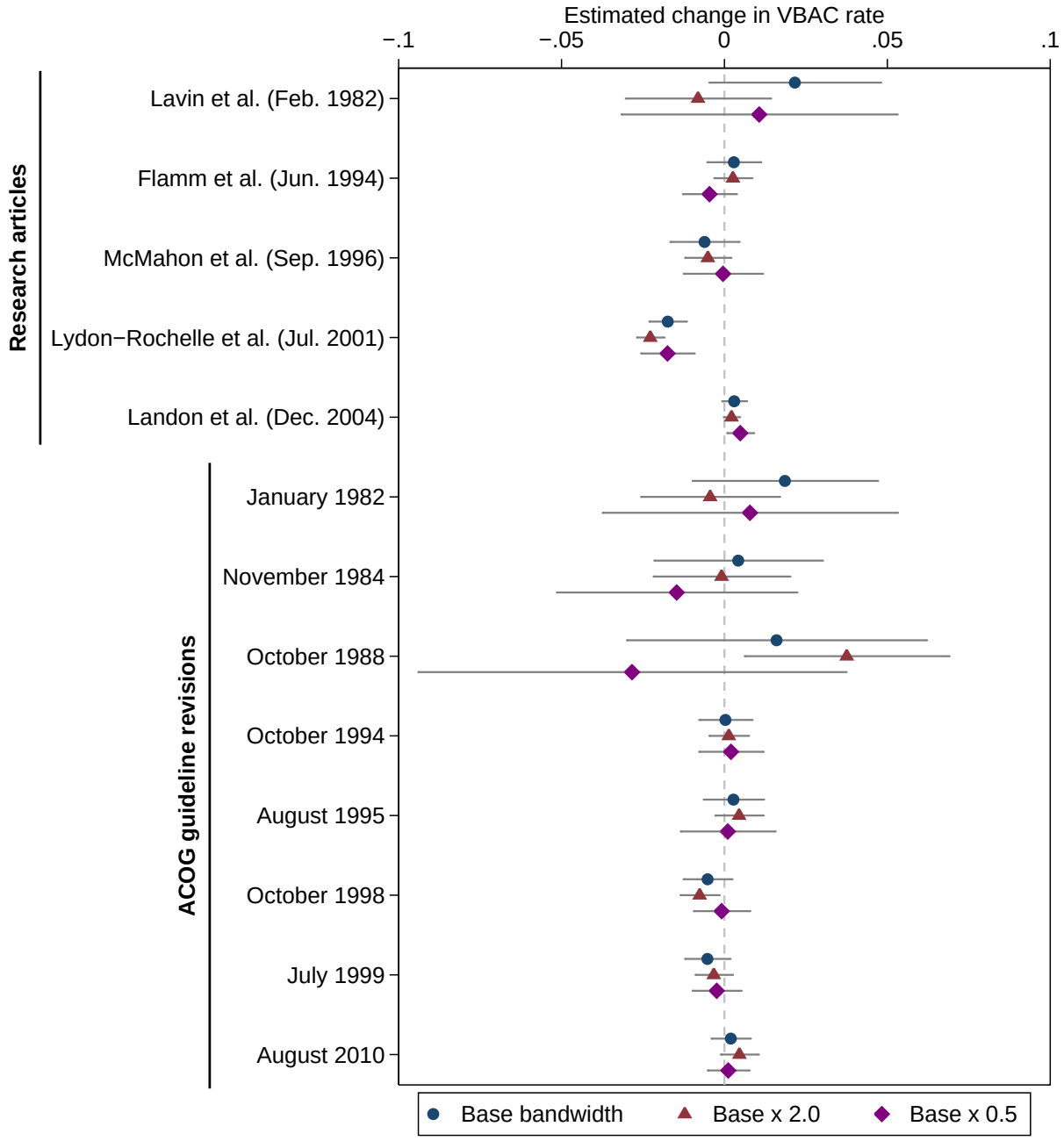
In addition, in the specifications using longer bandwidths, there are statistically significant estimated discontinuities after publication of the October 1988 and October 1998 ACOG guidelines. Both, however, occur in periods where the VBAC time series exhibits substantial curvature and, thus, RD specifications using long bandwidths are prone to giving biased results. Consistent with this theory, these discontinuities disappear in the quadratic specifications reported in Figure 1.20.

Figure 1.17: Dates of major events and monthly VBAC rates



Notes: This figure plots monthly VBAC rates from the NHDS in panel A and monthly VBAC rates from the NVSS in panel B. NVSS data for 2003 and later are adjusted for the gradual adoption of the 2003 certificate of live birth as described in Section 1.3. The red dashed lines mark months in which highly-cited research articles on the risk of uterine rupture were published. The purple dotted lines mark months in which the American College of Obstetricians and Gynecologists published new practice guidelines. Events occurring prior to January 1990 are plotted in panel A, and events occurring after that date are plotted in panel B.

Figure 1.18: Estimated discontinuities in the level of the VBAC rate around major events (linear specification)



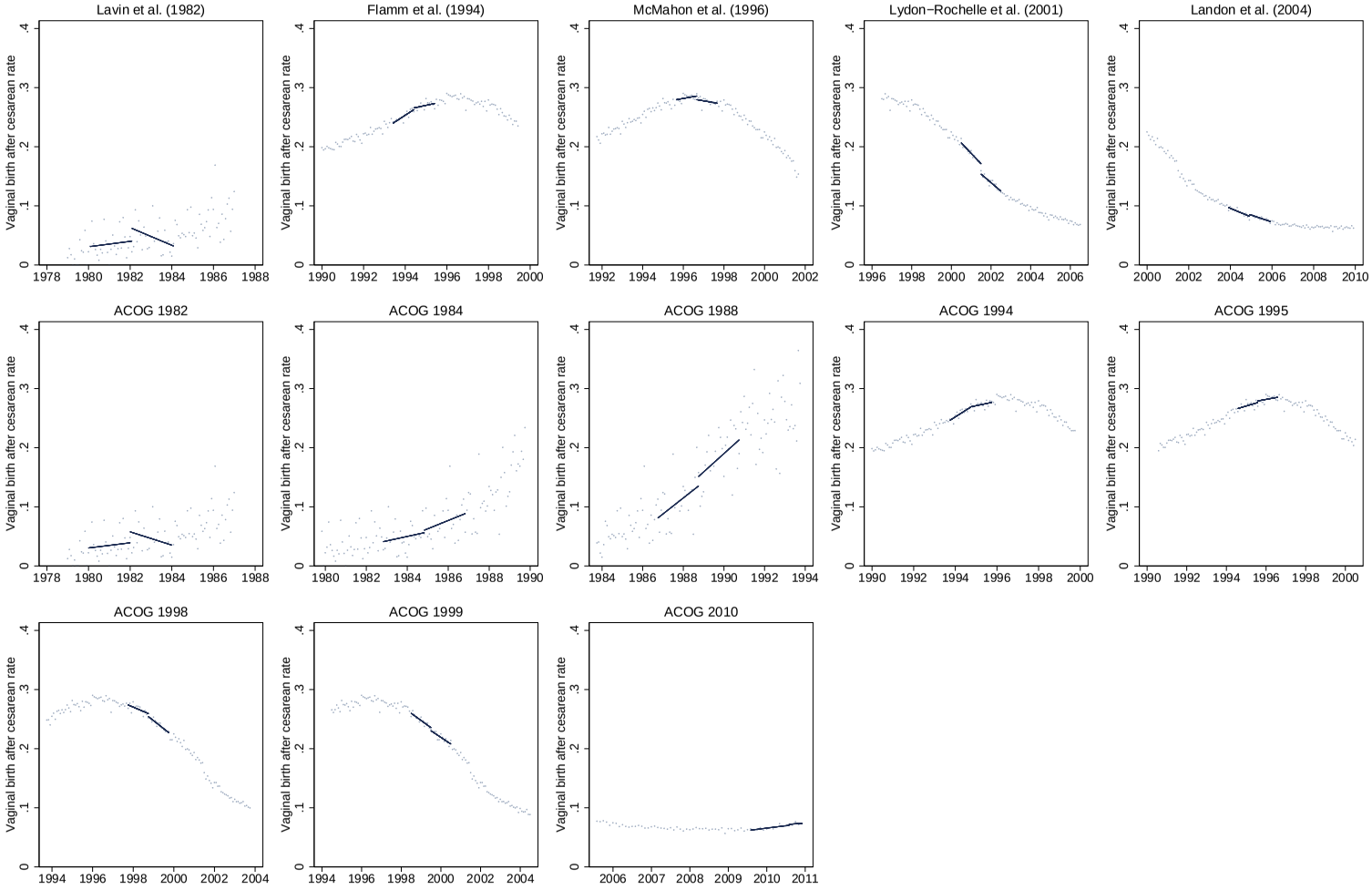
Notes: This figure plots the estimated discontinuity in the VBAC rate obtained by estimating the local linear regression specification in equation (1.10) for each of the events listed above using three different bandwidths. Gray bars depict 95 percent confidence intervals. Estimation for events occurring before January 1990 uses NHDS data and a base bandwidth of 24 months; the regressions are weighted using the NHDS sampling weights. Estimation for events occurring on or after January 1990 uses NVSS data and a base bandwidth of 12 months; regressions including births from 2003 or later are adjusted for the gradual adoption of the 2003 certificate of live birth as described in the text. In all cases, estimation excludes births occurring during the month in which the event occurs since it is unknown whether those births occurred before or after the event.

Table 1.12: Local linear and local quadratic regression estimates around high-profile events

Column:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Base bandwidth			**Base bandwidth × 2.0**				**Base bandwidth ×0.5**		
Discontinuity type:	level	level	N	level	level	slope	N	level	level	N
Polynomial order:	linear	quad		linear	quad	quad		linear	quad	
A. Estimated effects of highly-cited publications										
Lavin et al. (1982)	0.022	0.006	4,798	-0.008	0.028^{+}	-0.034^{+}	8,674	0.011	0.051	2,489
	(0.013)	(0.022)		(0.011)	(0.016)	(0.020)		(0.022)	(0.039)	
Flamm et al. (1994)	0.003	-0.011*	838,965	0.003	0.002	-0.016^{+}	1,675,920	-0.005	-0.008	419,734
	(0.004)	(0.005)		(0.003)	(0.004)	(0.009)		(0.004)	(0.009)	
McMahon et al. (1996)	-0.006	0.003	817,543	-0.005	-0.008	-0.008	1,643,727	-0.000	0.015*	407,312
	(0.005)	(0.006)		(0.004)	(0.006)	(0.012)		(0.006)	(0.008)	
Lydon-Rochelle et al. (2001)	-0.017***	-0.024***	903,700	-0.023***	-0.016***	0.001	1,809,557	-0.017***	-0.019**	449,048
	(0.003)	(0.004)		(0.002)	(0.003)	(0.006)		(0.004)	(0.007)	
Landon et al. (2004)	0.003	0.006*	1,009,760	0.002^{+}	0.005**	0.003	2,036,712	0.005*	0.005*	505,731
	(0.002)	(0.003)		(0.001)	(0.002)	(0.004)		(0.002)	(0.003)	
B. Estimated effects of ACOG guidelines										
January 1982	0.019	-0.002	4,783	-0.004	0.020	-0.017	8,544	0.008	0.022	2,475
	(0.014)	(0.023)		(0.011)	(0.017)	(0.021)		(0.023)	(0.028)	
November 1984	0.004	-0.023	5,565	-0.001	0.011	0.005	12,541	-0.015	-0.003	2,735
	(0.013)	(0.024)		(0.011)	(0.016)	(0.020)		(0.019)	(0.033)	
October 1988	0.016	-0.042	10,656	0.038*	-0.010	0.038	20,735	-0.028	-0.084*	5,520
	(0.023)	(0.036)		(0.016)	(0.025)	(0.031)		(0.033)	(0.043)	
October 1994	0.000	0.003	833,732	0.001	-0.001	-0.007	1,672,178	0.002	-0.007	414,746
	(0.004)	(0.006)		(0.003)	(0.004)	(0.009)		(0.005)	(0.007)	
August 1995	0.003	0.001	821,068	0.004	0.003	0.012	1,653,309	0.001	0.004	404,105
	(0.005)	(0.009)		(0.004)	(0.005)	(0.011)		(0.007)	(0.009)	
October 1998	-0.005	0.001	830,163	-0.008*	-0.004	-0.010	1,672,552	-0.001	-0.009	413,116
	(0.004)	(0.005)		(0.003)	(0.004)	(0.009)		(0.004)	(0.006)	
July 1999	-0.005	-0.002	843,484	-0.003	-0.002	0.022**	1,698,951	-0.002	0.001	415,762
	(0.004)	(0.004)		(0.003)	(0.004)	(0.008)		(0.004)	(0.009)	
August 2010	0.002	-0.005	747,332	0.005	-0.004	0.055	1,312,973	0.001	-0.001	465,546
	(0.003)	(0.005)		(0.003)	(0.005)	(0.061)		(0.003)	(0.005)	

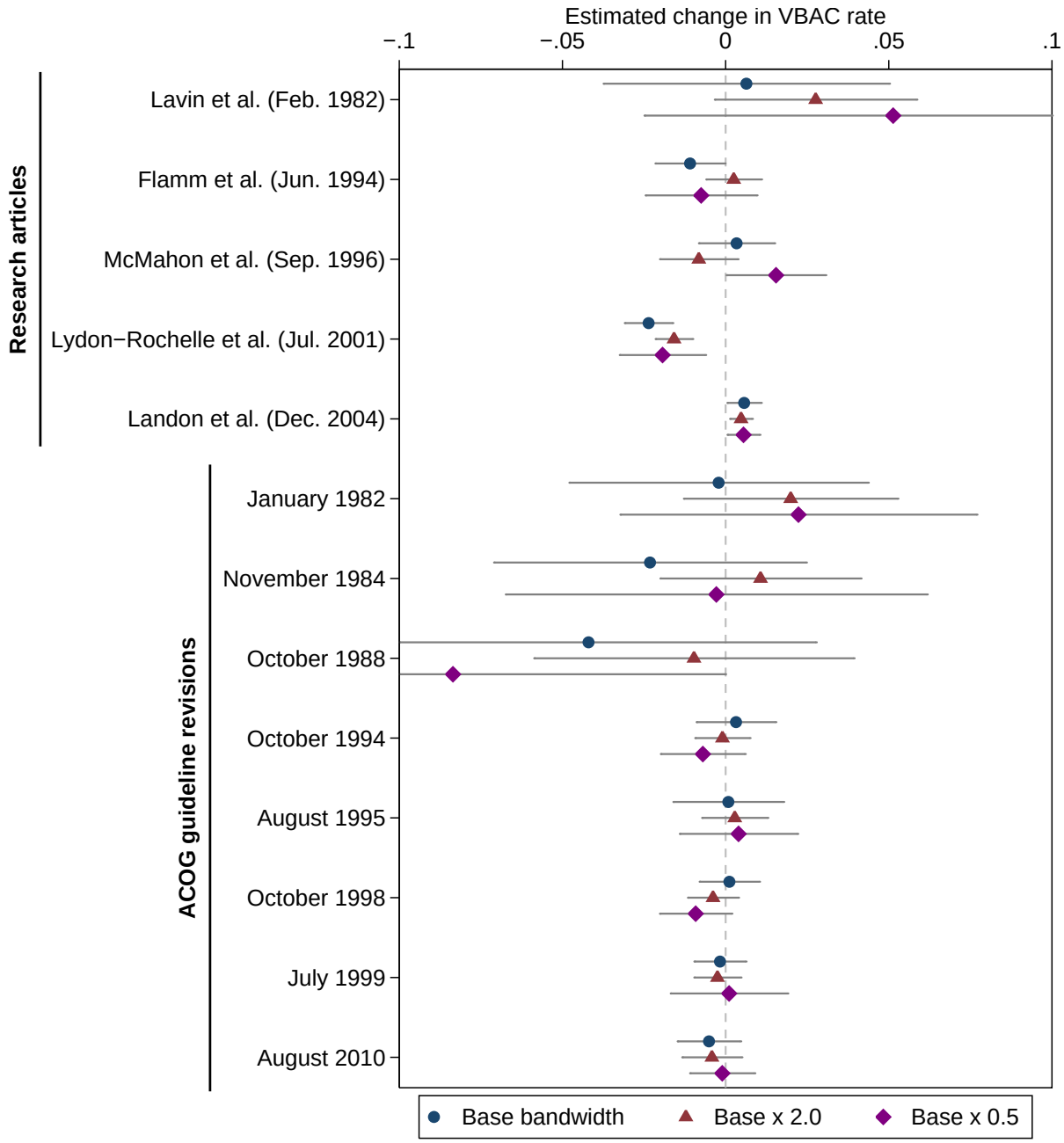
Notes: This table reports the estimated discontinuity in the level of or trend in the VBAC rate obtained by estimating the local linear specification in equation (1.10) or the local quadratic specification in equation (1.11) for each of the listed events. See the text for estimation details. Standard errors are clustered at the monthly level and displayed in parentheses. Statistical significance is denoted as follows: $^{+}$ $p < .1$, * $p < .05$, ** $p < .01$, *** $p < .001$.

Figure 1.19: Monthly VBAC rates and local linear regression fits around major events



Notes: This figure plots monthly VBAC rates for a ten year window surrounding each event date and superimposes the predicted values from estimating the local linear regression specification in equation (1.10). Estimation for events occurring before January 1990 uses NHDS data and a bandwidth of 24 months; the regressions and plotted rates are weighted using the NHDS sampling weights. Estimation for events occurring on or after January 1990 uses NVSS data and a bandwidth of 12 months; regressions and plotted rates that include births from 2003 or later are adjusted for the gradual adoption of the 2003 certificate of live birth as described in the text. In all cases, estimation excludes births occurring during the month in which the event occurs since it is unknown whether those births occurred before or after the event.

Figure 1.20: Estimated discontinuities in the level of the VBAC rate around major events (quadratic specification)



Notes: This figure plots the estimated discontinuity in the VBAC rate obtained by estimating the local quadratic regression specification in equation (1.11) for each of the events listed above using three different bandwidths. Gray bars depict 95 percent confidence intervals. Estimation for events occurring before January 1990 uses NHDS data and a base bandwidth of 24 months; the regressions are weighted using the NHDS sampling weights. Estimation for events occurring on or after January 1990 uses NVSS data and a base bandwidth of 12 months; regressions including births from 2003 or later are adjusted for the gradual adoption of the 2003 certificate of live birth as described in the text. In all cases, estimation excludes births occurring during the month in which the event occurs since it is unknown whether those births occurred before or after the event.

are reported in Figure 1.21. The corresponding point estimates are reported in tabular form in Table 1.12, and the fitted regression curves are plotted in Figure 1.22. As expected, these estimates are considerably less precise, but they provide little evidence that there are large delayed effects on practice. The only statistically significant estimate is for July 1999, and this estimate actually has the opposite of the expected sign, as the consensus is that this guideline was less favorable to VBAC than the one preceding it (MacDorman et al., 2011; Guise et al., 2010b). Examining the fitted regressions in Figure 1.22, it appears that this single anomalous result occurs because the bandwidth overlaps with the sharp drop in VBAC rates around the publication of Lydon-Rochelle et al. (2001), which distorts the fit. There is also, notably, no evidence that the effect of Lydon-Rochelle et al. (2001), gets larger over time; the estimated change in slope is essentially zero. I conclude, therefore, that there is little to no evidence that the immediate effects of these events understate their long-run effects, although the power limitations of these analyses are important to keep in mind.

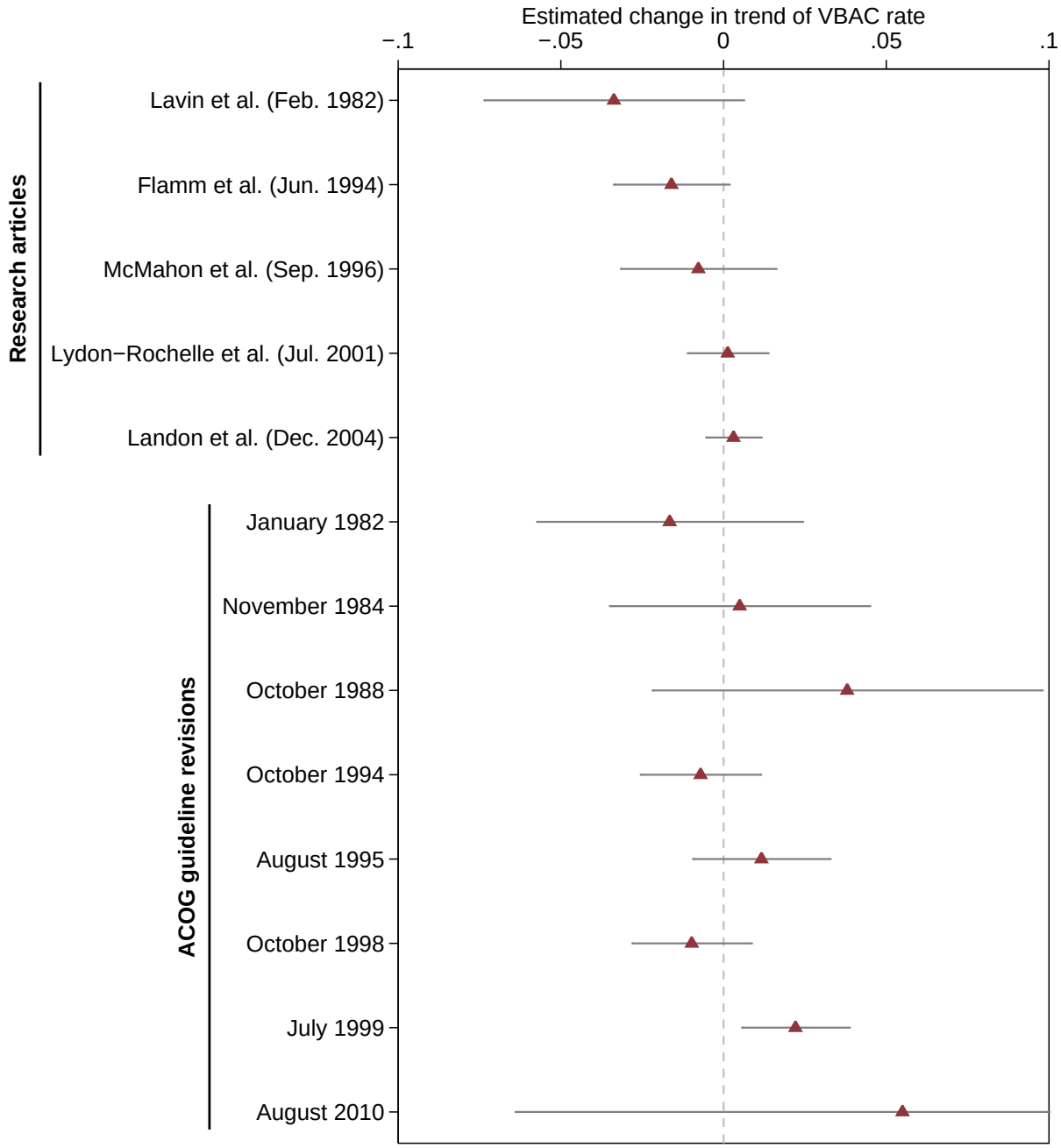
In sum, the evidence presented in this section suggests that these high-profile events did not play an important role in driving the evolution of VBAC rates over this period. Even the one exception to this general pattern, Lydon-Rochelle et al. (2001), only reduced VBAC rates by two percentage points, a very small effect when compared to the time series changed observed. In this sense, it is the exception that proves the rule. That is, it demonstrates that sufficiently important events of this kind can indeed cause immediate changes in the VBAC rate, but that even when such responses do appear, they tend to be small.

1.8 Conclusion

In this paper, I presented event study evidence that physicians and hospitals change their treatment patterns in response to idiosyncratic experiences with individual patients and that the responses to such events appear to occur at or above the level of an individual hospital. I demonstrated that this behavior is plausibly rational in light of disagreement in the clinical literature about the risk of uterine rupture.

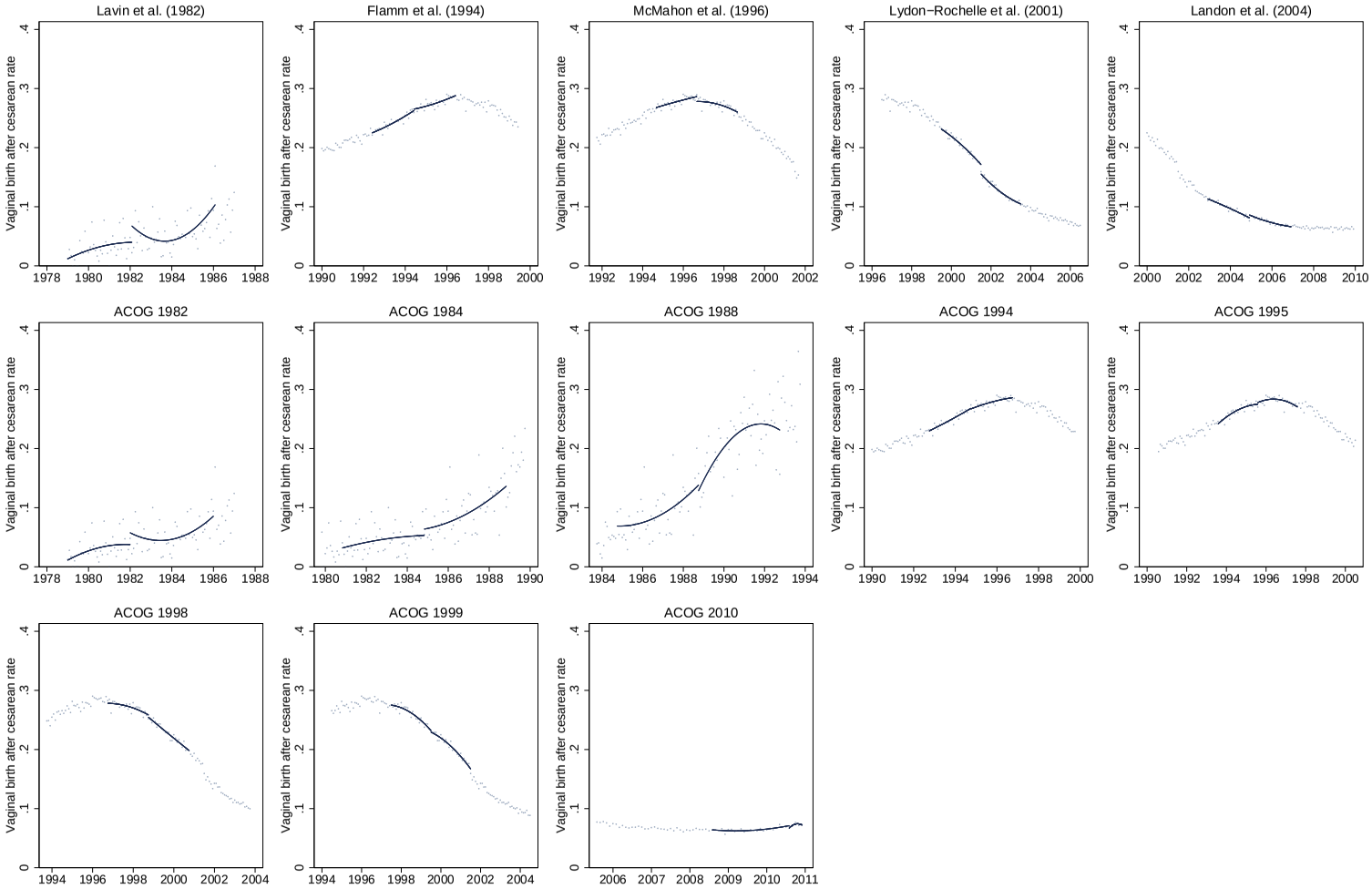
When embedded in a rational model of provider learning, the event study estimates imply the existence of substantial and long-lasting cross-provider variation in treatment decisions, even when

Figure 1.21: Estimated discontinuities in the trend of the VBAC rate around major events



Notes: This figure plots the estimated discontinuity in the trend of the VBAC rate obtained by estimating the local quadratic regression specification in equation (1.11) for each of the events listed above. Gray bars depict 95 percent confidence intervals. Estimation for events occurring before January 1990 uses NHDS data and a bandwidth of 48 months; the regressions are weighted using the NHDS sampling weights. Estimation for events occurring on or after January 1990 uses NVSS data and a bandwidth of 24 months; regressions including births from 2003 or later are adjusted for the gradual adoption of the 2003 certificate of live birth as described in the text. In all cases, estimation excludes births occurring during the month in which the event occurs since it is unknown whether those births occurred before or after the event.

Figure 1.22: Monthly VBAC rates and local quadratic regression fits around major events



Notes: This figure plots monthly VBAC rates for a ten year window surrounding each event date and superimposes the predicted values from estimating the local quadratic regression specification in equation (1.11). Estimation for events occurring before January 1990 uses NHDS data and a bandwidth of 24 months; the regressions and plotted rates are weighted using the NHDS sampling weights. Estimation for events occurring on or after January 1990 uses NVSS data and a bandwidth of 12 months; regressions and plotted rates that include births from 2003 or later are adjusted for the gradual adoption of the 2003 certificate of live birth as described in the text. In all cases, estimation excludes births occurring during the month in which the event occurs since it is unknown whether those births occurred before or after the event.

providers all face the same risk of uterine rupture. Furthermore, when the level of cross-provider dispersion in the risk of rupture is calibrated at a level that appears plausible based on the clinical literature, the model can match the actual level of cross-provider variation in treatment decisions, at least when learning is assumed to occur in a multi-hospital community.

More speculatively, I showed that if providers selectively remember uncommon bad outcomes relative to more common good outcomes, a model calibrated to match the event study estimates can explain both the rise and fall in TOLAC rates over the period studied. By contrast, the conventional account of the time series changes over this period, which focuses on a small number of high-profile events, appears inconsistent with the evidence.

One of the most striking findings of this paper is that idiosyncratic differences in experience can generate long-lasting cross-provider variation among fundamentally similar providers, even when learning is rational. This conclusion seems likely to hold beyond the particular medical setting I examined here. Indeed, there are at least two reasons to believe that idiosyncratic differences in experience may play an even more important role in other settings.

First, my setting features a large patient population; there are on the order of 500,000 deliveries by women with a prior cesarean section each year. As a result, studies on the consequences of TOLAC are plentiful and frequently feature relatively large sample sizes, so true sampling uncertainty makes a very limited contribution to physicians' prior uncertainty about the risk of rupture. In settings with a sparser clinical literature, sampling uncertainty might make a large contribution to physicians' prior uncertainty, magnifying their responsiveness to idiosyncratic events. The large patient population also means that, at least when learning occurs at the community level, experience relatively quickly reaches the point where variation in beliefs due to idiosyncratic differences in experience starts to decline. In a setting where experience accrued more slowly, variation would initially emerge more gradually, but would then be more long-lasting.

Second, the technologies of labor management and cesarean section were not changing particularly rapidly over the period I studied. In a setting with substantial ongoing technological change, the information content of clinical studies would "depreciate" over time, which would generally reduce the precision of physicians' prior beliefs and thereby magnify the effect of idiosyncratic differences in experience. Experience would similarly depreciate over time, potentially keeping physicians perpetually at the intermediate experience levels that generate the greatest amount of

variation in beliefs and practice.

This paper also contributes to our understanding of the link between cross-provider variation in the productivity of alternative treatments and cross-provider variation in utilization. Chandra and Staiger (2007) establish, in the context of cardiac catheterization, that the qualitative relationship between productivity and utilization is consistent with a model in which productivity differences drive differences in practice. Their results do not, however, address whether the observed differences in productivity are large enough to explain the observed differences in practice. By contrast, the results presented here demonstrate that plausible levels of cross-provider variation in the productivity of TOLAC (stemming from variation in the risk of rupture) are indeed large enough to explain the observed degree of cross-provider variation in TOLAC rates. This evidence that dispersion in productivity can quantitatively account for observed cross-provider variation in treatment should increase our confidence that this mechanism is important.

The results also imply that understanding why different clinical studies give different results could have great value if provider learning is indeed approximately rational. If it were determined that different clinical studies give different results primarily because different providers face truly different levels of risk (and the sources of that heterogeneity were identified), providers could jump directly to the correct practice pattern, rather than approaching it only gradually as they learn about their risk from experience. If, on the other hand, it were determined that different studies give different risk estimates primarily because of measurement error, physicians could then be much more confident that the mean estimate from the clinical literature is correct, reducing the uncertainty embodied in their common prior. This increased certainty would in turn reduce their responsiveness to idiosyncratic differences in experience which, under this scenario, would be desirable. In either case, the benefits could be substantial.

Finally, my finding that, when physicians exhibit selective memory, the gradual accumulation of experience with uterine rupture can explain the observed decline in TOLAC rates, although more speculative than the other results, is intriguing. In particular, it suggests that efforts to explain why some changes in practice diffuse quickly, others slowly, and some not at all should consider the structure of physician learning alongside more traditional economic factors like adoption costs, payment incentives, and patient demand.

Chapter 2

Technology Diffusion in Healthcare: The Welfare Implications of Use, Overuse, and Business Stealing[1]

The health sector in the United States has expanded dramatically over the last 50 years. In 1960, the United States spent 5.2 percent of GDP on health care, while by 2011, that share had more than tripled to 17.9 percent. The dominant factor driving this growth is generally believed to have been the development and diffusion of new health care technologies for treating a wide range of ailments (Newhouse 1992; CBO, 2008). As argued by Cutler (2004), many of these new technologies have generated striking improvements in the length and quality of life, benefits that have more than justified their substantial costs. On the other hand, health care markets feature a variety of well-known market imperfections that can cause technologies to spread beyond the population for which they generate positive returns (Chandra and Skinner, 2012). Overuse of sufficient magnitude can diminish or even completely dissipate the benefits generated by an otherwise promising new technology.

Diffusion of technologies that require adopting hospitals to incur large fixed costs also raises a separate set of questions traditionally studied in industrial organization. Simple economics implies that adoption by an additional hospital improves social welfare only if it leads to a sufficiently large

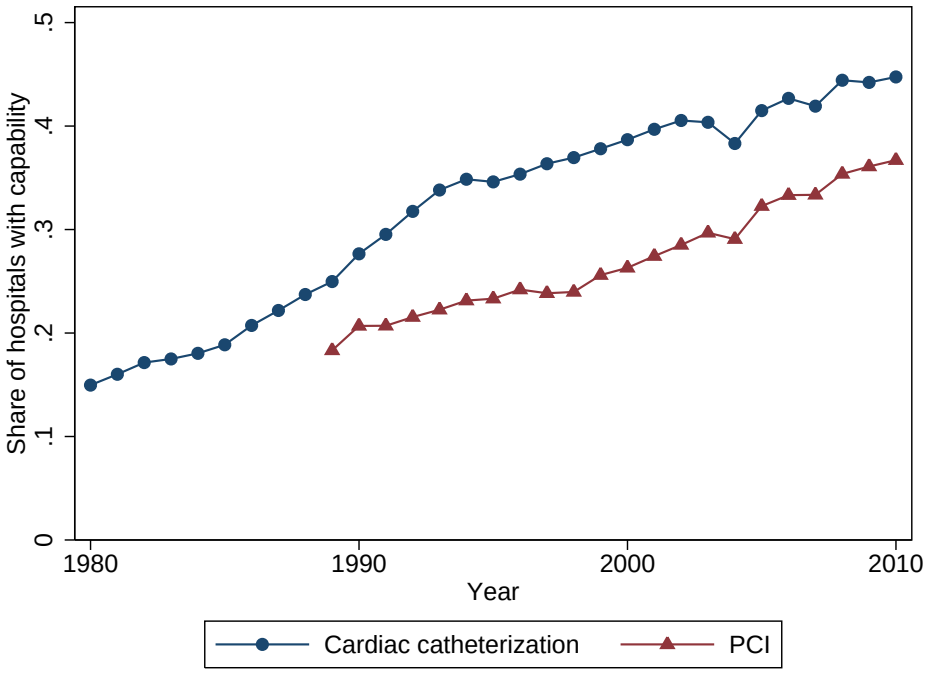
[1]This chapter is joint work with Amitabh Chandra.

market-wide increase in utilization of the technology. If adopting hospitals simply "steal" business from hospitals that have already adopted the technology, then society loses the fixed costs of adoption for no corresponding gain. Applied to the problem of technology adoption, the classic work of Mankiw and Whinston (1986) implies that excessive adoption is very likely to occur when adoption is unrestricted because new adopters can earn large profits even if they are primarily stealing business from incumbents. In practice, of course, adoption often faces important restrictions, like inelastic supply of key complementary inputs (Cutler et al., 2010) or legal restrictions, so excessive adoption is not a forgone conclusion, but it remains a real possibility.

In this paper, we examine the welfare consequences of hospital adoption of percutaneous coronary intervention (PCI), a commonly-used treatment for coronary artery disease. As depicted in Figure 2.1, the share of hospitals with PCI capabilities has risen rapidly over the last three decades and roughly one-third of hospitals have such capabilities today. This setting raises all of the questions posed above. Overuse is a real concern, as the health benefits of PCI (but not the private returns to medical providers) fall off sharply beyond the most appropriate group of patients. Likewise, PCI adoption requires a substantial capital investment, so excessive adoption has the potential to cause large welfare losses. But adoption also frequently faces legal barriers, notably state "certificate of need" requirements, so the Mankiw-Whinston conclusions do not directly apply.

We first present a simple theoretical analysis of the welfare consequences of PCI adoption that formalizes the tradeoff between the potential benefits of adoption – improved access to PCI for those who need it – and its costs – the required capital investment and expanded potential for overtreatment. Then, using Medicare claims data covering the universe of hospital stays by fee-for-service Medicare beneficiaries for 1992-2008, we implement an event study (dynamic difference-in-differences) design to estimate changes in PCI utilization around instances of hospital adoption of PCI. Unsurprisingly, we find that entering hospitals quickly acquire a substantial PCI business of 25 procedures per quarter. We find, however, that much of this business is taken from other hospitals, and the combined increase in PCI utilization among the hospital and its competitors is just 10 procedures per quarter. When we then decompose this market-wide increase in utilization according to the medical indication for the PCI, we find that only 2 of these 10 new PCIs are in patients currently experiencing an acute myocardial infarction (heart attack), the diagnosis for which the benefits of PCI are known to be large. The remainder are non-AMI patients, for whom

Figure 2.1: Diffusion of cardiac catheterization and PCI capabilities, 1980-2010



Notes: The share of hospitals with catheterization and PCI capabilities was estimated using the American Hospital Association Annual Survey. The hospital sample was restricted to those hospitals identifying themselves as general medical and surgical hospitals. The share of hospitals failing to respond to the survey in any given year rises from 10 percent of hospitals to 20 percent over the period examined. Hospitals that fail to respond in a given year are excluded from that year's estimate. Alternative approaches to handling missing data yield similar results.

evidence suggests the benefits of PCI are relatively small.

In the final section of the paper, we use our theoretical framework to integrate these utilization estimates with pre-existing evidence on the costs and benefits of PCI. Summarizing existing evidence on the marginal costs and benefits of PCI, we calculate that each additional PCI in an AMI patient generates social benefits of nearly $36,000, while each PCI performed in a non-AMI patient generates a social loss of close to $4,400. As a result, although our estimate of the cumulative per adoption benefit of increased utilization in AMI patients ($5.2 million) exceeds our estimate of the fixed cost of adoption ($3.0 million), the substantial welfare losses attributable to non-AMI patients ($2.6 million) cause adoption over the period studied to be close to welfare neutral overall.

We also demonstrate that this social calculus differs markedly from the private calculus of an individual hospital. Hospitals earn profits of $1.9 million on AMI patients, less than the fixed costs of adoption. By contrast, they earn very large profits of $7.7 million on non-AMI patients, which transform PCI adoption from a money-loser into a hugely profitable undertaking for a hospital. This pattern of private benefits implies that policy changes that solely discourage utilization of PCI among non-AMI patients cannot achieve the socially efficient outcome in which hospitals adopt PCI but only serve AMI patients. Optimal policy pairs changes for non-AMI patients with increases in PCI reimbursement for AMI patients that allow hospitals to capture more of the social surplus generated by PCI in these patients.

Our paper adds to a small economics literature on the diffusion of PCI in particular and the welfare consequences of entry in health care markets generally. Cutler and Huckman (2003) examine the diffusion of PCI in New York from the early 1980s through 2000. Cutler et al. (2010) examine an expansion in the number of cardiac surgery programs in Pennsylvania following relaxation of the state's "certificate of need" requirements in 1996. Unlike these earlier papers, we are able to examine the consequences of adoption nationwide. Not only is this national scope of intrinsic interest, but it also enables us to implement the event study design used in this paper, which is likely to be less vulnerable to confounds than earlier work. We also examine the divergence between the current structure of incentives in this market and the structure of incentives that would maximize social welfare.

The paper proceeds as follows. Section 1 provides a brief description of our clinical setting, and Section 2 presents a simple theoretical framework for thinking about the welfare consequences

of adoption in this setting. Section 3 describes our data, and Section 4 describes our empirical strategy. Section 5 presents our results. Section 6 presents a welfare analysis of the consequences of adoption, and the final section concludes.

2.1 Clinical background

The clinical context of this paper is the treatment of coronary artery disease, which is the leading cause of death in the United States. In coronary artery disease, fatty deposits known as "plaques" accumulate on the walls of the coronary arteries, the arteries that supply blood to the heart muscle. Plaques can interfere with blood flow, depriving the downstream heart muscle of oxygen and causing angina (chest pain). Plaques can also rupture, precipitating a clotting response that severely restricts blood flow through the artery. This reduction in blood flow can cause "unstable" (sudden and serious) angina and, potentially, death of the affected heart muscle. Events of this type are collectively referred to as acute coronary syndromes (ACSs), and cases in which the loss of blood flow leads to the death of the affected heart muscle are known as an acute myocardial infarctions (AMIs), or, colloquially, heart attacks.

The main focus of this paper is a category of procedures used in the treatment of coronary artery disease known as percutaneous coronary intervention (PCI). PCI falls under the broader heading of cardiac catheterization. In a cardiac catheterization, the operating physician inserts a thin tube known as a catheter through an incision in the patient's groin or wrist and guides the catheter through the patient's arteries into the coronary arteries (or other heart structure of interest). Once in place, the catheter is used to inject a contrast dye that is opaque to x-ray radiation, which permits the physician to use a specialized x-ray camera to visualize the location and extent of any coronary blockages. In many cases, physicians then use specialized tools attached to the end of the catheter to clear any blockages found; interventional procedures of this kind are collectively referred to as PCI. In the most common such procedure, which is known as angioplasty, the physician inflates a small balloon inside the affected artery to compress the accumulated plaque. A small wire mesh tube known as a stent may also be inserted to keep the artery from closing back up over time.

The benefits of PCI relative to purely medical therapy differ according to the manifestation of coronary artery disease being treated. There is clear evidence that, for patients with acute coronary

syndromes, PCI soon after onset of symptoms both improves survival and improves quality of life by reducing the subsequent incidence of AMI and angina. By contrast, for patients with non-acute manifestations of coronary artery disease – namely, stable angina – there is no evidence that PCI improves survival, although it may reduce the subsequent incidence of angina.[2]

Hospitals that wish to conduct cardiac catheterization procedures (including PCI) require specialized facilities known as cardiac catheterization laboratories. Building a catheterization laboratory requires a capital outlay of on the order of $3 million in today's dollars (Lieu et al., 1996; Gehrki, 2004). In addition, as of 2011, construction of catheterization laboratories was subject to state "certificate of need" requirements in 26 states (NCSL, 2012). Thus, in large portions of the country, construction also requires demonstrating to state regulators that such a facility is "necessary."

2.2 Theoretical framework

To frame the empirical exercise and discussion, we present a simple framework for analyzing the welfare consequences of hospital PCI adoption. Consider a market with multiple types of potential patients $\theta \in \Theta$, where the mass of each type of patient is given by some measure $\mu(\cdot)$. While the theoretical development that follows is fully general, in our application Θ will consist of just two types of patients: AMI patients and non-AMI patients. The medical benefit of PCI, monetized in some appropriate fashion, varies with θ and is given by $\tau(\theta)$. Hospitals can incur setup costs k at $t = 0$ to equip a PCI-capable catheterization laboratory, which will remain in operation for the next $\bar{T}$ units of time (interpreted as the lifetime of the equipment). Having paid the setup cost, PCI can be provided at marginal cost $c(\theta)$ for patients of type θ.[3] The probability that patients receive PCI may vary with the number of hospitals present in the market; let $v_N(\theta)$ denote the (flow) rate at which patients of type θ receive PCI when N hospitals have entered the market. We

[2]We provide a more detailed discussion of the clinical evidence in this area, including a discussion of the magnitude of the clinical benefits of PCI, in Section 2.6.

[3]In assuming that the marginal costs and benefits are not a function of the number of entrants, we have assumed that there are no economies of scale either in treatment quality or cost. There is, however, some evidence that such economies of scale do exist, at least with respect to treatment quality (Hannan et al., 2005). In general, if economies of scale are present, calculations using this framework will overstate the benefits of entry. Determining how to account for these factors in our empirical exercise is a possible avenue for further work.

leave the form of $v_N(\theta)$ fully general rather than deriving it from a particular model of consumer demand and market structure.

In this setup, total social welfare when N hospitals elect to enter the PCI market may be written as

$$W(N) = \int_0^{\bar{T}} \left[\int_\Theta (\tau(\theta) - c(\theta)) v_N(\theta) d\mu(\theta) \right] e^{-rt} dt - kN,$$

where r is an appropriate social discount rate. The change in social welfare created by an additional hospital entering the market when N hospitals have already elected to enter is then given by

$$\Delta W(N, N+1) = \underbrace{\left[\int_\Theta (\tau(\theta) - c(\theta))(v_{N+1}(\theta) - v_N(\theta)) d\mu(\theta) \right]}_{\text{flow net benefits of PCI}} \underbrace{\left[\frac{1 - e^{-r\bar{T}}}{r} \right]}_{\text{discount factor}} - \underbrace{k}_{\text{fixed cost}}, \tag{2.1}$$

where we have integrated over t. The integral gives the (flow) net benefits associated with PCI provision, while the term multiplying it is the appropriate discount and scaling factor for a discount rate of r and a period length of $\bar{T}$. The social desirability of additional adoption therefore hinges on whether the the discounted net benefits of the additional utilization (the first term) exceed the setup costs required for entry (the second term).

Examining the flow net benefit term in equation (2.1) more closely, we see that the benefits of additional utilization depend on two factors. The first is the magnitude of the market-wide increase in utilization of PCI, which is determined by the change in utilization, $v_{N+1}(\theta) - v_N(\theta)$, for each type θ and the population size $\mu(\Theta)$. In order to justify the setup costs and increase welfare, adoption must sufficiently increase utilization, rather then simply redirecting volume from incumbent hospitals to the entrant. In their well-known theoretical work, Mankiw and Whinston (1986) show that the latter is a substantial concern. They show that in homogeneous product markets with free entry, these "business stealing" effects mean that entry will almost always exceed the social optimum. In the present setting, of course, entry into the PCI market is frequently subject to legal restrictions (as noted above) or relies upon inputs in inelastic supply (Cutler et al., 2010). For this reason, whether hospital entry into the PCI market raises or lowers welfare remains an empirical question.

The second factor is the (marginal) cost-effectiveness of the incremental care: $\tau(\theta) - c(\theta)$. In a standard market, we would expect the net benefits of the additional utilization to be positive for the types $\theta \in \Theta$ that receive care (i.e. those for which $v_N(\theta) > 0$). If these net benefits

are indeed positive, adoption that increases utilization by a sufficiently large amount will always be welfare-improving. In health care markets, however, the presence of insured consumers and imperfect provider payment systems may cause overtreatment: utilization that extends beyond those patients for whom the benefits exceed the (marginal) costs. This means that additional utilization resulting from entry can directly reduce welfare, turning the standard logic on its head.

We seek to use this framework to evaluate the welfare effects of post-1992 PCI adoption. The following sections present empirical evidence on whether hospital PCI adoption increases access to PCI and, if so, by how much and for whom. In Section 2.6, we combine these estimates with estimates of $\tau(\theta) - c(\theta)$ and k drawn from the literature in order to estimate the full net benefits of hospital entry into the PCI market over the period studied.[4]

2.3 Data

Our main data source is the Medicare Provider Analysis and Review (MedPAR) files for 1992-2008, which provide data on the universe of hospital and skilled nursing facility discharges of fee-for-service Medicare beneficiaries.[5] The file reports a rich set of information on each stay, including the Center for Medicare and Medicaid Services (CMS) provider number for the facility at which the stay occurs, the dates of admission and discharge, basic patient characteristics (including zipcode of residence), and ICD-9-CM codes for the diagnoses and procedures associated with the stay. The specific ICD-9-CM diagnosis and procedure codes used to identify the diagnosis and procedure categories of interest are reported in Appendix B.1.

We use these data to construct a hospital-level panel of admission volumes over the period 1992Q1 to 2008Q2.[6] The universe of hospitals covered by the panel includes all acute care short-

[4]Observe that we do not require information on how entry affects prices to assess the welfare consequences of entry. Any welfare effect of reduced prices is fully captured in the observed effects on utilization.

[5]The MedPAR file also reports discharge records for individuals enrolled in certain classes of Medicare Advantage plans. However, the precise categories of plans included vary over time and these records are thought to be less complete (Asper and Mann, 2011). For consistency, therefore we limit our analysis to discharges of patients covered by traditional Medicare. More generally, although the lack of coverage of Medicare Advantage claims is not ideal, it is likely not important for the present analysis.

[6]The panel does not extend through the end of 2008 because the MedPAR files report stays on the basis of date of discharge, rather than date of admission. As a result, some admissions that occur late in 2008 will be reported in the 2009 MedPAR file. Ending the panel in 2008Q2 ensures that we achieve virtually complete coverage for the included quarters.

term hospitals that provide general medical and surgical services, which excludes skilled nursing facilities, long-term care hospitals, and various specialty hospitals (e.g. psychiatric hospitals, cancer hospitals, and specialty surgical hospitals). We identify these facilities using information from the CMS Provider of Services file and information on hospital case mix from the MedPAR file itself. The precise criteria are described in detail in Appendix B.1. A single facility may, for a variety of reasons, report admissions under more than one CMS provider number during the period studied. We group such provider numbers into a single "consolidated" record for the purposes of analysis. The algorithms for identifying and consolidating related provider numbers are also described in Appendix B.1

The event study design we use in this paper requires knowing the date each entering hospital adopts PCI. We consider a hospital to have adopted PCI in the first quarter in which the hospital reports: (a) at least one PCI; and (b) a total of at least five such procedures during the current quarter and the following three quarters. The second criterion helps distinguish isolated procedure coding errors from true PCI adoption. Because this criterion uses a full year of data to confirm adoption, adoption status is observed only through 2007Q3.

The theoretical analysis provided in the previous section shows that we will wish to estimate the effect of PCI adoption not only the adopting hospital's own volume, but also "market-wide." We compute market-wide volume in practice by adding each hospital's volume to the volume of an identified set of competitors. To identify the set of hospitals with which a given hospital competes, we turn to a simple model of patient hospital choice. Formally, we suppose that the utility of admission to hospital h for patient i living in zipcode z is given by

$$u_{izh} = \xi_h + \gamma_{hz} + \epsilon_{ihz}.$$

The coefficients $\{\xi_h\}$ represent the quality of each hospital h, while the coefficients $\{\gamma_{hz}\}$ represent the idiosyncratic desirability of each hospital in each zipcode (which is based, presumably, on the zipcode's proximity to the hospital, among other factors). Assuming that patients maximize utility and that ϵ_{ihz} follows an extreme value type I distribution, the market shares implied by this demand system take the standard form:

$$s_{zh} = \frac{\exp[\xi_h + \gamma_h z]}{1 + \sum_{h'=1}^{H} \exp[\xi_{h'} + \gamma_{h'z}]}.$$

To identify the competitors of hospital h, we consider the reallocation of patients arising from a small increase in the desirability of hospital h (that is, a small increase in ξ_h). It is straightforward to show that

$$\frac{\partial s_{zh}}{\partial \xi_h} = s_{zh}(1 - s_{zh}),$$

and that for $h' \neq h$,

$$\frac{\partial s_{zh'}}{\partial \xi_h} = -s_{zh}s_{zh'}.$$

Summing over zipcodes z, the marginal (normalized) increase in the volume of hospital h in this thought experiment is $\sum_z w_z s_{zh}(1 - s_{zh})$, where w_z is the share of the population living in zipcode z. The portion of this increase that represents business "stolen" from a given hospital $h' \neq h$ is similarly calculated to be $\sum_z w_z s_{zh}s_{zh'}$. For each hospital h, we then construct the hospital's competitor set C_h as the smallest set of hospitals that accounts for 95 percent of the total stolen business. That is, the set C_h is the smallest set that satisfies:

$$\sum_{h' \in C_h} \sum_z w_z s_{zh}s_{zh'} \geq 0.95 \sum_z w_z s_{zh}(1 - s_{zh}).$$

All of the zipcode population sizes (w_z) and zipcode market shares (s_{zh}) can be estimated directly from the data. We do so in the obvious way using data on all hospital admissions over the full sample period that report a valid zipcode, except that, for computing market shares in zipcode z, we exclude admissions to hospitals with a market share of less than 1 percent. "Trimming" the zipcode market shares in this way has two advantages. First, it dramatically reduces the number of hospitals with positive market share in each zipcode, and therefore dramatically reduces the computational burden of calculating the sets C_h. Second, cases in which a patient goes to a hospital that has a very low market share in his zipcode predominantly arise when the patient is traveling or when the patient's zipcode is recorded incorrectly. Including such admissions would, therefore, generate a misleadingly broad view of a hospital's set of competitors, which would introduce considerable noise into the market-wide volume estimates and reduce the precision of the regression results. In any case, we demonstrate that our qualitative results are insensitive to varying the 1 percent threshold.

An alternative way of defining each hospital's "market" would be to include all zipcodes in which the hospital has a market share above some threshold. While this approach has the advantage of

being very simple, it also has an important disadvantage. As noted above, a non-trivial share of admission records report an invalid zipcode or a zipcode that does not correspond to the patient's location at the time the hospital admission became necessary. As a result, under this approach, some admissions will be allocated to the wrong geographic market or to no market at all. In contrast, there are no missing provider numbers, and, by definition, the patient must have been physically present at the hospital when admitted. The patient's provider number will thus reliably map every hospital admission to the correct geographic "market." While the zipcode errors mean that these markets may be defined with some error, it seems likely that the trimming approach described above will ensure that those errors are small.

2.4 Estimation

We estimate the effect of adoption on subsequent hospital volume using a dynamic difference-in-differences (Jacobson et al., 1993) specification, which has frequently been referred to as an "event study" design in recent work (e.g. Almond et al. (2011) and Hilger (2012)). Our main estimating equation takes the form:

$$Y_{hq} = \psi_h + \phi_q + 1\{q < Q_h - \underline{d}\}\tau_{-\underline{d}-1} + \sum_{\substack{d \in \{-\underline{d},\ldots,0,\ldots,\bar{d}\} \\ d \neq -1}} 1\{q - Q_h = d\}\tau_d + \epsilon_{hq}, \tag{2.2}$$

where h indexes hospitals and q indexes quarters, Y_{hq} is the volume measure of interest, Q_h is the quarter in which hospital h adopts PCI, $\{\psi_h\}$ is a full set of hospital fixed effects, $\{\phi_q\}$ is a full set of quarter fixed effects, ϵ_{hq} is the error term, $\underline{d}$ is the number of leads of adoption included, and $\bar{d}$ is the number of lags.

Provided that the standard common trends assumption holds, the coefficients τ_d for $d \geq 0$ can be interpreted as the effect of PCI adoption on volume d quarters after adoption. The common trends assumption also implies that $\tau_d = 0$ for all $d < 0$. By testing this restriction, we can evaluate the plausibility of this crucial assumption.

In practice, we set the number of leads $\underline{d}$ to 8. The number of lags $\bar{d}$ is set to the maximum number of post-adoption periods observable in the sampled population of hospitals. In presenting the results, however, we focus on the first 12 post-adoption quarters. As one ventures further

into the post-adoption period, the precision of the estimates shrinks (because fewer hospitals are observed at long post-adoption horizons) and any small violations of the common trends assumption that do exist are magnified.

It will frequently be desirable, for reasons of both precision and simplicity, to combine estimates at multiple post-adoption horizons into a single summary measure. In practice, we focus on the following "pooled" estimator:

$$\hat{\tau}_{\text{pooled}} = \frac{1}{9} \sum_{d=4}^{12} \hat{\tau}_d. \tag{2.3}$$

The formula excludes $d \in \{0, 1, 2, 3\}$ so as to exclude the immediate post-adoption "ramp up" and permit interpretation of $\hat{\tau}_{\text{pooled}}$ as the "long-run" effect of PCI adoption on volumes.

The characteristics of the available data introduce one final complication. Because we identify PCI adoption directly from data on hospital volumes, we cannot identify the time of adoption for a hospital adopting PCI on or before 1992Q1, nor do we have all the information necessary to code the "lead" regressors for hospital-quarters falling less than $9 (= \underline{d} + 1)$ quarters before 2007Q3. To address these problems, we exclude hospitals that adopted PCI on or before 1992Q1 and exclude hospital-quarters falling after 2005Q2 when estimating equation (2.2).

2.5 Results

Before proceeding to the main results, we present Table 2.1, which reports descriptive statistics for the hospital panel. For consistency with the regression sample, Table 2.1 covers the period 1992Q1 to 2005Q2. In order to accurately portray the full population of hospitals, however, Table 2.1 does include hospitals that adopt PCI on or before 1992Q1 despite the fact that these hospitals are excluded from the regression sample.

We start by examining the effect of PCI adoption on the adopting hospital. Figure 2.2 plots the coefficients obtained from estimating equation (2.2) with the hospital's volume of PCI admissions as the left-hand-side variable. PCI volume is flat in the pre-period (which is unsurprising since volume is essentially zero at hospitals that have not yet adopted) and increases sharply in the quarters following adoption to around 25 procedures per quarter. Table 2.2 reports the corresponding summary estimate, which is defined in equation (2.3), for this set of coefficients.

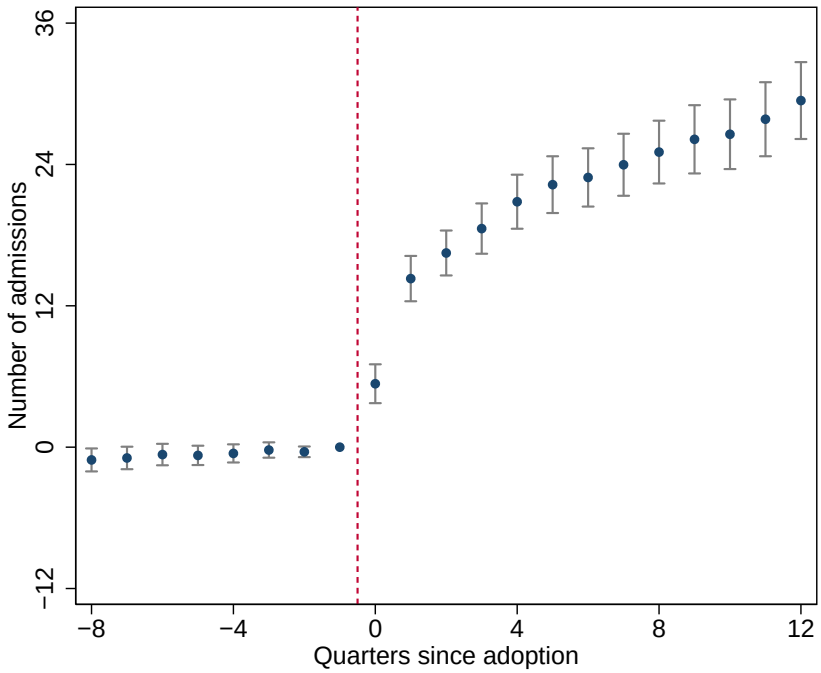
We next investigate whether the increase in PCI volume at the adopting hospital represents an

Table 2.1: Hospital panel descriptive statistics

Hospital-quarters included:	All		Post-adoption	
Volume measures	Mean	SD	Mean	SD
Hospital-level quarterly volumes				
Total admissions	598	690	1,409	861
Cardiovascular admissions	196	255	498	339
PCI admissions	15	42	63	67
with AMI primary diagnosis	4	11	17	17
with non-AMI primary diagnosis	11	32	46	53
Market-wide quarterly volumes				
Total admissions	11,517	7,490		
Cardiovascular admissions	3,916	2,517		
PCI admissions	409	285		
with AMI primary diagnosis	106	73		
with non-AMI primary diagnosis	303	219		
Sample sizes				
Hospital-quarters	253,848		58,769	
Hospitals	5,094			
PCI adoption events	613			
Pre-1992Q1 PCI adoptions	848			

Notes: This table reports descriptive statistics for the sample used to estimate equation (2.2) plus the sample of hospitals adopting PCI on or before 1992Q1. The first two columns report means and standard deviations for the full sample, while the latter two columns report data for hospital-quarters in which the hospital has PCI capabilities. Cardiovascular admissions are those with a cardiovascular primary diagnosis. PCI admissions are those that report a percutaneous coronary intervention in any procedure field. The ICD-9-CM codes used to define these categories are described in Appendix B.1. Hospital-level quarterly volumes reflect volumes at a given hospital, while market-level quarterly volumes reflect total volumes at the given hospital and its "competitors" as defined in Section 2.3.

Figure 2.2: Dynamic difference-in-differences estimates of the effect of PCI adoption on the adopting hospital's PCI volume



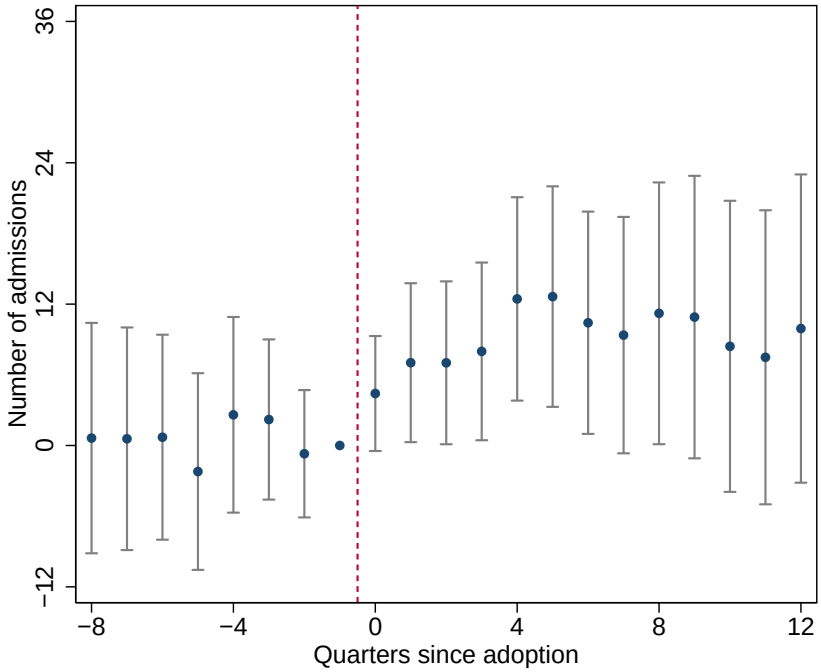
Notes: This figure plots estimates of the coefficients $\{\tau_d\}_{d\in\{-8,-7,\ldots,11,12\}}$ obtained from estimating equation (2.2); as in equation (2.2), the coefficient for $d = -1$ is normalized to zero. The dependent variable is the total number of PCIs performed in each hospital-quarter. The error bars depict 95 percent confidence intervals for the plotted point estimate and are computed using standard errors clustered at the hospital level.

Table 2.2: Summary difference-in-differences estimates of the effect of PCI adoption on hospital-level and market-wide PCI volumes

Column:	(1)	(2)	(3)
	Primary diagnosis		
Analysis level	All	AMI	Non-AMI
Hospital-level	25.0***	7.6***	17.4***
	(1.3)	(0.3)	(1.0)
	[0.10]	[0.05]	[0.13]
Market-wide	10.3*	1.8	8.6*
	(4.9)	(1.2)	(4.0)
	[0.45]	[0.34]	[0.62]
Auxiliary statistics			
Business-stealing ratio	0.587	0.767	0.509
Hospital-quarter observations	209,015	209,015	209,015

Notes: This table reports summary results from estimating equation (2.2). For each analysis level and primary diagnosis level, the reported point estimate is the "pooled" long-run volume effect defined in (2.3). For the hospital-level analyses, the dependent variable is the total number of PCIs performed on patients with the listed primary diagnosis in each hospital-quarter. For the market-wide analyses, the dependent variable is the total number of PCIs performed on patients with the listed primary diagnosis at the hospital or its competitors in each hospital-quarter (where the set of competitors is defined as described in Section 2.3). AMI is an abbreviation for acute myocardial infarction (heart attack). The p-value from a test of the null hypothesis of pre-period common trends (i.e. that $\tau_2 = \tau_3 = \cdots = \tau_8 = 0$) is displayed in brackets. Standard errors clustered at the hospital level are reported in parentheses. Statistical significance is denoted as follows: * $p < .05$, ** $p < .01$, *** $p < .001$.

Figure 2.3: Dynamic difference-in-differences estimates of the effect of PCI adoption on market-wide PCI volume



Notes: This figure plots estimates of the coefficients $\{\tau_d\}_{d\in\{-8,-7,\ldots,11,12\}}$ obtained from estimating equation (2.2); as in equation (2.2), the coefficient for $d = -1$ is normalized to zero. The dependent variable is the total number of PCIs performed at the hospital or its competitors in each hospital-quarter (where the set of competitors is defined as described in Section 2.3). The error bars depict 95 percent confidence intervals for the plotted point estimate and are computed using standard errors clustered at the hospital level.

increase in the total number of patients undergoing PCI in the hospital's market, or whether it instead reflects "business stealing" from the hospital's competitors. To address this question, we re-estimate equation (2.2), except that we place the combined PCI volume of the hospital and its competitors on the left-hand side of the regression.

Figure 2.3 reports the results, plotted on the same scale as Figure 2.2 to facilitate comparison. The figure shows that, while there is indeed an increase in market-wide PCI volume following adoption, this increase is considerably smaller than the increase in volume experienced by the adopting hospital. Comparing the two summary estimates reported in Table 2.2, we conclude that approximately 60 percent of the PCI volume at the adopting hospital is taken from the hospital's competitors, and the market-wide increase in quarterly PCI volume is only about 10 procedures.

As the theoretical framework provided in Section 2.2 makes clear, evaluating the welfare im-

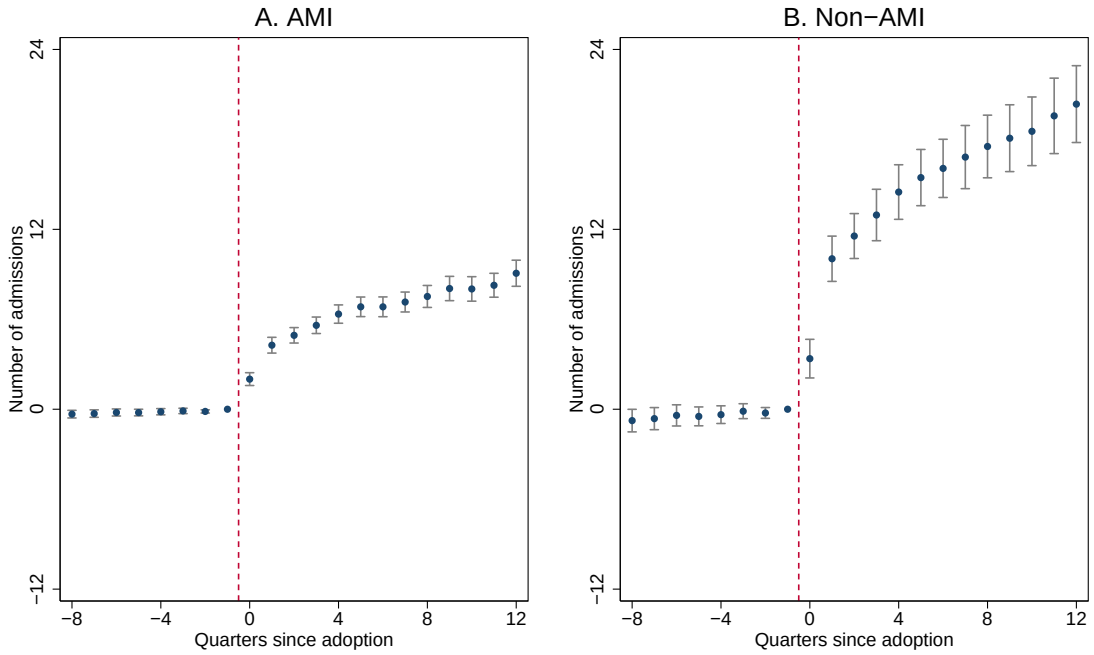
plications of this expansion in PCI availability requires understanding which categories of patients are newly receiving PCI. As described previously, the available clinical evidence indicates that PCI can have substantial benefits for patients with acute coronary syndromes (i.e. AMI or unstable angina), while the benefits to patients with chronic manifestations of coronary artery disease are considerably smaller. Guided by this medical evidence, therefore, we seek to separately estimate the increase in PCI volume for patients with and without acute illness. We operationalize this in practice by dividing patients receiving PCI into two groups: those with a primary diagnosis of AMI and those with other primary diagnoses. This distinction differs from the ideal clinical categorization, as it includes the relatively small number of patients with unstable angina in the "other" category. Unfortunately, this imprecision is unavoidable in light of prevailing ICD-9-CM coding practices, which make it impossible to distinguish the relatively small group of patients with unstable angina from a much larger group of patients with chronic coronary artery disease.[7]

Figure 2.4 provides this diagnosis breakdown of the increase in PCI volume at the adopting hospital alone. For the adopting hospital, the increase in PCI volume for patients with AMI diagnoses is large and highly statistically significant, although it is less than half as large as the increase in PCI volume for patients with non-AMI diagnoses. Figure 2.5 reports the same diagnosis breakdown at the market level. At the market level, the estimated increase in PCI volume for patients with AMI diagnoses shrinks dramatically. While the summary estimate for AMI patients that is reported in Table 2.2 remains positive, it is not statistically significant. By contrast, even at the market-level, there is a large and statistically significant increase in PCI volume for patients without an AMI diagnosis. We conclude, therefore, that hospital PCI adoption most likely causes a small increase increase in market-wide PCI utilization for AMI patients, but causes a substantial increase in utilization for other patients.

Having established our main results, we turn to a pair of more subtle issues that may affect the interpretation of those results. The first issue we explore is whether entry by the adopting hospital deters entry by other hospitals. If this were the case, then the event study results do not answer the question of interest: how adding one additional PCI-capable hospital to the market affects utilization in equilibrium. To look for deterrent effects of entry, we estimate equation (2.2)

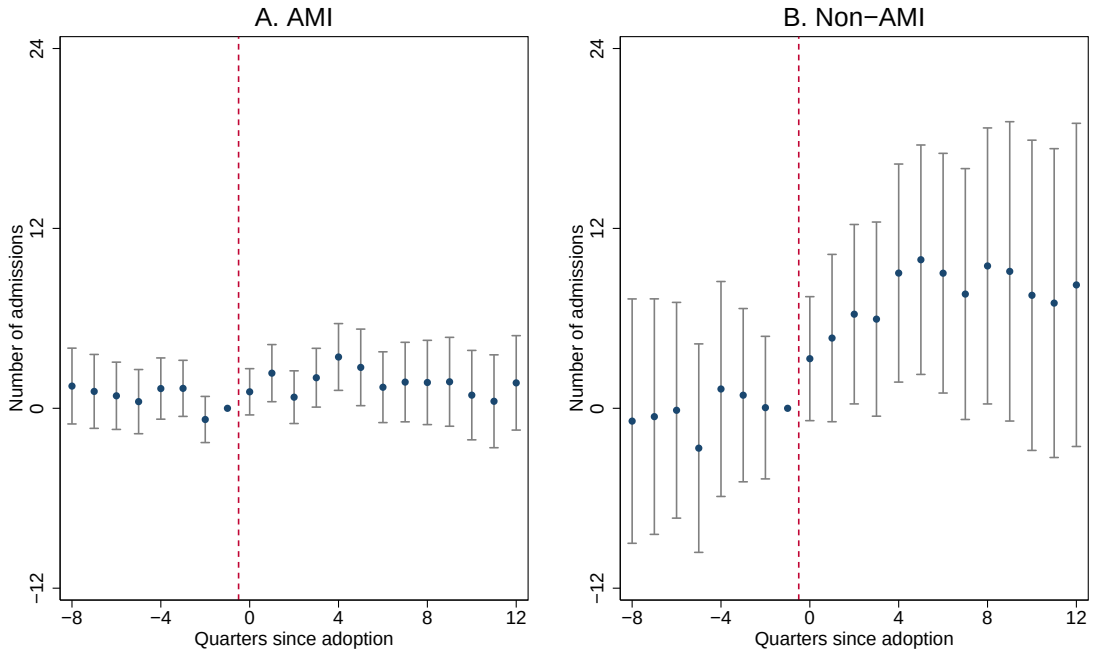
[7]ICD-9-CM diagnosis codes for unstable angina do exist. However, other features of the records on which these codes are used suggest that they are frequently planned, non-emergency admissions and, thus, not incidents of ACS.

Figure 2.4: Dynamic difference-in-differences estimates of the effect of PCI adoption on the adopting hospital's PCI volume by primary diagnosis



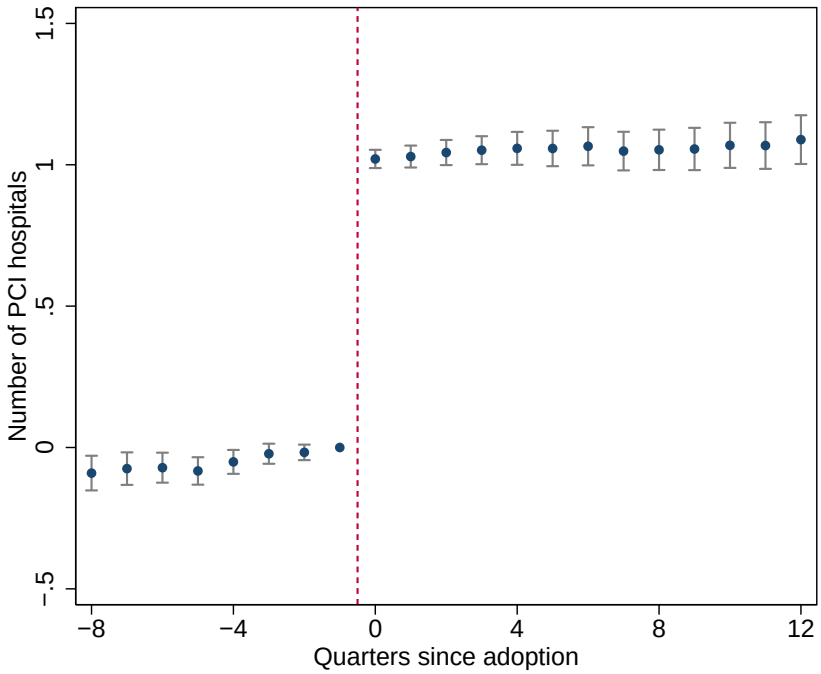
Notes: This figure plots estimates of the coefficients $\{\tau_d\}_{d\in\{-8,-7,\ldots,11,12\}}$ obtained from estimating equation (2.2); as in equation (2.2), the coefficient for $d = -1$ is normalized to zero. The dependent variable is the total number of PCIs performed on patients with the listed primary diagnosis in each hospital-quarter. AMI is an abbreviation for acute myocardial infarction (heart attack). The error bars depict 95 percent confidence intervals for the plotted point estimate and are computed using standard errors clustered at the hospital level.

Figure 2.5: Dynamic difference-in-differences estimates of the effect of PCI adoption on market-wide PCI volume by primary diagnosis



Notes: This figure plots estimates of the coefficients $\{\tau_d\}_{d\in\{-8,-7,\ldots,11,12\}}$ obtained from estimating equation (2.2); as in equation (2.2), the coefficient for $d = -1$ is normalized to zero. The dependent variable is the total number of PCIs performed on patients with the listed primary diagnosis at the hospital or its competitors in each hospital-quarter (where the set of competitors is defined as described in Section 2.3). AMI is an abbreviation for acute myocardial infarction (heart attack). The error bars depict 95 percent confidence intervals for the plotted point estimate and are computed using standard errors clustered at the hospital level.

Figure 2.6: Dynamic difference-in-differences estimates of the effect of PCI adoption on the market-wide number of PCI programs



Notes: This figure plots estimates of the coefficients $\{\tau_d\}_{d\in\{-8,-7,\ldots,11,12\}}$ obtained from estimating equation (2.2); as in equation (2.2), the coefficient for $d = -1$ is normalized to zero. The dependent variable is the total number of active PCI programs among the hospital itself and its competitors in each hospital-quarter (where the set of competitors is defined as described in Section 2.3). The error bars depict 95 percent confidence intervals for the plotted point estimate and are computed using standard errors clustered at the hospital level.

with the market-wide number of PCI programs on the left-hand-side. The results are reported in Figure 2.6 and column (1) of Table 2.3. We find that entry coincides with an immediate and permanent increase in the market-wide number of hospitals with PCI capabilities of almost exactly one, so we conclude that there is no evidence that a given hospital's entry has a deterrent effect on its competitors.

The second issue we explore is whether the increased PCI utilization we observe reflects substitution away from the main alternative surgical treatment for coronary artery disease, coronary artery bypass graft (CABG) surgery. Cutler and Huckman (2003) report evidence suggesting just such a pattern of substitution as PCI diffused in New York. If substantial substitution were indeed occurring, it would be important to account for the reduction in resources devoted to CABG in the welfare calculations presented in the next section. To quantify any such substitution, we once again

Table 2.3: Summary difference-in-differences estimates of the effect of PCI adoption on the hospital-level and market-wide number of PCI programs and CABG volume

Column:	(1)	(2)
	Dependent variable	
Analysis level	# of PCI hospitals	CABG volume
Hospital-level	N/A	10.9***
		(0.7)
		[0.00]
Market-wide	1.1***	-0.5
	(0.0)	(2.4)
	[0.01]	[0.05]
Hospital-quarter observations	209,015	209,015

Notes: This table reports summary results from estimating equation (2.2). For each regression, the reported point estimate is the "pooled" medium-run volume effect defined in (2.3). For the analysis presented in column (1), the dependent variable is the total number of hospitals with PCI programs in that quarter (where the set of competitors is defined as described in Section 2.3). For the market-level analysis presented in column (2), the dependent variable is the total number of CABGs performed in each hospital-quarter. For the market-wide analysis presented in column (2), the dependent variable is the total number of CABGs performed at the hospital or its competitors in each hospital-quarter. The p-value from a test of the null hypothesis of pre-period common trends (i.e. that $\tau_2 = \tau_3 = \cdots = \tau_8 = 0$) is displayed in brackets. Standard errors clustered at the hospital level are reported in parentheses. Statistical significance is denoted as follows: * $p < .05$, ** $p < .01$, *** $p < .001$.

re-estimate equation (2.2), first with CABG volume of each individual hospital on the left-hand-side and then with market-wide CABG volume on the left-hand-side.

The results are reported in Figure 2.7 and in column (2) of Table 2.3. We see that PCI-adopting hospitals actually experience a substantial *increase* in CABG volume of approximately 10 procedures per quarter; it appears that hospital PCI adoption frequently coincides with creation or expansion of hospitals' cardiac surgery programs.[8] While the market-level results suggest that all of this increase in CABG utilization reflects business stealing, we can safely conclude that there is no evidence that the market-wide increase in PCI utilization causes substitution away from CABG.[9]

Finally, we present two sets of specification checks. In our first set of checks, we evaluate the plausibility of the common trends assumption that justifies estimating equation (2.2). To do so, we examine the coefficients on the leads of PCI adoption included in equation (2.2). If the common trends assumption holds, these coefficient estimates should all be approximately equal to zero (and, consistent with that, exhibit no clear trend). Visual examination of Figures 2.2-2.5 suggests that the estimated coefficients do indeed satisfy this condition. We can also formally test the hypothesis that the coefficients on the leads are all zero. The p-values from joint F-tests of these restrictions are reported in brackets in the relevant cells of Table 2.2.[10] The p-values only approach standard thresholds for statistical significance in the hospital-level specifications, and, in these cases, the point estimates demonstrate that any pre-trend that does exist is exceedingly small.

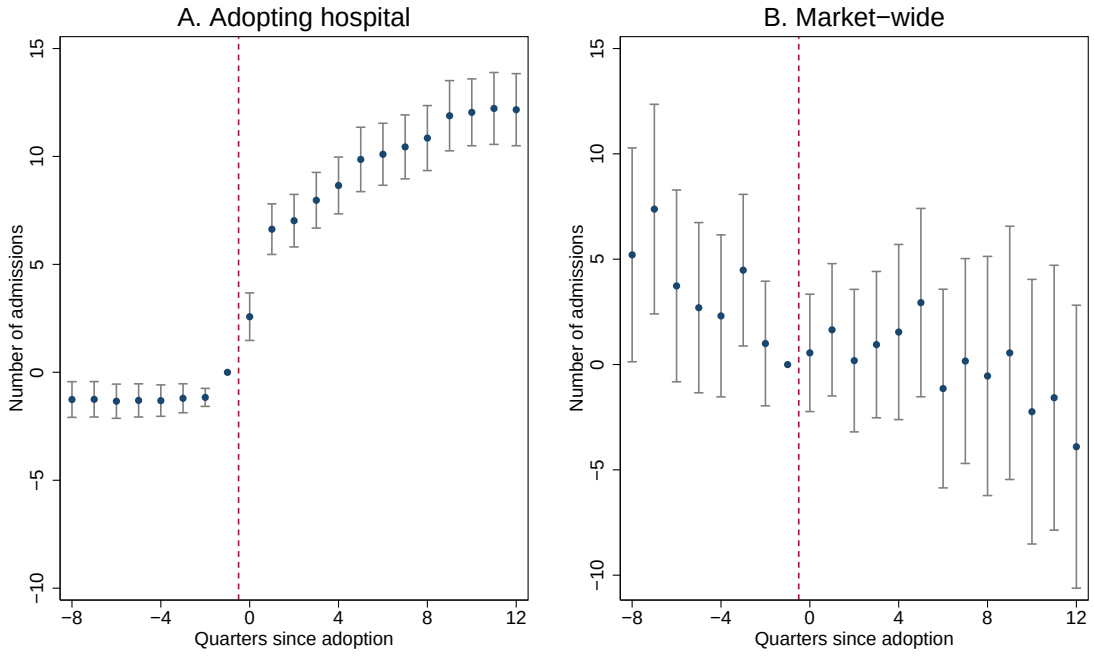
In our second set of checks, we examine whether the market-level results are sensitive to the

[8]This pattern of coincidental adoption may reflect PCI-adopting hospitals' desire to have emergency CABG "backup" on hand in the event of complications during PCI. Early clinical guidelines (e.g. Ryan et al. (1993)) recommended that all hospitals offering elective PCI be able to perform on-site CABG, and this recommendation has only been meaningfully weakened recently (e.g. Levine et al. (2011)). Although adherence to this rule is not universal, many hospitals do appear to abide by it. Alternatively, it may be that hospitals adopting PCI do so in the midst of a broader effort to expand their line of cardiology services.

[9]In fact, although the point estimate reported in Table 2.3 is slightly negative, there is visual evidence of a downward trend in market-level CABG volume for adopting hospitals, and the p-value reported in Table 2.3 indicates that this trend is statistically significant. Extrapolating this trend suggests that the markets of hospitals adopting PCI actually experience a small *increase* in CABG volume. In any case, there is no evidence of net substitution away from CABG.

[10]When running these joint hypothesis tests, we do not include the lead corresponding to quarters more than 8 quarters before adoption: τ_{-9} in equation (2.2). The F-tests for the market-level specifications would typically reject if this coefficient were included because τ_{-9} is generally negative and statistically significant. This indicates that market-level volumes for PCI adopting hospitals were trending upward during the period more than two years prior to adoption. We do not view this earlier pre-trend as cause for concern given that the trend has ceased by the time of PCI adoption.

Figure 2.7: Dynamic difference-in-differences estimates of the effect of PCI adoption on hospital-level and market-wide CABG volumes



Notes: This figure plots estimates of the coefficients $\{\tau_d\}_{d\in\{-8,-7,\ldots,11,12\}}$ obtained from estimating equation (2.2); as in equation (2.2), the coefficient for $d = -1$ is normalized to zero. The dependent variable in panel A is the total number of CABGs performed in each hospital-quarter. The dependent variable in panel B is the total number of CABGs performed at the hospital or or its competitors in each hospital-quarter (where the set of competitors is defined as described in Section 2.3). The error bars depict 95 percent confidence intervals for the plotted point estimate and are computed using standard errors clustered at the hospital level.

threshold used to "trim" zipcode market shares when defining each hospital's competitor set. Table 2.4 reports the market-level results for the base 1 percent threshold as well as for a trimming threshold half as large and a trimming threshold twice as large. The table indicates that the point estimates are qualitatively similar no matter the threshold used. Smaller thresholds substantially reduce precision, however, and when the smaller threshold considered is used, the market-wide increases in overall PCI volume and non-AMI PCI volume are no longer statistically significant.[11]

2.6 Welfare analysis

We now turn to the question of whether entry over this period improved welfare, which we address by calibrating equation (2.1) from Section 2.2. In the first several parts of this section, we use the utilization estimates from the last section and existing evidence on the benefits and costs of PCI to pin down the unknown quantities in equation (2.1). The resulting parameter values are are summarized in Table 2.5. Then, in the last part of this section, we use these parameters to estimate the welfare gain or loss resulting from entry.

2.6.1 Calibrating the utilization response

We start by selecting, for each patient type $\theta \in \{\text{AMI}, \text{non-AMI}\}$, values for $(v_{N+1}(\theta) - v_N(\theta))\mu(\theta)$, which is the change in the flow of PCI utilization for patients of type θ. This quantity can be computed directly from the market-wide utilization results presented in the last section. Specifically, we start with the long-run change in market-wide Medicare PCI utilization reported in Table 2.2. We then scale these estimates up to account for the share of patients covered by insurers other than Medicare. Using the National Hospital Discharge Survey (NHDS) for 1992-2005 (the years included in the sample used to estimate equation (2.2)), we find that Medicare patients account for 42 percent of PCI utilization among patients with an AMI primary diagnosis and 50 percent of PCI utilization among patients with a non-AMI primary diagnosis.[12] The resulting estimates of the

[11] The loss in precision with a lower threshold occurs because the number of hospitals in each provider's competitor set grows as the threshold falls. A great number of competitors increases the scale and, therefore, the variance of market-wide volume, which decreases the precision of the estimated effects.

[12] The NHDS data files were obtained from the the Inter-university Consortium for Political and Social Research, study number 24281.

Table 2.4: Sensitivity analysis of the estimated effect of PCI adoption of hospital-level and market-wide PCI volumes

Column:	(1)	(2)	(3)
	Primary diagnosis		
Analysis level	All	AMI	Non-AMI
Hospital-level	25.0***	7.6***	17.4***
	(1.3)	(0.3)	(1.0)
	[0.10]	[0.05]	[0.13]
Market-wide (0.5 percent threshold)	8.4	2.0	6.3
	(6.0)	(1.5)	(4.9)
	[0.12]	[0.09]	[0.23]
Market-wide (1 percent threshold)	10.3*	1.8	8.6*
	(4.9)	(1.2)	(4.0)
	[0.45]	[0.34]	[0.62]
Market-wide (2 percent threshold)	10.9**	2.1*	8.8*
	(4.1)	(1.0)	(3.5)
	[0.64]	[0.42]	[0.83]
Hospital-quarter observations	209,015	209,015	209,015

Notes: This table reports summary results from estimating equation (2.2). For each analysis level and primary diagnosis level, the reported point estimate is the "pooled" medium-run volume effect defined in (2.3). For the hospital-level analyses, the dependent variable is the total number of PCIs performed on patients with the listed primary diagnosis in each hospital-quarter. For the market-wide analyses, the dependent variable is the total number of PCIs performed on patients with the listed primary diagnosis at the hospital or its competitors in each hospital-quarter (where the set of competitors is defined as described in Section 2.3 using the "trimming" threshold listed). AMI is an abbreviation for acute myocardial infarction (heart attack). The p-value from a test of the null hypothesis of pre-period common trends (i.e. that $\tau_2 = \tau_3 = \cdots = \tau_8 = 0$) is displayed in brackets. Standard errors clustered at the hospital level are reported in parentheses. Statistical significance is denoted as follows: * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 2.5: Base parameters for welfare calculations

	Diagnosis group		
Parameter	AMI	Non-AMI	**Source**
A. Change in all-payer quarterly PCI utilization			
Market-wide	4.2	17.0	Estimates in Table 2.2 and NHDS.
	(2.8)	(8.0)	
Hospital-level	18.1	34.6	Estimates in Table 2.2 and NHDS.
	(0.8)	(2.0)	
B. Health benefits of PCI			
Short-run survival without PCI (s_θ^0)	0.87	1	AMI: Keeley et al. (2003); Non-AMI: Several sources. See text.
Short-run survival gain with PCI (Δ_θ^s)	0.04	0	AMI: Keeley et al. (2003); Non-AMI: Several sources. See text.
Mortality hazard (λ_θ^s)	0.081	0.073	CDC and NHDS data and Cutler et al. (1998).
Angina incidence without PCI (a_θ^0)	0.39	N/A	AMI: Keeley et al. (2003); Non-AMI: N/A since $\Delta^s_{\text{Non-AMI}} = 0$.
Short-run angina reduction with PCI (Δ_θ^a)	0.16	0.14	AMI: Keeley et al. (2003); Non-AMI: RITA-2 and COURAGE.
Angina benefit decay rate (λ_θ^a)	0.281	0.281	RITA-2 and COURAGE trials.
QALY cost of angina (ν_a)	0.2		Tengs and Wallace (2000).
Value of a QALY	\$100,000		Cutler (2004).
C. Costs of PCI			
Marginal cost ($c(\theta)$)	\$4,942	\$11,675	Calculations based on Medicare payment schedule.
Fixed cost (k)	\$3 million		Lieu et al. (1996) and Gehrki (2004).
Equipment lifetime ($\bar{T}$)	10 years		Lieu et al. (1996).
D. Other parameters			
Discount rate (r)	0.03		Conventional.
Medicare marginal revenue from PCI	\$5,814	\$13,736	Calculations based on Medicare payment schedule.
Private-insurer marginal revenue from PCI	\$9,535	\$22,527	Calc. based on Medicare payment schedule and Robinson (2011).

Notes: These parameters are used for the base case in the welfare calculations presented in Table 2.6. The detailed justification for the selection of each set of parameters is provided in the text. Standard errors for the utilization responses are reported in parentheses.

market-wide, all-insurer increase in quarterly PCI utilization are reported in panel A of Table 2.5.

2.6.2 Calibrating the health benefits

We next turn to calibrating the health benefits of PCI, $\tau(\theta)$. Receipt of PCI may improve survival, improve quality of life, or both; the main source of quality-of-life improvements in this context is a reduction in the incidence of angina. In order to monetize these benefits, we follow a standard approach in health economics and compute the number of quality-adjusted-life-years (QALYs) generated by each treatment path.[13]

We use the following simple framework for our QALY calculations. For a patient of type θ with treatment status $p \in \{0, 1\}$, the probability of survival until time t is denoted by $S^p_\theta(t)$, where $t = 0$ is the time of PCI. The incidence of angina among those alive at each time t is given by $A^p_\theta(t)$. Angina is assumed to carry a penalty in QALY units of ν_a. The discounted number of quality-adjusted life years (QALYs) gained by switching from pure medical management to PCI is therefore given by

$$\tau(\theta) = \int_0^\infty [1 - \nu_a A^1_\theta(t)] S^1_\theta(t) e^{-rt} dt - \int_0^\infty [1 - \nu_a A^0_\theta(t)] S^0_\theta(t) e^{-rt} dt, \tag{2.4}$$

where r is the social discount rate as in equation (2.1). We set $r = 0.03$ throughout.

To facilitate calibration, we specify simple parametric functional forms for the survivor functions $S^p_\theta(t)$ and the angina incidence paths $A^p_\theta(t)$, taking care to choose functions that are flexible enough to capture the salient features of the clinical literature. We assume that the survival function in the absence of PCI takes the form $S^0_\theta(t) = s^0_\theta \exp[-\lambda^s_\theta t]$, which features an initial instantaneous survival probability of s^0_θ and a subsequent mortality hazard of λ^s_θ. We further suppose that any survival benefits of PCI appear immediately after treatment and that any new survivors have a life expectancy similar to that of other survivors. This implies that the survival function for individuals receiving PCI can be written as $S_1(t) = [s^0_\theta + \Delta^s_\theta] \exp[-\lambda^s_\theta t]$, where Δ^s_θ can be interpreted as the short-run survival benefit of PCI. We parametrize angina incidence by $A^p_\theta(t) = a^0_\theta + p\Delta^a_\theta \exp[-\lambda^a_\theta t]$, which implies that PCI causes an immediate reduction in angina incidence of size Δ^a_θ that decays

[13] The number of QALYs associated with spending a given path of health states is defined as the number of years of life in perfect health that an individual would accept in exchange for that path. See Dolan (2000) for a review of this approach to valuation.

over time at a rate λ_θ^a.

Substituting these functional forms into equation (2.4) and doing a small amount of algebra, we obtain the following expression for $\tau(\theta)$:

$$\tau(\theta) = \frac{\Delta_\theta^s(1 - \nu_a a_\theta^0)}{\lambda_\theta^s + r} + \frac{\Delta_a^\theta \nu_a (s_\theta^0 + \Delta_\theta^s)}{\lambda_\theta^s + \lambda_\theta^a + r}. \tag{2.5}$$

The first term of equation (2.5) represents the QALY value of any survival improvement, while the second represents the QALY value of any quality-of-life gains. We obtain values for the parameters that appear in equation (2.5) from the clinical literature.

Clinical benefits for AMI patients

For patients with AMI, the clinical evidence is clear that PCI improves both the length and quality of life. The most readily applicable evidence is for patients with the most serious type of AMI, ST-elevation myocardial infarction (STEMI), which is named for the electrocardiogram abnormality that distinguishes it from other AMIs. For these patients, the authoritative meta-analysis by Keeley et al. (2003) finds that prompt receipt of PCI reduces mortality at 6-18 months by 4 percentage points from 13 percent to 9 percent, so we set $\Delta_{\text{AMI}}^s = 0.04$ and $s_{\text{AMI}}^0 = 0.87$. Keeley et al. (2003) also find that PCI reduces the incidence of recurrent ischemia (i.e. new incidents of insufficient coronary blood flow) by 16 percentage points from 39 percent to 23 percent. Recurrent ischemia can take a variety of forms, ranging from angina up to a new AMI. Likely conservatively, we treat these reductions in recurrent ischemia as reflecting reductions in the incidence of persistent angina and thus set $\Delta_{\text{AMI}}^a = 0.16$ and $a_{\text{AMI}}^0 = 0.39$. There is little evidence available on the persistence of these gains, so we calibrate λ_{AMI}^a on the basis of the evidence for non-AMI patients that is discussed below.[14]

[14]In evaluating this parametrization, one important question is whether the survival and quality-of-life benefits of PCI in STEMI patients are a reasonable guide to the benefits realized by other heart attack (known as NSTEMI) patients. Answering this question is complicated by the fact that the available trials for NSTEMI patients do not directly compare PCI with medical management. Rather, most of the available trials compare two broad treatment regimes: a "conservative" regime emphasizing medical management; and an "invasive" regime emphasizing revascularization (i.e. PCI and bypass surgery).

Bavry et al. (2006) present a comprehensive meta-analysis of such trials and concludes that treatment according to the "invasive" pathway reduces mortality by 1.6 percentage points and the incidence angina-related hospitalizations by 7.8 percentage points, both at a mean follow-up period of 24 months. In light of the fact that patients in the "invasive" arm of these trials are only 25 percentage points more likely to receive revascularization and assuming that the benefits found are not disproportionately due to bypass surgery rather than PCI, these results suggests that the benefits of PCI for NSTEMI patients are similar in magnitude to those for STEMI patients.

To calibrate the mortality hazard λ^s_{AMI}, we turn once again to the NHDS and calculate that the average age of AMI patients receiving PCI during the study period was 62.2 years. By interpolating life tables created by the CDC (Arias, 2012) for 1999-2001, we compute that the typical individual of this age during this period had a remaining life expectancy of 19.8 years. In a Medicare cohort, Cutler et al. (1998) find that life expectancy after an AMI is approximately 55 percent of that for the general population.[15] Applying this same factor, we compute a life expectancy for the patients in this study of 10.9 years. The NHDS indicates that 23 percent of AMI patients received PCI during the study period, which allows us to calculate that $\lambda^s_{\text{AMI}} = 0.081$.

Clinical benefits for non-AMI patients

By contrast to AMI patients, the benefits of PCI for non-AMI patients – almost entirely patients with "stable" coronary artery disease – appear to be considerably smaller.[16] For these patients, the uniform conclusion of recent meta-analyses (Wijeysundera et al., 2010; Steriopulos and Brown, 2012; Thomas et al., 2013) and the position taken by current practice guidelines (Fihn et al., 2012) is that PCI provides no improvement in survival relative to medical management. Thus, we set $\Delta^s_{\text{Non-AMI}} = 0$, and since early mortality is negligible for these patients, we set $s^0_{\text{Non-AMI}} = 1$.

While the consensus among cardiologists is that PCI does not improve survival in this group of patients, it is generally thought to reduce the incidence of angina. To quantify these benefits, we turn to the two largest trials in this area: the RITA-2 trial (Henderson et al., 2003) and the COURAGE trial (Weintraub et al., 2008).[17] RITA-2 reports that PCI reduces the incidence

A meta-analysis by Hoenig et al. (2010) presents a dissenting view and concludes that there is no statistically significant evidence of reductions in mortality, recurrent AMI, or recurrent angina among patients randomized to invasive or conservative pathways. However, the Hoenig et al. point estimates are similar to those presented by Bavry et al. (2006) as well as those obtained by Mehta et al. (2005) and Biondi-Zoccai et al. (2005) and fail to reach statistical significance mainly because Hoenig et al. include a narrower set of studies.

[15] We could directly repeat the Cutler et al. (1998) exercise in our Medicare data, and we may do so in future work.

[16] Note, once again, that due to data limitations, the non-AMI category includes some patients experiencing true unstable angina. The benefits of PCI for these patients are likely more similar to the benefits for AMI patients than to the benefits for the rest of the non-AMI group. They are, however, a small portion of the non-AMI group.

[17] As noted above, there are three major meta-analyses in this literature: Wijeysundera et al. (2010), Steriopulos and Brown (2012), and Thomas et al. (2013). These analyses reach varying conclusions regarding the effect of PCI on the incidence of angina. These analyses share a key flaw, however, that they combine studies using widely-differing follow-up horizons. As the discussion of RITA-2 and COURAGE indicates, the effect of PCI on angina appears to decay relatively rapidly, which is likely to render the results of this type of meta-analysis uninterpretable.

of angina by 17 percentage points at 3 months, but by just 6 percentage points after 5 years. COURAGE reports that PCI reduces the incidence of angina after 3 months by 11 percentage points, but by just 3 percentage points after 3 years. On the basis of this evidence, we set the initial reduction in the incidence of angina at an intermediate value of $\Delta_{\text{Non-AMI}} = 0.14$ and choose a decay parameter intermediate between those implied by the two studies of $\lambda^a_{\text{Non-AMI}} = 0.281$.[18] Because, as noted above, there is no evidence on the rate at which the angina benefits to AMI patients decay over time, we apply this same decay parameter to those patients.

To calibrate the mortality hazard $\lambda^s_{\text{Non-AMI}}$, we use an approach very similar to that used for AMI patients. NHDS calculations indicate that the average age of the non-AMI patients was 64.4 years. The corresponding life expectancy in the general population from the CDC life tables is 18.2 years. We suppose that these patients' life expectancy is 75 percent of that of individuals in the general population, which is higher than the 55 percent figure from Cutler et al. (1998) that we used in the calculation for the AMI patients. We use the higher value here to reflect the fact that these patients are likely somewhat healthier than the AMI population. Given the arbitrariness of this choice, we explore the sensitivity of our results to alternative assumptions.[19] Our base assumption yields a value for the mortality hazard of $\lambda^s_{\text{Non-AMI}} = 0.073$.

Parameters common to AMI and non-AMI patients

We draw two final parameters from other sources. For the QALY penalty associated with chronic angina (ν_a), we follow Cutler and Huckman (2003), who use an estimate based on Tengs and Wallace (2000) that a year with moderate angina is worth 0.2 QALYs less than a year in perfect health. Since some of the reductions in angina incidence observed in the clinical literature, particularly for non-AMI patients, may reflect milder angina, we explore the robustness of our results to smaller values. To convert the resulting QALY totals into monetary terms, we assume that a year of life in perfect health is valued at $100,000, following Cutler (2004). Murphy and Topel (2006) argue that considerably higher values may be appropriate, and we consider the sensitivity of our results to an

[18] We compute the average decay parameter between the two studies as the harmonic mean rather than the simple arithmetic mean since the inverse of the decay parameter is the more meaningful quantity in our setting.

[19] We could reduce the arbitrariness here by repeating the Cutler et al. (1998) exercise in the present sample. We may do so in future work

alternative value of $200,000.

2.6.3 Calibrating costs

We base our estimates of the marginal costs of PCI, $c(\theta)$, on the amount paid by Medicare, which, at least in principle, approximate provider costs (Newhouse, 2002). For non-AMI patients, we start with the average payment that Medicare makes for a PCI hospitalization during fiscal year 2013. For AMI patients, we wish to account for the fact that these patients would likely have already been hospitalized and, thus, would have incurred the basic costs of hospitalization even without PCI. Thus, for AMI patients, we obtain our initial estimate by subtracting the average payment Medicare makes for a medically-managed AMI hospitalization from the average payment for a PCI hospitalization.[20] We add an additional $1,000 to each of these figures to account for the cost of physician services that are reimbursed separately. The fact that providers are willing to accept this reimbursement to perform PCI suggests that these amounts exceed hospitals' marginal costs of PCI. We therefore reduce these figures by 15 percent to obtain our final cost estimates, which are displayed in Panel C of Table 2.5.

We consult two sources to obtain an estimate of the capital investment required to enter the PCI market. A 2004 article in a major industry newsletter (Gehrki, 2004) reports total construction and equipment costs for a new catheterization facility of $2-2.5 million; taking the middle of the range and adjusting for inflation, this suggests a total cost of $2.7 million in today's dollars. Lieu et al. (1996) report an estimated equipment cost of $1.6 million and an estimated construction cost of $1.5 million, both in 1993 dollars. They assume that the equipment investment needs to be repeated every 10 years and the construction costs every 30 years. Under the assumption that the hospital can continue to make use of the building even if decides to exit to the PCI market after 10 years, this implies a total cost in today's dollars of $3.5 million.[21] Guided by these estimates, we use an intermediate figure of $k = \$3$ million, and we set the replacement horizon to $\bar{T} = 10$ years

[20] In detail, we compute the average cost of a PCI hospitalization by computing the average payment level across DRGs 246-251, weighting each DRG's contribution to the average by the number of admissions categorized in that DRG during fiscal year 2012. We obtain the average payment for a medically-managed AMI hospitalization by averaging across DRGs 280-285, weighting each DRG's contribution in the same fashion. The data required for these calculations are available for download from the Center for Medicare and Medicaid Services website.

[21] For this calculation, we use a 3 percent discount rate and assume the building depreciates linearly.

as suggested by Lieu et al. (1996).

2.6.4 Welfare analysis results and discussion

The first row of Table 2.6 reports the welfare estimates we obtain under the base parameterization. Focusing first on the net benefits of a single PCI procedure (in columns (1) and (2) of the table), we see that PCI for AMI and non-AMI patients has markedly different welfare implications. For AMI patients, a single procedure generates net benefits of almost $36,000, while, for non-AMI patients, PCI generates a net loss of approximately $4,400. Turning to the cumulative social welfare effects of treatment (in columns (3) and (4) of the table), we see that the cumulative net benefits for AMI patients reach $5.2 million per adopter despite the relatively modest utilization response we estimated for this group, an amount more than enough to cover the fixed costs of adoption. The social surplus generated by this group is, however, completely offset by the $2.6 million per entrant welfare loss associated with the increase in PCI utilization among non-AMI patients. As a result, our best estimate is that adoption was approximately welfare neutral over this period. As indicated by the reported standard errors, this estimate is subject to considerable uncertainty (primarily attributable to uncertainty in the estimated effect of entry on PCI utilization by AMI patients), and we cannot rule out the possibility that adoption either increased or reduced welfare by economically meaningful amounts. We can, however, conclude with high confidence that the net benefits of entry would have been larger (and, more likely than not, positive) had the utilization increases been limited to AMI patients.

In Table 2.7, we explore the sensitivity of our basic social welfare conclusions to alternative parameterizations. The results demonstrate that our basic conclusions – that adoption was most likely close to welfare neutral and that the welfare costs of increased utilization of PCI by non-AMI patients are sizable – are robust. The one exception to this general pattern is when we increase the value of a year of life from $100,000 to $200,000. In this case, even modest reductions in the incidence of angina for non-AMI patients are sufficient to justify the cost of providing PCI, and, as a result, the overall net benefits swing to being strongly positive.

Our finding that increased PCI utilization among non-AMI patients reduced welfare suggests that it would have been socially desirable for hospitals to have forgone treating these patients. This leads naturally to the question of how efforts to discourage treatment for these patients would affect

Table 2.6: Estimated effect of hospital PCI adoption on social welfare and entrant profits

Column:	(1)	(2)	(3)	(4)	(5)	(6)
	Per case net effect (dollars)		**Cumulative net effect (millions)**		**Setup costs (millions)**	**Total effect (3) + (4) − (5)**
Effect on...	AMI	Non-AMI	AMI	Non-AMI		
Social welfare	35,846	-4,394	5.2	-2.6	3.0	-0.4
			(3.5)	(1.2)		(3.7)
Entrant profits	3,040	6,418	1.9	7.7	3.0	6.6
			(0.1)	(0.4)		(0.5)

Notes: Columns (1) and (2) report the net effect of a single PCI with the listed diagnosis on social welfare or entrant profits. Columns (3) and (4) report the corresponding discounted net effect on social welfare or entrant profits over the facility's lifetime. Column (5) reports the setup cost k. Column (6) reports the total effect on social welfare or profits. Standard errors reflecting uncertainty in the estimated utilization responses are displayed in parentheses.

Table 2.7: Estimated effect of hospital PCI adoption on social welfare under alternative parametrizations

Column:	(1)	(2)	(3)	(4)	(5)	(6)
	Case net benefit (dollars)		**Cumulative net benefits (millions)**		**Setup costs (millions)**	**Welfare effect (3) + (4) − (5)**
Scenario	AMI	Non-AMI	AMI	Non-AMI		
A. Base parametrization						
Base results	35,846	-4,394	5.2	-2.6	3.0	-0.4
			(3.5)	(1.2)		(3.7)
B. Alternative data sources for non-AMI angina benefits						
Based on COURAGE trial only	33,771	-6,455	4.9	-3.8	3.0	-1.9
			(3.3)	(1.8)		(3.7)
Based on RITA-2 trial only	37,548	-762	5.5	-0.4	3.0	2.0
			(3.7)	(0.2)		(3.7)
C. Alternative parameterizations of non-AMI mortality hazard						
Life expectancy 55% of general pop.	35,846	-4,866	5.2	-2.9	3.0	-0.6
			(3.5)	(1.3)		(3.7)
Life expectancy 100% of general pop.	35,846	-4,030	5.2	-2.4	3.0	-0.1
			(3.5)	(1.1)		(3.7)
D. Alternative health benefit valuation parameters						
Value of QALY is $200,000	76,634	2,887	11.2	1.7	3.0	9.9
			(7.5)	(0.8)		(7.5)
QALY cost of angina is half as large	33,541	-8,035	4.9	-4.7	3.0	-2.8
			(3.3)	(2.2)		(4.0)
E. Alternative marginal costs						
25 percent smaller than base case	37,082	-1,475	5.4	-0.9	3.0	1.5
			(3.6)	(0.4)		(3.6)
25 percent higher than base case	34,611	-7,313	5.0	-4.3	3.0	-2.2
			(3.4)	(2.0)		(3.9)
F. Alternative equipment lifetimes						
Lasts only 5 years	35,846	-4,394	2.8	-1.4	3.0	-1.6
			(1.9)	(0.7)		(2.0)
Lasts 15 years	35,846	-4,394	7.3	-3.6	3.0	0.7
			(4.9)	(1.7)		(5.2)

Notes: Columns (1) and (2) report the net benefit $\tau(\theta) - c(\theta)$ of a single PCI with the listed diagnosis. Columns (3) and (4) report the corresponding discounted social benefits over the facility's lifetime. Column (5) reports the setup cost. Column (6) reports the overall welfare effect of entry $\Delta W(N, N+1)$. Standard errors reflecting uncertainty in the estimated utilization responses are displayed in parentheses.

hospital adoption incentives. We can address this question directly using the estimated hospital-level utilization responses reported in Table 2.2 (which we scale up using data from the NHDS in the same fashion as before) together with information on marginal costs and payment rates. The required information on costs and Medicare payment rates was already described in the last subsection. To obtain payment rates relevant to privately-insured patients (who constitute nearly all of the non-Medicare PCI market), we apply an estimate reported by Robinson (2011) that private payment rates for angioplasty exceed Medicare rates by 64 percent. Although the Robinson estimates apply to a selected group of 61 hospitals that participated in one of two value-based purchasing initiatives, they provide a useful starting point, and we examine the sensitivity of our results to alternative parameter choices.

The results of this exercise are reported in the second row of Table 2.6. We see that the additional profits earned from PCI in non-AMI patients are $7.7 million per adopter, far in excess of the profits earned from AMI patients. Moreover, without the profits earned from non-AMI patients, the total returns to adoption would not cover the required setup costs. Observe that these effects on profits are very precisely estimated since they rely only on the hospital-level utilization responses, which are themselves very precisely estimated. As shown in Table 2.8, these conclusions are also highly robust to alternative choices of parameter values.

This pattern of results suggests that policy reforms that seek solely to reduce the utilization of PCI in non-AMI patients (e.g. by lowering payment rates for these patients) cannot achieve the first-best outcome in which hospitals adopt PCI but decline to treat non-AMI patients. Any policy that was sufficiently successful in discouraging use of PCI among non-AMI patients would cause hospitals to decline to adopt PCI altogether, with the result that society would not realize the incipient welfare benefits for AMI patients. Rather, optimal policy would pair measures to reduce utilization of PCI in non-AMI patients with increases in payment rates for AMI patients in order to allow hospitals to capture more of the surplus generated by PCI in the latter group.

2.7 Discussion and conclusion

Our analysis demonstrated that the typical hospital entering the PCI market over the period 1992-2005 engaged in significant business stealing, but still generated a meaningful market-wide increase

Table 2.8: Estimated effect of hospital PCI adoption on entrant profits under alternative parametrizations

Column:	(1)	(2)	(3)	(4)	(5)	(6)
	Per case profit (dollars)		**Cumulative profits (millions)**		**Setup costs (millions)**	**Total profits (3) + (4) − (5)**
Scenario	AMI	Non-AMI	AMI	Non-AMI		
A. Base parametrization						
Base results	3,040	6,418	1.9	7.7	3.0	6.6
			(0.1)	(0.4)		(0.5)
B. Alternative marginal costs						
25 pct. smaller than base case	4,275	9,337	2.7	11.2	3.0	10.8
			(0.1)	(0.6)		(0.7)
25 pct. higher than base case	1,804	3,499	1.1	4.2	3.0	2.3
			(0.1)	(0.2)		(0.2)
C. Alternative equipment lifetimes						
Lasts only 5 years	3,040	6,418	1.0	4.1	3.0	2.1
			(0.0)	(0.2)		(0.2)
Lasts 15 years	3,040	6,418	2.7	10.7	3.0	10.4
			(0.1)	(0.6)		(0.6)
D. Alternative hospital pricing power estimates						
50 pct. over Medicare	2,565	5,465	1.6	6.5	3.0	5.1
			(0.1)	(0.4)		(0.4)
80 pct. over Medicare	3,581	7,508	2.2	9.0	3.0	8.2
			(0.1)	(0.5)		(0.5)

Notes: Columns (1) and (2) report the net effect of a single PCI with the listed diagnosis on entrant profits. Columns (3) and (4) report the corresponding discounted profits over the facility's lifetime. Column (5) reports the setup cost. Column (6) reports the total effect on profits. Standard errors reflecting uncertainty in the estimated utilization responses are displayed in parentheses.

in PCI volume. These increases in utilization were, however, weighted toward patients with low clinical benefits of PCI. As a result, our best estimate is that even though the average hospital adopting PCI generated benefits among AMI patients that were large enough to justify the fixed costs of adoption, overuse among non-AMI patients more than dissipated these social benefits. We also showed that the structure of incentives offered by public and private payers made this pattern of post-adoption behavior highly profitable and rendered efficient post-adoption behavior unprofitable. The implication of our results is that achieving the first-best outcome would require not only reducing hospitals' incentives to perform PCI in non-AMI patients, but also replacing the lost cross-subsidy by increasing payment rates for PCI in AMI patients.

Our conclusions regarding the welfare consequences of the diffusion of PCI differ from prior work examining the experience in New York by Cutler and Huckman (2003). We differ for two main reasons. First, Cutler and Huckman find that, in New York, PCI substituted for bypass surgery as it diffused, which substantially reduced costs; we find no evidence of such substitution here. This difference may arise because Cutler and Huckman examine an earlier period than we examine here or because of differences between New York and the rest of the country. Second, we rely on more recent clinical evidence to quantify the reductions in angina due to PCI in non-AMI patients. This more recent evidence suggests that those benefits are both smaller and, more importantly, much less durable than the evidence cited by Cutler and Huckman.[22]

Our overall conclusion that hospital entry into the PCI market was close to welfare neutral is similar to that reached by Cutler et al. (2010) in their study of entry into the the cardiac surgery market in Pennsylvania during the late 1990s. The factors driving this common conclusion differ, however. Cutler et al. find that the fixed costs of entry were offset by an reallocation of patients to higher-quality surgeons, while in our case the fixed costs of entry were offset by the net benefits of increased utilization.[23]

In closing, we note that the issues raised in this paper are not unique to PCI. The health sector

[22]There is some disagreement on whether the benefits of PCI in non-AMI patients were actually larger in the past or whether the later studies are simply higher quality. See Wijeysundera et al. (2010) for a discussion.

[23]As noted earlier, we have assumed for the purposes of our analysis that the reallocation of patients across providers has no direct welfare consequences. Although we may seek to relax this assumption in future work, we do note that an analysis of the multi-site COURAGE trial of PCI found no evidence of cross-site differences in the benefit of PCI relative to medical therapy (Kolm et al., 2013).

continues to grapple with the diffusion of technologies that feature both high fixed costs and social returns that are highly heterogeneous across patient groups. High-profile current examples include proton beam radiation therapy for cancer and robot-assisted surgery. The experience with PCI suggests that, while these technologies may have great promise, realizing that promise will require careful tailoring of the incentives hospitals face when deciding who to treat.

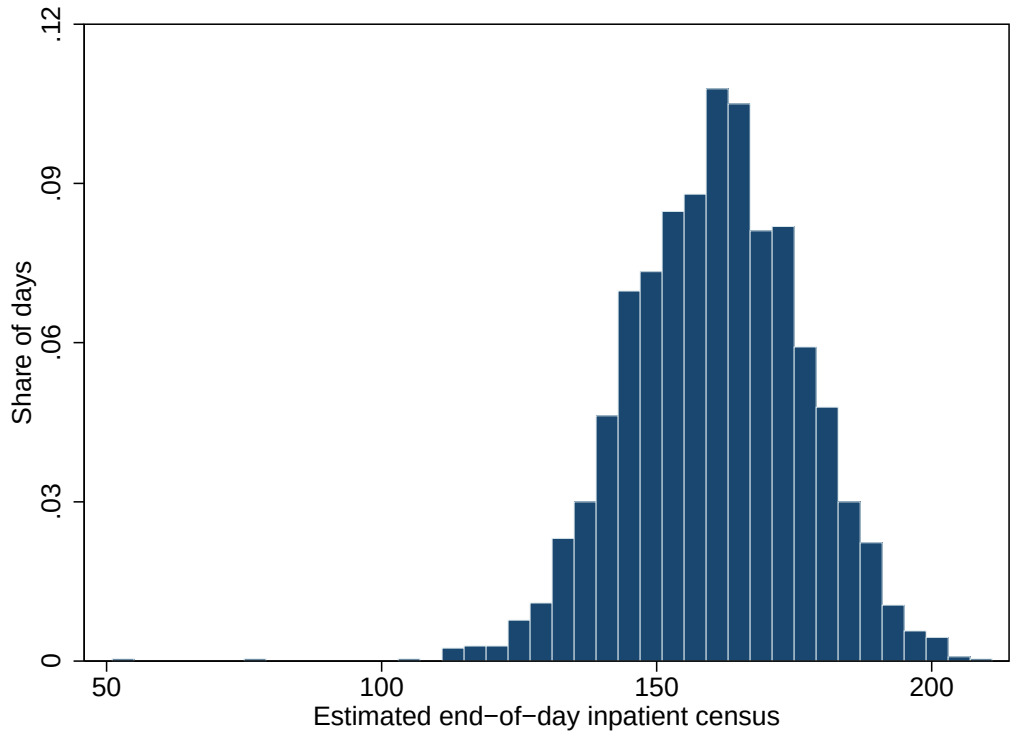
Chapter 3

How does hospital congestion affect care and outcomes for patients with acute coronary illness?

Many major categories of illness arise unpredictably and require immediate treatment. As a result, hospitals face large day-to-day variation in patient demand for their services. For example, Figure 3.1 plots the distribution of end-of-day inpatient censuses for a medium-sized hospital in the state of Massachusetts over several years. The end-of-day census at the 75th percentile of the plotted distribution (171) is 14 percent larger than that at the 25th percentile (150). Patients arriving on a day when the hospital sits at the 25th percentile therefore encounter a hospital that can offer a more generous bundle of per-patient resources than patients who arrive on a 75th percentile day. In this paper, I exploit this quasi-random variation in resource levels to estimate the marginal return of scaling a hospital's per-patient bundle of resources up or down.

Estimates of this marginal return to an increase in per-patient hospital resources are useful for answering at least two types of economic questions. First, these estimates are directly relevant to evaluating the efficiency of hospital capacity investment decisions, which is important for policy analysis of state regulatory ("certificate of need") restrictions on hospital investment and technology acquisition. Second, economic theory often makes sharp predictions about how the productivity of medical care varies across medical providers (e.g. Chandra and Staiger (2007)). Because it exploits

Figure 3.1: Distribution of end-of-day inpatient census at a medium-sized hospital



Notes: This figure plots the distribution of the total number of inpatients present at the end of each day in a medium-sized hospital in Massachusetts from calendar quarter 2002Q4 through calendar quarter 2009Q2. The dataset used to compute the quantities plotted in this figure are described in detail in Section 3.3.

within-provider variation in the resource intensity of treatment, my research design can provide estimates of how the marginal product of additional resources varies across different categories of providers, which can be used to directly test such theories.

I implement my empirical approach in data covering the universe of emergency department, outpatient observation, and inpatient stays that occurred in Massachusetts during the years 2003-2009. These data provide basic information on patient characteristics and detailed information on the care patients receive during any given stay. I am also able to link these hospital records to the universe of death certificates for Massachusetts for the years 2002-2009, which permits me to examine the effect of congestion on post-discharge mortality outcomes. Using these data, I construct an analysis sample of new hospital arrivals who are experiencing acute coronary illnesses. I focus on this group of patients because they are a large and important group of patients for whom the timing of hospital presentation is likely to be approximately random and for whom timely

receipt of care is important.

To provide a framework for my empirical analysis, I develop a simple structural model that permits me to clearly state the conditions under which day-to-day variation in hospital census can be used to identify the effect of resource congestion on patient care and outcomes. The key assumption is that, after purging variation associated with calendar date and a rich set of hospital-time fixed effects (in my case, hospital-year, hospital-month, and hospital-day-of-week effects), the residual variation in hospital census is independent of both: (1) the characteristics of newly arriving patients; and (2) factors determining the hospital's staffing choice set. (The importance of this latter independence assumption has been overlooked in the prior literature using similar research designs.) I then use my model to derive the main estimating equation used in this paper, which relates patient care and outcomes to two dimensions of hospital congestion: the patient census in the hospital's intensive care unit (ICU); and the patient census outside of the hospital's ICU. I parametrize each census measure as a share of its mean for the appropriate hospital-year, rescaled by the interquartile range of this share in the full sample; a one unit change in these "scaled census" variables can therefore be interpreted as approximately equivalent to moving from the 25th to the 75th percentile of demand for hospital resources.

Moving to the results, I first present evidence on the validity of the identifying assumptions. I demonstrate that shocks to residual hospital census fade rapidly and completely, which is consistent with the hypothesis that they are driven by high-frequency random variation in patient demand, rather than low-frequency variation in hospital capacity or patient demand. This is reassuring, as the presence of variation from either of the latter two sources would likely invalidate the research design. I next present direct evidence on both aspects of the identifying assumption. In particular, I demonstrate that the demographic and clinical characteristics of newly arriving patients are independent of current hospital census. I also show that the total number of scheduled surgeries done at the hospital and the number of physicians present at the hospital are only very weakly related to hospital census, which suggests that hospital staffing options are similar on high and low congestion days.

I then examine the effect of congestion on an arriving patient's probability of being admitted to the hospital's ICU. I find that a one unit change in scaled ICU census reduces the probability of ICU admission by one percentage point. The reduction in the probability of ICU admission is

proportionally largest for the least sick patients, which is qualitatively consistent with a model in which hospitals ration ICU beds by expected clinical benefit. I find no effect of non-ICU census on the probability of ICU admission, but higher patient census in the rest of the hospital does appear to very slightly lengthen the duration of a patient's ICU stay. I also demonstrate that the effect of rising ICU census on the probability of ICU admission is highly non-linear, as should be expected given that the hospitals face a hard capacity constraint in the number of ICU beds.

I next investigate how congestion affects procedure receipt, focusing on cardiac catheterization and percutaneous coronary intervention (PCI), two important procedures for this pouplation of patients. Focusing on the sub-group of hospitals with full catheterization capabilities, I find that a one unit increase in scaled ICU census reduces the probability of catheterization and PCI receipt 1-2 days after arrival by one percentage point. Effects on receipt of cardiac catheterization fade by the time of discharge, but a substantial fraction of the effect on PCI persists. The patients for whom prompt receipt of catheterization and PCI has the largest returns appear to face smaller (or no) delays in the receipt of care, once again consistent with efficient rationing. These results are, however, noisy.

Finally, I examine the effect of congestion on patient health outcomes, specifically, patient mortality and subsequent hospital admission. I examine both outcomes at horizons of up to one year. I find no statistically significant evidence that congestion worsens patient outcomes, but I am unable to rule out effects that would be clinically and economically significant.

A substantial prior literature in medicine and operations research uses related empirical strategies to identify the effect of congestion on patient care and outcomes (Strauss et al., 1986; Selker et al., 1987; Iapichino et al., 2004; Richardson, 2006; Pines et al., 2006, 2007, 2009; Fee et al., 2007; Kc and Terwiesch, 2009; Anderson et al., 2011, 2012; Stelfox et al., 2012; Kim et al., 2013; Sun et al., 2013). In economics, two papers have taken closely-related approaches. Bartel et al. (2011) use monthly variation in nurse-to-patient ratios to estimate the effect of nurse staffing on patient outcomes and find that higher nurse staffing levels improve patient outcomes. Freedman (2012) uses variation in the availability of neonatal intensive care unit (NICU) beds at the time of birth to estimate the effect of NICU availability on NICU utilization and patient outcomes. He finds that bed availability has an important effect on the probability of NICU admission for marginal infants and that the benefits of admission appear to justify the costs.

Relative to the existing literature, my paper makes four contributions. First, I clarify the conditions under which empirical strategies of this kind identify the causal effect of congestion on care and outcomes, and I present detailed evidence on whether those conditions hold in my setting. In contrast, prior work has neglected to examine whether hospital staffing options are similar between high and low congestion days. Second, unlike nearly all prior work other than Bartel et al. (2011) and Freedman (2012), I am able to estimate the effect of congestion in a large population of hospitals, rather than a single hospital or a small group of hospitals, meaning that I can obtain more precise and potentially more representative estimates. Third, unlike all prior contributions except Freedman (2012), I have obtained linked vital records data that permit me to examine effects of congestion on post-discharge survival outcomes. Finally, my results extend the literature to a particular group of patients who have not previously been studied in detail.

The paper proceeds as follows. Section 1 describes my clinical setting. Section 2 introduces my empirical approach and derives my estimating equation. Section 3 describes my data, and Section 4 presents my results. Section 5 discusses and concludes.

3.1 Clinical background

I examine the effect of hospital congestion on patient outcomes in the context of acute manifestations of coronary artery disease (CAD). These patients can be broken into three basic groups. The first is those with acute myocardial infarction (AMI) or, colloquially, a heart attack. AMI occurs when an acute blockage in one of the coronary arteries (the arteries that supply the heart muscle) causes the death of downstream heart tissue. In this paper, I will sometimes examine two subgroups of AMI patients: patients with ST-elevation myocardial infraction (STEMI), so named for the particular electrocardiogram abnormality that defines this condition; and non-ST-elevation myocardial infarction (NSTEMI). In general, STEMI patients face a worse prognosis and benefit the most from aggressive early treatment (O'Gara et al., 2013). The second group is those with unstable angina, sudden and serious chest pain resulting from an acute blockage of a coronary artery, but which does not lead to death of the affected heart muscle. Because unstable angina and AMI share the common underlying cause of an acute blockage of a coronary artery, they are frequently grouped together under the heading of "acute coronary syndromes" (ACS). The final

group of patients I examine consists of those presenting at the hospital on an emergency basis with symptoms of coronary artery disease (typically angina), but who are determined not to be suffering from a current acute coronary blockage.

One dimension of resource utilization I examine in this paper is admission to a hospital's intensive care unit (ICU). An intensive care unit is a specialized group of beds in a hospital that serves the hospital's sickest patients. ICU beds typically offer more and more highly-trained nurses, more intensive monitoring technologies, and life-sustaining technologies like ventilators. Many large hospitals operate multiple ICUs that specialize in different types of patients, typically including one that specializes in patients with serious cardiac illnesses. Boundaries between specialized ICUs are porous, however, and patients may be placed in any one of a hospital's ICUs when capacity constraints bind. For this reason, I combine all of a hospital's ICUs together and refer to a single measure of "ICU census" in this paper.[1]

I also examine patient receipt of cardiac catheterization procedures. In cardiac catheterization, a catheter (small tube) is inserted through an incision in the patient's wrist or groin and then guided into the patient's coronary arteries. Once in place, the catheter can be used to inject contrast dye that is opaque to x-rays, permitting visualization of blood flow through the coronary arteries using a specialized x-ray camera. If blockages are identified, the physician can then use the catheter to perform a procedure called angioplasty, in which a small balloon is inflated to compress the blockage against the walls of the artery. A small wire mesh tube called a stent may also be inserted to keep the artery from closing up again over time. Angioplasty, stenting, and a small number of related procedures are collectively referred to as percutaneous coronary intervention (PCI). Catheterization procedures must be performed in a specialized procedure room known as a cardiac catheterization lab, which are typically in limited supply.

Contemporary treatment guidelines state that STEMI patients (without a contraindication) should undergo immediate catheterization, with the goal of performing PCI to clear the blocked coronary artery (O'Gara et al., 2013). Patients with NSTEMI or unstable angina may also may be appropriate candidates for early catheterization and PCI (Anderson et al., 2013). The benefits of catheterization and PCI are much smaller for patients without AMI or unstable angina, although

[1] As a practical matter, it is also not clear that hospitals consistently distinguish between different ICU types in the billing records I use to measure ICU utilization.

it may still relieve symptoms (Fihn et al., 2012).

3.2 Empirical model

This paper seeks to understand how changes in hospitals' congestion levels affect patient care and patient outcomes. To clarify the conditions under which the natural experiment I analyze will provide answers to this causal question, I consider the following simple structural model. Throughout, let i index patients, h index hospitals, and t index time. I assume that I observe the treatment or health outcome of interest Y_{iht}, the start-of-period hospital census N_{ht}, and a set of patient characteristics X_{iht} for each arriving patient, which in practice includes a full set of age decile-sex interactions, a set of indicators for comorbid conditions identified using the algorithm of Elixhauser et al. (1998), and indicators for four primary diagnosis groups (ST-elevation myocardial infarction, non-ST-elevation myocardial infarction, unstable angina, and other coronary artery disease). Patient census N_{ht} may be multi-dimensional and, thus, may capture patient census in multiple distinct parts of the hospital.

Outcomes for patient i arriving at hospital h at time t are determined by the production function

$$Y_{iht} = g(N_{ht}, S_{ht}, U_{iht}),$$

where S_{ht} is an unobserved vector describing the hospital's current operational capabilities, and U_{iht} is an unobserved vector of the arriving patient's characteristics. The hospital chooses staffing to maximize some (unspecified) objective function before seeing the characteristics of the new patient arrivals at time t. This yields a decision rule for S_{ht} of the form

$$S_{ht} = h(N_{ht}, \epsilon^S_{ht}),$$

where ϵ^S_{ht} is a vector of non-load factors (e.g. vacation schedules) that determine the hospital's staffing choice set. This form admits the possibility that the hospital can "staff up" on days where patient census is high.

To identify the causal effects of interest, I make the following conditional independence assumption:

$$N_{ht} \perp\!\!\!\perp (U_{iht}, \epsilon^S_{ht}) \mid \psi_t, \{\phi^\tau_{h\times\tau(t)}\}_{\tau\in S}, X_{iht}, \tag{ID}$$

where ψ_t is an unobserved date effect and $\phi^{\tau}_{h\times\tau(t)}$ is an unobserved hospital-time effect for each $\tau \in S = \{\text{year}, \text{month}, \text{day-of-week}\}$. Assumption ID states that, after conditioning on the date effect ψ_t and the hospital time effects $\{\phi^{\tau}_{h\times\tau(t)}\}_{\tau\in S}$, the remaining variation in patient census N_{ht} is as good as randomly assigned; in particular, it is independent of the characteristics of the newly-arriving patient (U_{iht}) and the non-load factors that affect the hospital's staffing choice set (ϵ^{S}_{ht}).

The first independence restriction rules out the possibility that patients presenting on more congested days differ systematically from those presenting on less congested days. The second independence restriction rules out the possibility that the hospital's staffing options differ between more and less congested days. Concretely, it rules out the possibility that the availability of physicians and nurses is different on days when patient census happens to be high. The date effects ψ_t allow each different date to have distinct patterns of congestion, patient characteristics, and hospital capabilities, and the hospital-time effects $\{\phi^{\tau}_{h\times\tau(t)}\}_{\tau\in S}$ allow each hospital to have its own pattern of evolution over time and its own seasonal and weekly patterns.

Under assumption ID, the regression function $\mathbb{E}[Y_{iht}|N_{ht}, X_{iht}, \psi_t, \{\phi^{\tau}_{h\times\tau(t)}\}_{\tau\in S}]$ satisfies

$$\mathbb{E}[Y_{iht}|N_{ht} = n, X_{iht}, \psi_t, \{\phi^{\tau}_{h\times\tau(t)}\}_{\tau\in S}] = \int g(n, h(n, \epsilon^{S}), u) dF(\epsilon^{S}, u|X_{iht}, \psi_t, \{\phi^{\tau}_{h\times\tau(t)}\}_{\tau\in S}). \quad (3.1)$$

The right-hand-side of equation (3.1) is the average outcome that would be obtained by assigning the hospital the patient census vector n, permitting the hospital to select its optimal staffing level given that census vector, and then averaging over the distribution of patient types (U_{iht}) and hospital staffing choice set shocks (ϵ^{S}_{ht}) that occur given observed patient characteristics X_{iht} and day characteristics ($\psi_t, \{\phi^{\tau}_{h\times\tau(t)}\}_{\tau\in S}$). Thus, under Assumption ID, the regression function on the left-hand-side answers the causal question of interest.

For the purposes of estimation, I assume that the integral on the right-hand-side of equation (3.1) can be written in a form such that

$$\mathbb{E}[Y_{iht}|N_{ht}, X_{iht}, \psi_t, \{\phi^{\tau}_{h\times\tau(t)}\}_{\tau\in S}] = \alpha + \tilde{N}^{\text{ICU}}_{ht}\beta_{\text{ICU}} + \tilde{N}^{\text{OIP}}_{ht}\beta_{\text{OIP}} + X_{iht}\gamma + \psi_t + \sum_{\tau\in S}\phi^{\tau}_{h\times\tau(t)}, \quad (3.2)$$

where $\tilde{N}^{\text{ICU}}_{ht}$ and $\tilde{N}^{\text{OIP}}_{ht}$ are, respectively, the hospital's "scaled" ICU and non-ICU inpatient ("other

inpatient") census. For hospital unit $k \in \{\text{ICU}, \text{OIP}\}$, this scaled census variable is defined as

$$\tilde{N}_{ht}^{k} = \frac{N_{ht}^{k} / \bar{N}_{h,\text{year}(t)}^{k}}{\text{IQR}(N_{ht}^{k} / \bar{N}_{h,\text{year}(t)}^{k})}$$

where N_{ht}^{k} is the total number of patients in unit k of hospital h at the start of date t, $\bar{N}_{h,\text{year}(t)}^{k}$ is the mean number of patients in unit k of hospital h during the year in which t falls, and the denominator of this equation is the full-sample interquartile range of their ratio. Concretely, this measure is current unit k census as a share of its hospital-year mean, scaled such that a one unit change corresponds to moving from the 25th percentile to the 75th percentile in the full-sample distribution of shares.

The functional form in equation (3.2) is admittedly somewhat restrictive. In particular, this specification assumes that the effect of congestion on the outcomes of interest is constant across hospitals and over time. However, Chernozhukov et al. (2010) show that, in settings with a single fixed effect and a single binary treatment variable, fixed effects specifications of this form will still estimate a particular weighted average treatment effect even when the true effects are heterogeneous. While it seems plausible that such results would carry over in some form to this setting, confirming that is beyond the scope of this paper.

I estimate equation (3.2) using the ordinary least squares fixed effects estimator, and I cluster my standard errors at the hospital-date level, the level at which the congestion measures vary. Estimation poses two complications, one theoretical and one computational. The theoretical complication is that fixed effects estimation typically requires a stronger assumption than the contemporaneous exogeneity assumption made in Assumption ID.[2] The standard stronger assumption is strict exogeneity, in which the current error term is independent of the regressors for *all* observations associated with each fixed effect. In practice, however, Wooldridge (2010) demonstrates that, provided that contemporaneous exogeneity holds, the asymptotic bias associated with failures of strict exogeneity is inversely proportional to the number of observations identifying each fixed effect. In my application, this number will typically be quite large and the bias correspondingly small. I conclude, therefore, that this is not a substantial concern.

The computational complication is that equation (3.2) contains four high-dimensional fixed

[2]I thank Seth Freedman for bringing this point to my attention.

effects. Direct computation of the OLS estimates via a least squares dummy variables approach would therefore require solving a linear system of very high dimension, which is computationally infeasible. With a single set of fixed effects, the standard way of addressing this problem is to use a de-meaning transformation to compute the portion of the variation in the outcomes and regressors that is orthogonal to the fixed effects. The Frisch-Waugh-Lovell Theorem then implies that the desired estimates can be obtained from a regression of the de-meaned outcome on the de-meaned regressors. This method, however, is not directly applicable to models with multiple fixed effects.

To solve this problem, I turn to an algorithm proposed by Gaure (2013). In brief, the algorithm consists of iteratively de-meaning the outcome and regressors with respect to each set of fixed effects in sequence until convergence. It can be shown that this algorithm computes the portion of the variation in the outcome and regressors that is orthogonal to the space spanned by all four sets of fixed effects. As above, the Frisch-Waugh-Lovell Theorem then implies that the estimates of interest can be obtained from a regression of the transformed outcome on the transformed regressors. Additional details on this algorithm and a comparison to other approaches to estimation in the presence of multiple high-dimensional fixed effects (Abowd et al., 1999, 2002; Carneiro et al., 2012) are provided in Appendix C.1.

3.3 Data

This analysis draws on data from two sources: (1) records of all emergency department, outpatient observation, and inpatient hospital stays in Massachusetts during fiscal years 2003-2009; and (2) the universe of death certificates filed in Massachusetts during the period 2002-2009. In this section, I first describe the discharge data. I then describe how I construct the analysis sample from these data. Finally, I describe how I link the discharge data to the death certificates.

3.3.1 Hospital records

I obtained data on the universe of emergency department, outpatient observation, and inpatient hospital stays in Massachusetts during fiscal years 2003-2009 from the Massachusetts Center for Health Information and Analysis (formerly the Division of Health Care Finance and Policy).[3]

[3]Fiscal years are defined such that year Y starts in October of year $Y-1$.

Each discharge record includes the full set of fields typically present on hospital discharge records, including ICD-9-CM diagnosis and procedures codes, basic patient characteristics (like age, race, and insurance status), hospital identifiers,[4] a code for the type of discharge (e.g. to another hospital, to home, or due to death), and a code for the source of the patient's admission (e.g. home or another hospital). The emergency department and outpatient observation files also sometimes report Current Procedural Terminology (CPT) procedure codes in place of or in addition to ICD-9-CM procedure codes.

The records contain three other types of information that are essential to my analysis but are not available in all discharge databases. First, each inpatient discharge record reports detailed billing information for each stay, including UB-92/UB-04 "revenue codes," charges, and the number of units for each billed service. Crucially for my purposes, these codes permit computing the duration of any ICU admission during the hospital stay, as well as whether the patient received emergency department services prior to admission; the specific codes used are described in Appendix C.2. Second, all three data sources report the exact dates of patient arrival and discharge, which permits me to compute hospital census at a daily frequency and to link each arriving patient to the appropriate measure of congestion. Finally, these data report a unique patient identifier that is suitable for following patients across encounters (both within and between hospitals). This unique identifier is an encrypted version of the patient's Social Security Number and thus is also suitable for linkage to death certificates as discussed later in this section.

From these data, I identify the diagnoses and procedures of primary interest using the reported ICD-9-CM and CPT codes. The specific ICD-9-CM and CPT codes used are reported in Appendix C.2. As noted in the last section, I identify indicators of patient comorbid conditions for use as control variables using the algorithm proposed by Elixhauser et al. (1998).[5] I identify a small number of additional diagnosis and procedure categories for use in covariate balance tests using the Clinical Classification Software (CCS) groupings of ICD-9-CM codes defined by the Healthcare Cost and Utilization Project. Details on the use of CCS codes are reported in Appendix C.2.

[4] A small number of hospitals report distinct campuses separately for part of the period and jointly for part of the period. In the interest of consistency over time, I consolidate such hospitals into a single unit for the full period.

[5] For this purpose, I adapt software posted by the Healthcare Cost and Utilization Project at http://www.hcup-us.ahrq.gov/toolssoftware/comorbidity/comorbidity.jsp.

As described in Section 3.2, the key right-hand-side variables in my main specification are the hospital's ICU and non-ICU inpatient censuses at the beginning of the day the patient arrived. To compute ICU census at each point in time, I start with the universe of inpatient records for each hospital, and I identify all stays for which the billing records indicate an ICU stay using the procedure described in Appendix C.2. These billing records also report the length of the ICU stay measured in days, but, unfortunately, do not report when during the patient's hospital stay the ICU stay occurred.[6] I thus assume that all of a patient's ICU days fall at the beginning of the patient's hospital stay. Although this assumption is not ideal, there is no alternative given the available data, and it should be accurate in the sizable majority of cases. As for census in the rest of the hospital, the hospital's overall inpatient census at the beginning of each day can be computed unambiguously from the admission and discharge dates on each discharge record, so I derive non-ICU patient census by subtraction.

3.3.2 Analysis sample construction

I estimate equation (3.2) in a sample of newly arriving patients with an acute coronary illness. Specifically, I identify patients with a diagnosis of acute myocardial infarction, unstable angina, or another diagnosis indicating coronary artery disease. Some of the diagnosis codes used to identify patients in the latter two groups can also be used for patients being admitted to the hospital for scheduled procedures (generally PCI or bypass surgery). I thus exclude patient records that report these codes and show no evidence of interaction with the hospital's emergency department. Details of the ICD-9-CM codes used to define each diagnosis group and the method for identifying admissions through the emergency department are reported in Appendix C.2. Starting with this basic group of records, I limit the sample in the following ways.

First, I wish to limit the sample to patients for whom hospital arrival represents the beginning of that treatment "episode"; I do not wish to include transfers from other hospitals. I identify a

[6] There is one complication in interpreting the length of stay information reported by the billing records. My calculations require knowing the number of midnights the patient was present in the ICU, but Medicare requires hospitals to measure ICU length of stay as the arrival day plus the number of subsequent days that the patient is present at midnight (CMS, 2013), and private insurers appear to follow the Medicare definition. The number of midnights a patient was present in the ICU is thus ambiguous whenever the reported ICU length of stay is equal to one. I resolve this ambiguity by assuming that the patient stayed overnight in the ICU unless the patient's date of hospital admission and discharge are the same.

patient as having been transferred into the current hospital if either: (1) the current hospital codes the patient as having arrived via transfer in the admission source field; or (2) another hospital coded the patient as being discharged due to transfer on the day of arrival at the current hospital.

Second, in order to focus on hospitals that routinely treat patients with acute coronary illness, I restrict the sample to those hospitals that see at least two patient arrivals each day by patients with the diagnoses listed above or a diagnosis of non-cardiac chest pain. This restriction eliminates 16 hospitals from the dataset, approximately half of which are specialty hospitals of various types and the remainder of which are very small hospitals, many of which do not provide full emergency services. This restriction reduces the sample size by less than 5 percent.

Third, because stay records are reported to the state on the basis of the date of discharge, the database will fail to include many arrivals occurring near the end of fiscal year 2009. I thus limit the sample to arrivals occurring at least 90 days before October 1, 2009. Similarly, I only examine stays occurring after the start of fiscal year 2003 since any reported admissions from before that date will be a selected sample of longer stays. There are also a small number of episodes where a hospital fails to report some or all of its data. I exclude the hospital from the sample for the period of the reporting failure and for a 90 day window preceding the failure when the failure affects the hospital's inpatient records, but only a 30 day window when the failure affects only the hospital's emergency department or outpatient observation records. Finally, I exclude two patients whose age is missing.

Table 3.1 reports basic descriptive statistics for the final analysis sample. Slightly more than half of the sample is made up of patients who are experiencing an AMI, and most of the rest are experiencing unstable angina. There are evident day-of-week and seasonal patterns in arrival, which suggests that the types of patients arriving may vary at these temporal frequencies, justifying the inclusion of the corresponding hospital-time fixed effects in equation (3.2).

3.3.3 Linkage to vital records

Post-discharge patient survival is the main health outcome of interest in my analyses. I obtain data on these outcomes by linking the hospital records to Massachusetts death certificates for 2002-2009. I obtained these records via a request to the Registry of Vital Statistics in the Massachusetts Department of Public Health.

Table 3.1: Descriptive statistics for the analysis sample

	Mean	SD
A. Patient characteristics		
Age	69.182	14.485
Female	0.418	0.493
COPD	0.124	0.330
Diabetes	0.310	0.463
Hypertension	0.607	0.488
B. Primary diagnosis		
ST-elevation myocardial infarction	0.226	0.418
Non-ST elevation myocardial infarction	0.328	0.470
Unstable angina	0.256	0.436
Other coronary artery disease	0.190	0.392
C. Day of arrival		
Sunday	0.125	0.330
Monday	0.160	0.367
Tuesday	0.150	0.357
Wednesday	0.150	0.357
Thursday	0.148	0.355
Friday	0.143	0.350
Saturday	0.124	0.330
D. Quarter of arrival		
Quarter 1	0.258	0.438
Quarter 2	0.256	0.436
Quarter 3	0.218	0.413
Quarter 4	0.267	0.442
E. Encounter type		
Emergency department	0.129	0.336
Observation admission	0.062	0.240
Inpatient admission	0.809	0.393
F. Treatment characteristics		
Admitted to ICU	0.295	0.456
Catheterization	0.360	0.480
PCI	0.185	0.388
Length of stay	3.529	4.442
Transferred before admission	0.084	0.277
Transferred ever	0.256	0.437
N	149,428	

Note: This table reports descriptive statistics for the analysis sample of new hospital arrivals with acute manifestations of coronary artery disease described in Section 3.3.

Table 3.2: Effect of hospital congestion on the probability a valid SSN is recorded

		Coef./SE on Independent Variables	
Dependent variable	**Mean**	Scaled ICU census	Scaled OIP census
Missing patient SSN	0.0296	-0.0006	0.0005
		(0.0004)	(0.0006)
N	149,428	149,428	149,428

Note: This table reports the results of estimating equation (3.2) where the outcome is an indicator for whether the patient SSN is missing. The key independent variables are ICU and non-ICU census at the start of the day the patient arrived, scaled so that a one unit change can be interpreted as the effect of moving from the 25th percentile to the 75th percentile in the distribution of proportional deviations from average census. The regression also includes a full set of age decile-sex interactions, a full set of Elixhauser comorbidities, indicators for primary diagnosis as described in the text, and date, hospital x year, hospital x month, and hospital x day-of-week fixed effects. The estimation sample is the sample of new hospital arrivals with coronary artery disease described in Section 3.3. Standard errors are clustered at the hospital-date level. Statistical significance is denoted as follows: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

The linkage of the two data sources was achieved using patient Social Security Numbers (SSNs), which are present on both data sources. As noted above, the SSNs reported on the hospital discharge data were encrypted by the Center for Health Information and Analysis before they were transmitted to me in order to protect patient privacy. To achieve the linkage, therefore, the Department of Public Health transmitted the the death certificate file directly to a contractor of the Center for Health Information and Analysis. That contractor encrypted the SSNs, deleted the raw SSNs from the file, and then transmitted the file with the encrypted identifiers to me for linking.

More than 97 percent of the hospital discharge records in the analysis sample report a valid (encrypted) SSN and, thus, are eligible for linking. A possible concern is that whether a patient's record is included in the 3 percent of records without a valid SSN is a function of hospital congestion, which could bias my estimates of the effect of congestion on survival. To address this concern, I fit a version of equation (3.2) that uses an indicator for whether the patient's SSN was missing as the outcome. Table 3.2 reports the results; there is no evidence that whether the hospital records an SSN is an outcome of congestion, meaning that this is unlikely to be a source of bias.

Match quality is high. To evaluate the true positive rate of the match, I follow Zingmond et al. (2004) and estimate the share of discharge records on which the patient is reported to have died in the hospital that match a death certificate with a date of death equal to the date of discharge; I estimate that this share is equal to 93 percent. To evaluate the false match rate, I estimate the

share of matches in which patient sex as listed on the death certificate and patient sex as listed on the discharge record disagree; this share is just 0.3 percent.

3.4 Results

3.4.1 Examining the variation in congestion

As discussed in Section 3.2, I aim to identify the effect of hospital congestion on patient care and outcomes using high-frequency variation in hospital congestion levels. The key identifying assumption (Assumption ID) states that, after conditioning on date fixed effects, hospital-by-year, hospital-by-month, and hospital-by-day-of-week fixed effects, the remaining variation in hospital congestion is as good as randomly assigned. In this subsection, I present evidence supporting this identifying assumption.

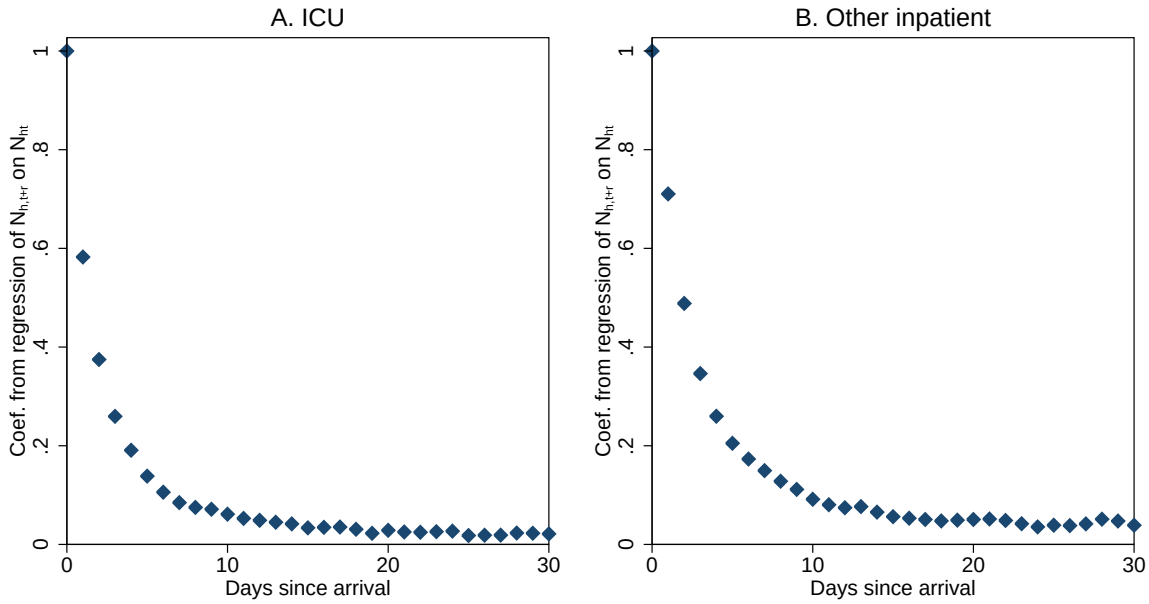
Substantial variation does remain in the hospital congestion measures included in equation (3.2) after accounting for the fixed effects. The R^2 from a regression of scaled ICU census on the fixed effects in equation (3.2) is just 0.14, while the corresponding figure for scaled non-ICU census is 0.36. An important question, however, is whether this residual variation is "high-frequency" variation in patient load that arises from random variation in patient arrivals or "low-frequency" variation that arises from persistent shifts in hospital capacity or patient demand. Variation of the latter type is unlikely to satisfy the identifying assumption and, as such, is undesirable. While the hospital-time effects included in equation (3.2) should purge most low-frequency variation, if they are not sufficiently rich, some could remain.

To address this question, I examine the persistence of variation in congestion via regressions of the form

$$N^k_{h,t+s} = \alpha + N^{\text{ICU}}_{ht}\beta^{\text{ICU}} + N^{\text{OIP}}_{ht}\beta^{\text{OIP}} + \psi_t + \sum_{\tau \in S} \phi^{\tau}_{h\times\tau(t)}, \tag{3.3}$$

where N^k_{ht} is the unscaled census of hospital h at time t in unit $k \in \{\text{ICU}, \text{OIP}\}$ and $s \geq 0$. For any given value of s and census type k on the left-hand-side, the coefficient β^k can be interpreted as the share of the deviation from normal in hospital unit k at time t that persists s days later. Figure 3.2 plots the resulting coefficients for each type of hospital census. The evidence is clear that congestion shocks dissipate quickly and essentially completely, indicating that the residual

Figure 3.2: Persistence of variation in hospital congestion



Notes: Panel A of this figure reports the coefficient β^{ICU} from estimating equation (3.3) with $N^{\text{ICU}}_{h,t+s}$ as the outcome for each $s \in \{0, 1, \ldots, 30\}$. Panel B reports the corresponding sequence of estimated coefficients for non-ICU census. In addition to the variables described in the last sentence, equation (3.3) includes the other measure of hospital census and date, hospital x year, hospital x month, and hospital x day-of-week fixed effects. The estimation sample is the sample of new hospital arrivals with coronary artery disease described in Section 3.3.

variation in congestion is almost entirely "high-frequency" variation as desired.[7]

I turn now to direct "covariate balance" tests of the identifying assumption. This assumption has two parts: that the (unobserved) characteristics of arriving patients are independent of residual variation in congestion; and that hospital staffing options are independent of residual variation in congestion. To assess the plausibility the first assumption, I estimate versions of equation (3.2) in which observed patient characteristics are the outcomes (and all patient controls are omitted from the right-hand-side). If congestion appears independent of a rich set of observed characteristics of arriving patients, then it is reasonable to believe that congestion is independent of unobserved patient characteristics as well. Table 3.3 presents the results of these analyses. Panel A shows that the basic demographics and clinical characteristics of arriving patients are not associated with

[7] A small portion of the shock to non-ICU census does persist 30 days later. However, I have verified in a simple model that even under the assumption that arrivals are completely independent from day to day, persistence of this magnitude is to be expected, essentially because the length-of-stay distribution has a very long right tail.

congestion, and panel B shows that summary indices obtained by using the full set of patient covariates included in equation (3.2) to predict various treatment and health outcomes of interest are also unrelated to congestion.[8] Panel C shows that patients arriving on high and low congestion days do not differ in their utilization of hospital care over the last year. Panel D presents evidence on whether hospitals are turning patients away when patient census is high, which could be problematic if certain types of patients are more likely to be rejected than others; I find no evidence of patient diversion. Given the precision of the reported estimates, I conclude that the results permit reasonable confidence that patients arriving on days on which hospital census is high are similar to those arriving on other days.

I turn next to evaluating whether the hospital's staffing choice set is similar on congested days relative to less-congested days. This assumption is difficult to test, as it does *not* rule out hospitals altering staffing in response to patient census, which means that observed indicators of hospital staffing may differ across high and low census days even if the assumption holds. Rather, it is inherently about the hospital's *menu* of staffing options remains constant when congestion changes.

With this important caveat in mind, I examine the effect of congestion on two proxies for hospital staffing options. First, I examine whether the hospital's slate of scheduled procedures appears to differ between high and low-census days, which provides an indicator of its general staffing configuration and operational state. Concretely, I compute for each day the number of patients admitted that day who undergo several common procedures that typically occur on a scheduled basis, and I divide that number by the typical number of such arrivals during that hospital-year; to avoid division by zero, I exclude hospital-years in which a hospital did none of the listed procedure. I then run a version of equation (3.2) with these volume measures on the left-hand-side (omitting patient covariates). Table 3.4 reports the results.[9] I find that when scaled ICU census increases by one unit, hospitals' hysterectomy volume appears to be approximately 3 percent higher, while when scaled non-ICU census increase by one unit, orthopedic volume and cholecystectomy volume

[8]The creation of summary indices of this form follows Baicker et al. (2006) and Chandra and Staiger (2007).

[9]To interpret these results, recall that, as described in Section 3.2, the scaled census variables are defined to be the relevant census measure as a share of mean census in that hospital-year, scaled by the interquartile range of this share in the full dataset. A one unit change in this variable is, therefore, roughly analogous to a change in this share from the 25th to the 75th percentile.

Table 3.3: Balance on patient characteristics

		Coef./SE on Independent Variables	
Dependent variable	**Mean**	Scaled ICU census	Scaled OIP census
A. Demographics and comorbidities ($N = 149,428$)			
Age	69.1821	-0.0068	-0.0209
		(0.0414)	(0.0549)
Female	0.4182	-0.0026	0.0017
		(0.0014)	(0.0019)
COPD	0.1244	-0.0011	-0.0023
		(0.0009)	(0.0013)
Diabetes	0.3102	0.0012	-0.0026
		(0.0013)	(0.0017)
Hypertension	0.6071	-0.0007	0.0009
		(0.0014)	(0.0019)
B. Summary predictive indices ($N = 149,428$)			
ICU admission	0.2900	0.0000	-0.0002
		(0.0004)	(0.0006)
Catheterization	0.3542	0.0003	0.0001
		(0.0004)	(0.0005)
One-year mortality	0.1626	-0.0004	-0.0002
		(0.0004)	(0.0006)
C. Prior-year utilization ($N = 119,311$)			
Any inpatient stay	0.3904	0.0009	-0.0004
		(0.0015)	(0.0021)
Number of inpatient stays	0.9508	0.0044	-0.0081
		(0.0059)	(0.0078)
Any hospital encounter	0.5503	0.0017	-0.0010
		(0.0016)	(0.0021)
Number of hospital encounters	1.8710	0.0231	-0.0111
		(0.0126)	(0.0169)
D. Arrival cohort size ($N = 149,428$)			
Arrival cohort size	2.5153	-0.0066	-0.0081
		(0.0054)	(0.0100)

Note: This table reports the results of estimating equation (3.2) with the listed outcome, omitting covariates. The comorbidity indicators are described in Appendix C.2. The summary predictive indices are predicted values from a logit regression of the listed outcome on the patient covariates included in equation (3.2). The independent variables are ICU and non-ICU census at the start of the day the patient arrived, scaled so that a one unit change can be interpreted as the effect of moving from the 25th percentile to the 75th percentile in the distribution of proportional deviations from average census. The regression also includes date, hospital x year, hospital x month, and hospital x day-of-week fixed effects. The estimation sample is the sample of new hospital arrivals with coronary artery disease described in Section 3.3. Standard errors are clustered at the hospital-date level. Statistical significance is denoted as follows: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 3.4: Balance on scheduled procedure volume

	Coef./SE on Independent Variables		
Dependent variable	Scaled ICU census	Scaled OIP census	**N**
Hip and knee procedures	-0.0032	-0.0250**	149,428
	(0.0059)	(0.0077)	
Spine surgery	-0.0482	0.0012	135,991
	(0.0274)	(0.0221)	
Hysterectomy	0.0298*	-0.0122	148,438
	(0.0127)	(0.0204)	
Cholecystectomy	-0.0027	-0.0364**	149,428
	(0.0086)	(0.0120)	

Note: This table reports the results of estimating equation (3.2) with the listed outcome, omitting covariates. The dependent variables are the total volume of the listed procedure performed that day on newly-arriving patients, scaled by the hospital-year mean volume. Hospitals with no volume of the listed procedure during the hospital-year are excluded. The key independent variables are ICU and non-ICU census at the start of the day the patient arrived, scaled so that a one unit change can be interpreted as the effect of moving from the 25th percentile to the 75th percentile in the distribution of proportional deviations from average census. The regression also includes date, hospital x year, hospital x month, and hospital x day-of-week fixed effects. The estimation sample is the sample of new hospital arrivals with coronary artery disease described in Section 3.3. Standard errors are clustered at the hospital-date level. Statistical significance is denoted as follows: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

appear to be lower to a similar extent. The other coefficients are statistically insignificant. For the reasons discussed above, it is not clear whether even the statistically significant coefficients necessarily imply violations of the identifying assumption. Given that the effects of census on scheduled surgery volume are small, however, I conclude that any violations of the identifying assumption are small and unlikely to meaningfully affect the conclusions of this analysis.

Table 3.5 reports similar balance tests where the left-hand-side variables are measures of the hospital's current physician complement. Panel A reports results where the left-hand-side variables are, respectively, the number of distinct physician identifiers listed on emergency department encounters starting that day, the number of distinct physician identifiers listed as attending physicians for inpatient admissions starting that day, and the number of distinct physician identifiers listed as operating physicians for inpatient admissions starting that day; I scale each by the average number of each type of physician identifier appearing each day. Panel B reports results where the left-hand-side variables are the total number of times that each physician identifier that is present that day

appears over the course of the year. The results indicate that the total number of distinct physicians appearing on the inpatient records of arriving patients is very slightly smaller when census is high than when it is low, and the physicians present seem to have a slightly different experience profile. Of course, the results pose the same interpretation problems as those presented in Table 3.4; they could be an *outcome* of high census if hospitals cancel various scheduled admissions when beds are tight. Suggestive evidence for this theory is that emergency department staffing, which should not be subject to such pressures, does not appear to vary with inpatient census. In any case, I take comfort in the fact that the estimated effects are small and conclude, as above, that these results suggest that any failure of the identifying assumption that does exist is small.

3.4.2 Effects on patient treatment

In this subsection, I examine how hospital congestion affects patient treatment. Before proceeding to the main results, however, I note that it will frequently be desirable to examine whether patients with illnesses of differing severity are affected differently by congestion. The most natural way to examine such questions is to estimate specifications separately by the diagnosis patients receive. Such an approach is only valid, however, if diagnosis is not itself an *outcome* of congestion. To rule out this possibility, I estimate a version of equation (3.2) in which the indicators for the patients' diagnosis are placed on the left-hand-side as outcomes rather than included on the right-hand-side as controls. Table 3.6 reports the results. There is no evidence that patients' final diagnosis is itself a function of congestion, so I conclude that subgroup analyses that stratify based on patient diagnosis will give valid and interpretable results.[10]

I first examine how patients' initial routing depends on hospital census. As I will throughout the results, I separately examine effects on patient care at hospitals that can provide a full range of care for patients with acute coronary illness – particularly 24-hour catheterization services – and those that cannot. I identify full-service hospitals empirically as those that transfer less than 10 percent of patients who arrive at that hospital with acute myocardial infarction or unstable angina. In practice, nearly all hospitals have transfer shares close to zero or close to 50 percent, so the classification of hospitals into the two categories is quite insensitive to the precise threshold used.

[10]Similarly, these results indicate that including indicators for patient diagnosis as control variables in equation (3.2) is acceptable.

Table 3.5: Balance on hospital staffing

		Coef./SE on Independent Variables	
Dependent variable	**Mean**	Scaled ICU census	Scaled OIP census
A. Scaled number of physicians present			
Emergency department	1.0039	-0.0013	0.0017
		(0.0007)	(0.0009)
Inpatient attending	1.0342	-0.0029**	-0.0057**
		(0.0009)	(0.0018)
Inpatient operating	1.0384	-0.0044***	-0.0080***
		(0.0013)	(0.0022)
B. Average annual volume of physicians present			
Emergency department	2663.2	-1.8	-3.2
		(1.8)	(2.5)
Inpatient attending	237.8	-0.1	-1.2***
		(0.2)	(0.3)
Inpatient operating	145.7	0.2	0.4*
		(0.1)	(0.2)
N	149,428	149,428	149,428

Note: This table reports the results of estimating equation (3.2) with the listed outcome, omitting covariates. In panel A, the dependent variables are, respectively, the total number of distinct physicians listed on emergency department encounters starting that day, listed as attending physicians on inpatient stays staring that day, or listed as operating physicians on inpatient stays starting that day, each scaled by the relevant hospital-year mean. In panel B, the dependent variables are, respectively, the average annual number of emergency department records on which each present emergency department physician appears, the average number of inpatient records on which each present attending physician appears, and the average number of inpatient records on which each present operating physician appears. The key independent variables are ICU and non-ICU census at the start of the day the patient arrived, scaled so that a one unit change can be interpreted as the effect of moving from the 25th percentile to the 75th percentile in the distribution of proportional deviations from average census. The regression also includes date, hospital x year, hospital x month, and hospital x day-of-week fixed effects. The estimation sample is the sample of new hospital arrivals with coronary artery disease described in Section 3.3. Standard errors are clustered at the hospital-date level. Statistical significance is denoted as follows: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 3.6: Effect of hospital congestion on patient diagnosis

		Coef./SE on Independent Variables	
Dependent variable	**Mean**	Scaled ICU census	Scaled OIP census
ST-elevation myocardial infarction	0.2260	0.0006	0.0009
		(0.0012)	(0.0016)
Non-ST elevation myocardial infarction	0.3282	0.0001	-0.0007
		(0.0012)	(0.0017)
Unstable angina	0.2560	-0.0004	-0.0013
		(0.0012)	(0.0017)
Other coronary artery disease	0.1898	-0.0002	0.0012
		(0.0011)	(0.0015)
N	149,428	149,428	149,428

Note: This table reports the results of estimating equation (3.2) where the outcome is an indicator for whether the patient SSN is missing. The key independent variables are ICU and non-ICU census at the start of the day the patient arrived, scaled so that a one unit change can be interpreted as the effect of moving from the 25th percentile to the 75th percentile in the distribution of proportional deviations from average census. The regression also includes a full set of age decile-sex interactions, a full set of Elixhauser comorbidities, and date, hospital x year, hospital x month, and hospital x day-of-week fixed effects. The estimation sample is the sample of new hospital arrivals with coronary artery disease described in Section 3.3. Standard errors are clustered at the hospital-date level. Statistical significance is denoted as follows: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 3.7 reports the results;[11] it appears that patients are slightly less likely to be admitted to the hospital when ICU census is high, but this effect is quite small. The point estimate for non-ICU census is similar, but insignificant. Precisely where the non-admitted patients are sent instead is unclear. All in all, however, these effects are very small, and unlikely to have meaningful consequences for patient care.

I next examine where patients are cared for once admitted to the hospital. Unsurprisingly, when scaled ICU census increases by one unit, patients are approximately 1 percentage point less likely to be admitted to the hospital's intensive care unit, and the total number of ICU days is commensurately lower. Subject to the limits of precision, this estimate appears similar across low and high-transfer hospitals. By contrast, ICU census appears to have little effect on either how long the patient ultimately remains in the hospital or whether the patient is transferred to another hospital during the stay. Census in the rest of the hospital seems to have little effect on care, except that it seems to very modestly lengthen stays in the hospital's ICU, likely a reflection of the fact that when beds are at a premium in the rest of the hospital, the hospital's best option is sometimes to hold patients in the ICU longer.

In light of the modest, but meaningful effects on the probability of ICU admission found above, it is interesting to examine which patients are being rationed out of the ICU. Table 3.9 reports results for the same set of outcomes, broken down by three patient groups: patients with ST-elevation myocardial infarction (STEMI), non-STEMI acute coronary syndrome patients, and patients experiencing manifestations of coronary artery disease that do not qualify as an acute coronary syndrome. Judged as a share of the baseline probability of ICU admission for each group, the reduction in the probability of ICU admission rises as patients become less sick, and the difference between the STEMI patients and the non-ACS patients is nearly statistically significant. This pattern of results is at least qualitatively consistent with a model in which hospitals are rationing ICU beds in accordance with illness severity (and, presumably, expected clinical benefit).

[11] One feature of Table 3.7 merits additional comment. Comparing the estimates for the full sample to the estimates in the two hospital sub-samples demonstrates that the full sample estimates are not simply sample-size-weighted averages of the coefficients in the sub-samples; the full sample estimate is much closer to the estimate for the high-transfer hospitals. The main reason for this is that the average effect estimated by a pooled regression is weighted by both the relative sample sizes and the amount of residual variance in the treatment variable in the two samples (see, for example, Angrist and Pischke (2009)). Because the low-transfer hospitals are typically larger, the residual variance in the scaled census is smaller for these hospitals, and so the estimated effect for the low-transfer hospitals receives considerably lower weight in the pooled estimate.

Table 3.7: Effect of congestion on initial patient routing by hospital type

Column:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Hospital sample:	All hospitals			Low transfer hospitals			High transfer hospitals		
		Independent variable			Independent variable			Independent variable	
Dependent variable	Mean	ICU cens.	OIP cens.	Mean	ICU cens.	OIP cens.	Mean	ICU cens.	OIP cens.
Admitted	0.8092	-0.0029**	-0.0020	0.9131	-0.0014	-0.0010	0.7464	-0.0030*	-0.0022
		(0.0011)	(0.0014)		(0.0021)	(0.0024)		(0.0012)	(0.0016)
Transferred before admission	0.0837	0.0014	0.0009	0.0033	0.0002	0.0004	0.1322	0.0014	0.0009
		(0.0009)	(0.0011)		(0.0005)	(0.0007)		(0.0010)	(0.0014)
Died before admission	0.0115	0.0000	-0.0005	0.0086	-0.0006	-0.0008	0.0133	0.0000	-0.0005
		(0.0003)	(0.0004)		(0.0008)	(0.0008)		(0.0003)	(0.0005)
Sent home without admission	0.0957	0.0015	0.0016	0.0751	0.0019	0.0014	0.1081	0.0015	0.0018
		(0.0008)	(0.0011)		(0.0019)	(0.0022)		(0.0009)	(0.0013)
N	149,428	149,428	149,428	56,265	56,265	56,265	93,163	93,163	93,163

Note: This table reports the results of estimating equation (3.2) with the listed outcome in the listed hospital sample. Low transfer hospitals are those that transfer (at any point after arrival) less than 10 percent of their acute coronary syndrome patients. The key independent variables are ICU and non-ICU census at the start of the day the patient arrived, scaled so that a one unit change can be interpreted as the effect of moving from the 25th percentile to the 75th percentile in the distribution of proportional deviations from average census. The regression also includes a full set of age decile-sex interactions, a full set of Elixhauser comorbidities, indicators for primary diagnosis as described in the text, and date, hospital x year, hospital x month, and hospital x day-of-week fixed effects. The underlying sample is the sample of new hospital arrivals with coronary artery disease described in Section 3.3. Standard errors are clustered at the hospital-date level. Statistical significance is denoted as follows: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 3.8: Effect of congestion on stay characteristics by hospital type

Column:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Hospital sample:	All hospitals			Low-transfer hospitals			High-transfer hospitals		
		Independent variable			Independent variable			Independent variable	
Dependent variable	Mean	ICU cens.	OIP cens.	Mean	ICU cens.	OIP cens.	Mean	ICU cens.	OIP cens.
Admitted to ICU	0.2951	-0.0091***	0.0019	0.3170	-0.0064	0.0006	0.2819	-0.0091***	0.0027
		(0.0011)	(0.0016)		(0.0033)	(0.0040)		(0.0012)	(0.0017)
Days in ICU	0.8511	-0.0307***	0.0167*	1.0461	-0.0529*	0.0520	0.7333	-0.0253***	0.0141*
		(0.0048)	(0.0076)		(0.0229)	(0.0291)		(0.0043)	(0.0067)
Length of stay	3.5288	0.0010	0.0222	4.7236	0.0474	0.1093*	2.8073	0.0003	0.0079
		(0.0089)	(0.0126)		(0.0389)	(0.0463)		(0.0083)	(0.0115)
Transferred	0.2562	-0.0008	-0.0007	0.0207	-0.0008	-0.0015	0.3984	-0.0008	-0.0005
		(0.0012)	(0.0016)		(0.0012)	(0.0015)		(0.0014)	(0.0019)
N	149,428	149,428	149,428	56,265	56,265	56,265	93,163	93,163	93,163

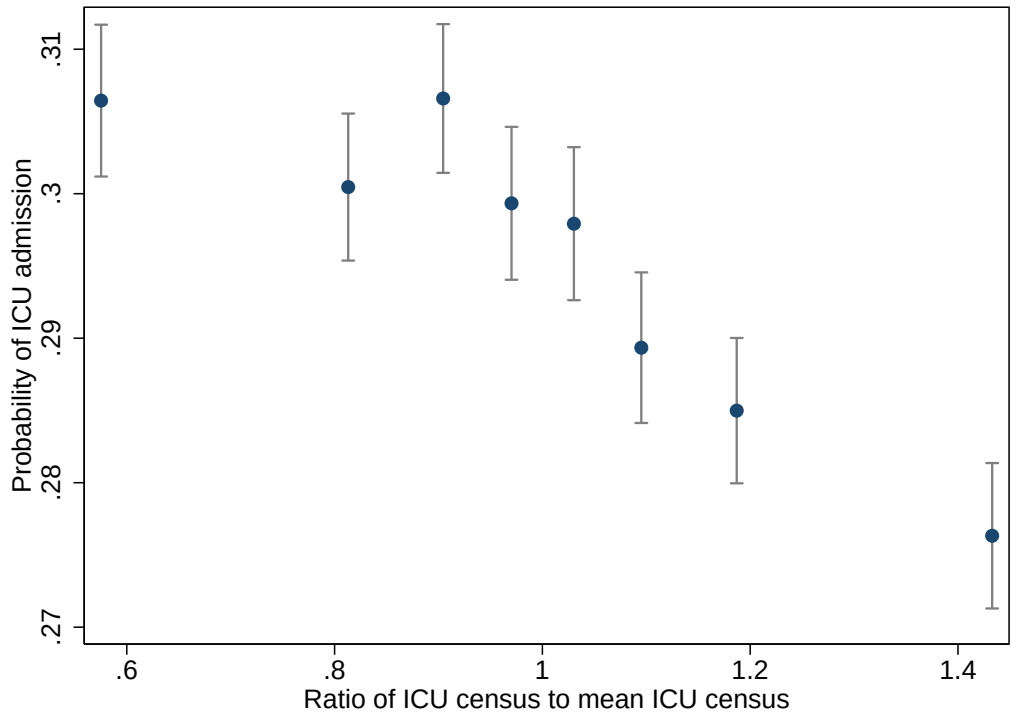
Note: This table reports the results of estimating equation (3.2) with the listed outcome in the listed hospital sample. Low transfer hospitals are those that transfer (at any point after arrival) less than 10 percent of their acute coronary syndrome patients. The key independent variables are ICU and non-ICU census at the start of the day the patient arrived, scaled so that a one unit change can be interpreted as the effect of moving from the 25th percentile to the 75th percentile in the distribution of proportional deviations from average census. The regression also includes a full set of age decile-sex interactions, a full set of Elixhauser comorbidities, indicators for primary diagnosis as described in the text, and date, hospital x year, hospital x month, and hospital x day-of-week fixed effects. The underlying sample is the sample of new hospital arrivals with coronary artery disease described in Section 3.3. Standard errors are clustered at the hospital-date level. Statistical significance is denoted as follows: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 3.9: Effect of congestion on stay characteristics by patient type

Column:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Patient diagnosis:	STEMI			Non-STEMI ACS			Other coronary artery disease		
		Independent variable			Independent variable			Independent variable	
Dependent variable	Mean	ICU cens.	OIP cens.	Mean	ICU cens.	OIP cens.	Mean	ICU cens.	OIP cens.
Admitted to ICU	0.4539	-0.0098***	0.0017	0.2986	-0.0095***	0.0031	0.0954	-0.0055**	0.0024
		(0.0026)	(0.0035)		(0.0015)	(0.0021)		(0.0019)	(0.0028)
Days in ICU	1.2442	-0.0388**	0.0122	0.8908	-0.0319***	0.0195	0.2610	-0.0121	0.0084
		(0.0122)	(0.0183)		(0.0066)	(0.0105)		(0.0068)	(0.0109)
Length of stay	3.1995	-0.0020	-0.0118	3.9763	0.0049	0.0276	2.5435	0.0044	0.0255
		(0.0210)	(0.0285)		(0.0123)	(0.0177)		(0.0172)	(0.0250)
Transferred	0.3374	-0.0010	-0.0043	0.2659	-0.0017	0.0013	0.1295	-0.0015	0.0029
		(0.0025)	(0.0033)		(0.0016)	(0.0021)		(0.0024)	(0.0033)
N	33,771	33,771	33,771	87,296	87,296	87,296	28,361	28,361	28,361

Note: This table reports the results of estimating equation (3.2) with the listed outcome for patients with the listed diagnosis. The key independent variables are ICU and non-ICU census at the start of the day the patient arrived, scaled so that a one unit change can be interpreted as the effect of moving from the 25th percentile to the 75th percentile in the distribution of proportional deviations from average census. The regression also includes a full set of age decile-sex interactions, a full set of Elixhauser comorbidities, indicators for primary diagnosis as described in the text, and date, hospital x year, hospital x month, and hospital x day-of-week fixed effects. The underlying sample is the sample of new hospital arrivals with coronary artery disease described in Section 3.3. Standard errors are clustered at the hospital-date level. Statistical significance is denoted as follows: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Figure 3.3: Non-linearity in the effect of ICU congestion on ICU admission probability



Notes: This figure reports the results of estimating equation (3.2) with scaled ICU census replaced by indicators for each octile of scaled ICU census and where the outcome is an indicator for ICU admission. The coefficient on each octile indicator is plotted at the average ratio of ICU census to hospital-year mean ICU census in that octile. Error bars depict 90 percent confidence intervals for the indicators; failure of confidence intervals for two estimated coefficients to overlap corresponds approximately to rejection of the hypothesis that they are the same at the 5 percent level. The regression includes a full set of age decile-sex interactions, a full set of Elixhauser comorbidities, indicators for primary diagnosis as described in the text, and date, hospital x year, hospital x month, and hospital x day-of-week fixed effects. The underlying sample is the sample of new hospital arrivals with coronary artery disease described in Section 3.3. Standard errors are clustered at the hospital-date level.

A check on whether these results do indeed represent a casual effect of congestion on ICU admission is whether they rise non-linearly with ICU census. At low levels of ICU census, where the ICU capacity constraint does not bind, the probability of ICU admission should be reasonably insensitive to changes in patient census, while the probability of admission should become highly sensitive to ICU census at higher census levels. To test this assumption, I estimate a version of equation (3.2) in which the scaled ICU census variable is replaced with indicator variables for octiles of scaled ICU census. Figure 3.3 plots the results and indicates that the responsiveness of the ICU admission probability does indeed demonstrate the predicted non-linear pattern.

I next examine the effect of congestion on the procedures patients undergo, specifically cardiac catheterization and PCI. Table 3.10 reports these results. There do appear to be effects on treatment receipt. These effects are apparently confined to the low-transfer hospitals, which is unsurprising since the catheterization capabilities of the high-transfer hospitals are generally limited or non-existent. Focusing on the low-transfer hospitals, a one unit increase in scaled ICU census appears to reduce the probability of catheterization and PCI by the first or second day of the stay by approximately one percentage point. For catheterization, this effect appears to fade out by discharge, while for PCI most of the effect remains at discharge. There is no statistically significant evidence that non-ICU census has an effect on treatment receipt, although the relevant coefficients are frequently negative.

Once again, to gain more insight into these effects, I break them down by patient diagnosis, focusing solely on low-transfer hospitals since these are the only hospitals for which there is evidence of an effect of congestion on utilization. Table 3.11 reports the results. For STEMI patients, there is no evidence that high ICU census reduces the probability of catheterization or PCI receipt; indeed the point estimates for this group of patients are generally positive. Rather, the effect of high ICU census appear concentrated among less ill patients. For the non-STEMI acute coronary syndrome patients, a one unit increase in scaled ICU census at arrival reduces the probability of PCI and catheterization early in the hospitalization by approximately one percentage point, although these effects largely fade by discharge. There is also suggestive evidence that high ICU census reduces catheterization and PCI receipt for the non-ACS patients, and these effects appear durable. Because prompt PCI is considered to be extremely important for STEMI patients, while timeliness of PCI is considered relatively less important for other ACS and non-ACS patients, these results appear consistent with rationing according to expected clinical benefit. As before, there is no evidence that non-ICU census has an effect on treatment receipt.

3.4.3 Effects on patient outcomes

I turn finally to evaluating the effects of congestion on two types of patient health outcomes: patient mortality risk and a composite outcome of mortality or subsequent hospital admission.[12] I evaluate

[12]Note that a subsequent admission might not technically be a "readmission" since some of the patients included in this sample are never admitted initially and instead seen only in a the hospital's emergency department or observation

Table 3.10: Effect of congestion on procedure receipt by hospital type

Column:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Hospital sample:	All hospitals			Low-transfer hospitals			High-transfer hospitals		
		Independent variable			Independent variable			Independent variable	
Dependent variable	Mean	ICU cens.	OIP cens.	Mean	ICU cens.	OIP cens.	Mean	ICU cens.	OIP cens.
A. Effect on cardiac catheterization receipt by...									
Arrival day + 0	0.1312	-0.0014*	-0.0002	0.2485	-0.0047	0.0015	0.0603	-0.0007	0.0002
		(0.0007)	(0.0010)		(0.0030)	(0.0034)		(0.0005)	(0.0008)
Arrival day + 1	0.2289	-0.0023**	-0.0021	0.4280	-0.0092**	-0.0068	0.1086	-0.0013	-0.0011
		(0.0008)	(0.0011)		(0.0034)	(0.0040)		(0.0007)	(0.0010)
Arrival day + 2	0.2865	-0.0015	-0.0019	0.5230	-0.0089*	-0.0072	0.1437	-0.0007	-0.0008
		(0.0009)	(0.0012)		(0.0035)	(0.0043)		(0.0008)	(0.0011)
Discharge	0.3602	-0.0002	-0.0013	0.6398	-0.0016	-0.0060	0.1913	-0.0000	-0.0003
		(0.0009)	(0.0013)		(0.0034)	(0.0042)		(0.0009)	(0.0012)
B. Effect on PCI receipt by...									
Arrival day + 0	0.0947	-0.0015**	0.0007	0.1898	-0.0067*	0.0012	0.0372	-0.0007	0.0009
		(0.0006)	(0.0008)		(0.0026)	(0.0031)		(0.0004)	(0.0006)
Arrival day + 1	0.1362	-0.0017**	0.0006	0.2898	-0.0101**	-0.0017	0.0434	-0.0006	0.0009
		(0.0006)	(0.0009)		(0.0031)	(0.0037)		(0.0004)	(0.0006)
Arrival day + 2	0.1579	-0.0017**	0.0003	0.3416	-0.0100**	-0.0031	0.0469	-0.0006	0.0008
		(0.0006)	(0.0009)		(0.0033)	(0.0040)		(0.0005)	(0.0007)
Discharge	0.1848	-0.0013	0.0005	0.4066	-0.0075*	-0.0034	0.0509	-0.0005	0.0011
		(0.0007)	(0.0010)		(0.0035)	(0.0042)		(0.0005)	(0.0007)
N	149,428	149,428	149,428	56,265	56,265	56,265	93,163	93,163	93,163

Note: This table reports the results of estimating equation (3.2) with the listed outcome in the listed hospital sample. Low-transfer hospitals are those that transfer (at any point after arrival) less than 10 percent of their acute coronary syndrome patients. The key independent variables are ICU and non-ICU census at the start of the day the patient arrived, scaled so that a one unit change can be interpreted as the effect of moving from the 25th percentile to the 75th percentile in the distribution of proportional deviations from average census. The regression also includes a full set of age decile-sex interactions, a full set of Elixhauser comorbidities, indicators for primary diagnosis as described in the text, and date, hospital x year, hospital x month, and hospital x day-of-week fixed effects. The underlying sample is the sample of new hospital arrivals with coronary artery disease described in Section 3.3. Standard errors are clustered at the hospital-date level. Statistical significance is denoted as follows: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 3.11: Effect of congestion on procedure receipt at low-transfer hospitals by patient type

Column:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Patient diagnosis:	ST-elevation MI			Non-STEMI ACS			Other coronary artery disease		
		Independent variable			Independent variable			Independent variable	
Dependent variable	Mean	ICU cens.	OIP cens.	Mean	ICU cens.	OIP cens.	Mean	ICU cens.	OIP cens.
A. Effect on cardiac catheterization receipt by...									
Arrival day+ 0	0.6311	0.0050	-0.0057	0.1803	-0.0059	0.0024	0.0971	0.0019	0.0080
		(0.0096)	(0.0110)		(0.0037)	(0.0043)		(0.0062)	(0.0072)
Arrival day + 1	0.7089	0.0071	-0.0090	0.4054	-0.0127**	-0.0041	0.2307	-0.0041	-0.0086
		(0.0088)	(0.0102)		(0.0044)	(0.0052)		(0.0084)	(0.0098)
Arrival day + 2	0.7368	0.0023	-0.0109	0.5199	-0.0109*	-0.0036	0.3286	-0.0062	-0.0095
		(0.0084)	(0.0097)		(0.0046)	(0.0056)		(0.0092)	(0.0110)
Discharge	0.7736	0.0042	-0.0100	0.6617	-0.0012	-0.0024	0.4439	0.0010	-0.0109
		(0.0080)	(0.0093)		(0.0044)	(0.0055)		(0.0098)	(0.0116)
B. Effect on PCI receipt by...									
Arrival day + 0	0.5651	-0.0016	-0.0037	0.1241	-0.0062	-0.0002	0.0374	-0.0053	0.0099*
		(0.0101)	(0.0113)		(0.0032)	(0.0038)		(0.0039)	(0.0048)
Arrival day + 1	0.6155	0.0051	-0.0028	0.2530	-0.0111**	-0.0012	0.0944	-0.0075	-0.0015
		(0.0097)	(0.0110)		(0.0041)	(0.0048)		(0.0061)	(0.0074)
Arrival day + 2	0.6339	0.0016	-0.0093	0.3177	-0.0091*	-0.0033	0.1373	-0.0079	0.0082
		(0.0096)	(0.0108)		(0.0044)	(0.0053)		(0.0072)	(0.0087)
Discharge	0.6574	0.0034	-0.0057	0.3991	-0.0034	-0.0050	0.1910	-0.0115	0.0049
		(0.0095)	(0.0107)		(0.0048)	(0.0056)		(0.0081)	(0.0096)
N	10,560	10,560	10,560	34,638	34,638	34,638	11,067	11,067	11,067

Note: This table reports the results of estimating equation (3.2) with the listed outcome, for patients with the listed diagnosis, and with the sample limited to hospitals that transfer (at any point after arrival) less than 10 percent of their acute coronary syndrome patients. The key independent variables are ICU and non-ICU census at the start of the day the patient arrived, scaled so that a one unit change can be interpreted as the effect of moving from the 25th percentile to the 75th percentile in the distribution of proportional deviations from average census.The regression also includes a full set of age decile-sex interactions, a full set of Elixhauser comorbidities, indicators for primary diagnosis as described in the text, and date, hospital x year, hospital x month, and hospital x day-of-week fixed effects. The underlying sample is the sample of new hospital arrivals with coronary artery disease described in Section 3.3. Standard errors are clustered at the hospital-date level. Statistical significance is denoted as follows: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

these outcomes at horizons of 30, 90, 180, and 365 days after hospital arrival. Due to the required post-arrival follow-up period, the sample sizes are modestly smaller for these analyses.[13,14]

Table 3.12 reports the effect of congestion on patient outcomes for the full sample and for the two hospital sub-samples previously examined. There is no evidence of an effect of either congestion measure on mortality in the full sample. Due to limited precision, however, I cannot rule out the hypothesis that a one unit change in the census measures has an effect on mortality on the order of 0.2-0.3 percentage points, an effect that would be clinically and economically meaningful. For the composite outcome of mortality and hospital admission, on the other hand, the point estimates are generally negative and, as a result, I can come close to ruling out the hypothesis that congestion leads to clinically and economically meaningful degradation in patient outcomes. Overall, however, the most reasonable conclusion is that these analyses are underpowered for the health outcomes of interest.

In light of the evidence from the last subsection that the pattern of changes in procedure receipt varies with patient diagnosis at low-transfer hospitals, Table 3.13 breaks down the outcome results by patient diagnosis at low-transfer hospitals. There is no statistically significant evidence that either measure of hospital congestion affects patient outcomes, but the estimates are quite imprecise. Furthermore, even though the point estimates for the effect of both congestion measures are almost uniformly negative, in essentially all cases, I cannot exclude the hypothesis that congestion leads to increases in mortality or hospital admission that would be clinically and economically meaningful. I conclude once again, therefore, that these analyses do not have sufficient power to reach firm conclusions about the effect of congestion on health outcomes.

---

unit.

[13]In particular, as described previously, I only observe hospital discharges occurring in 2009Q3 or earlier. I therefore only treat my database of hospital *admissions* as being complete for admissions occurring in 2009Q2 or earlier. A 365-day follow-up window with respect to admission outcomes is therefore only possible for admission occurring in 2008Q2 or earlier. Note that because death certificates are available through the end of 2009, observing mortality outcomes is not the binding constraint.

[14]I also exclude patient records that do not report a valid SSN.

Table 3.12: Effect of congestion on patient outcomes at various time horizons by hospital type

Column:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Hospital sample:	All hospitals			Low-transfer hospitals			High-transfer hospitals		
		Independent variable			Independent variable			Independent variable	
Dependent variable	Mean	ICU cens.	OIP cens.	Mean	ICU cens.	OIP cens.	Mean	ICU cens.	OIP cens.
A. Patient death by...									
30 days	0.0739	0.0007	-0.0001	0.0606	-0.0012	0.0008	0.0816	0.0009	-0.0002
		(0.0008)	(0.0010)		(0.0021)	(0.0023)		(0.0009)	(0.0012)
90 days	0.1047	0.0002	-0.0009	0.0869	-0.0025	-0.0002	0.1152	0.0005	-0.0009
		(0.0009)	(0.0012)		(0.0024)	(0.0027)		(0.0010)	(0.0013)
180 days	0.1312	0.0005	0.0001	0.1102	-0.0026	0.0002	0.1435	0.0010	0.0005
		(0.0010)	(0.0013)		(0.0026)	(0.0030)		(0.0011)	(0.0015)
365 days	0.1698	0.0002	0.0002	0.1444	0.0003	0.0002	0.1846	0.0005	0.0007
		(0.0011)	(0.0014)		(0.0029)	(0.0034)		(0.0012)	(0.0016)
B. Patient death or hospital admission by...									
30 days	0.2484	-0.0006	-0.0016	0.2349	-0.0067	-0.0051	0.2563	-0.0001	-0.0009
		(0.0013)	(0.0018)		(0.0037)	(0.0042)		(0.0014)	(0.0020)
90 days	0.4150	-0.0028	-0.0028	0.3926	-0.0081*	-0.0004	0.4281	-0.0022	-0.0031
		(0.0015)	(0.0020)		(0.0041)	(0.0048)		(0.0016)	(0.0022)
180 days	0.5301	-0.0019	-0.0026	0.5061	-0.0079	0.0003	0.5441	-0.0012	-0.0032
		(0.0015)	(0.0020)		(0.0042)	(0.0049)		(0.0016)	(0.0022)
365 days	0.6541	-0.0025	-0.0008	0.6292	-0.0035	0.0055	0.6687	-0.0022	-0.0018
		(0.0014)	(0.0019)		(0.0041)	(0.0046)		(0.0015)	(0.0021)
N	126,739	126,739	126,739	46,692	46,692	46,692	80,047	80,047	80,047

Note: This table reports the results of estimating equation (3.2) with the listed outcome in the listed hospital sample. Low-transfer hospitals are those that transfer (at any point after arrival) less than 10 percent of their acute coronary syndrome patients. The key independent variables are ICU and non-ICU census at the start of the day the patient arrived, scaled so that a one unit change can be interpreted as the effect of moving from the 25th percentile to the 75th percentile in the distribution of proportional deviations from average census. The regression also includes a full set of age decile-sex interactions, a full set of Elixhauser comorbidities, indicators for primary diagnosis as described in the text, and date, hospital x year, hospital x month, and hospital x day-of-week fixed effects. The underlying sample is the sample of new hospital arrivals with coronary artery disease described in Section 3.3. Standard errors are clustered at the hospital-date level. Statistical significance is denoted as follows: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 3.13: Effect of congestion on patient outcomes at low-transfer hospitals by patient type

Column:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Primary diagnosis:	ST-elevation MI			Non-STEMI ACS			Other coronary artery disease		
		Independent variable			Independent variable			Independent variable	
Dependent variable	Mean	ICU cens.	OIP cens.	Mean	ICU cens.	OIP cens.	Mean	ICU cens.	OIP cens.
A. Patient death by...									
30 days	0.1403	0.0005	0.0088	0.0510	-0.0039	0.0013	0.0163	0.0012	0.0026
		(0.0081)	(0.0092)		(0.0025)	(0.0028)		(0.0029)	(0.0035)
90 days	0.1637	-0.0035	0.0013	0.0814	-0.0047	0.0005	0.0324	-0.0014	0.0030
		(0.0084)	(0.0098)		(0.0029)	(0.0033)		(0.0043)	(0.0047)
180 days	0.1815	-0.0019	-0.0007	0.1074	-0.0046	0.0010	0.0526	-0.0072	0.0029
		(0.0086)	(0.0100)		(0.0033)	(0.0038)		(0.0052)	(0.0062)
365 days	0.2043	-0.0005	0.0014	0.1465	-0.0027	0.0010	0.0821	0.0033	0.0006
		(0.0089)	(0.0102)		(0.0037)	(0.0043)		(0.0062)	(0.0076)
B. Patient death or hospital admission by...									
30 days	0.2862	-0.0153	0.0071	0.2273	-0.0074	-0.0030	0.2110	0.0014	-0.0168
		(0.0108)	(0.0122)		(0.0047)	(0.0053)		(0.0093)	(0.0115)
90 days	0.4084	-0.0117	-0.0120	0.3920	-0.0091	0.0068	0.3799	-0.0003	-0.0055
		(0.0114)	(0.0129)		(0.0053)	(0.0061)		(0.0109)	(0.0134)
180 days	0.4962	-0.0068	-0.0147	0.5107	-0.0095	0.0066	0.5010	-0.0086	-0.0084
		(0.0115)	(0.0130)		(0.0054)	(0.0062)		(0.0111)	(0.0135)
365 days	0.5949	-0.0063	-0.0139	0.6358	-0.0049	0.0104	0.6403	-0.0018	-0.0005
		(0.0112)	(0.0130)		(0.0052)	(0.0058)		(0.0106)	(0.0130)
N	8,606	8,606	8,606	28,826	28,826	28,826	9,260	9,260	9,260

Note: This table reports the results of estimating equation (3.2) with the listed outcome for patients with the listed primary diagnosis, with the sample limited to hospitals that transfer (at any point after arrival) less than 10 percent of their acute coronary syndrome patients. The key independent variables are ICU and non-ICU census at the start of the day the patient arrived, scaled so that a one unit change can be interpreted as the effect of moving from the 25th percentile to the 75th percentile in the distribution of proportional deviations from average census.The regression also includes a full set of age decile-sex interactions, a full set of Elixhauser comorbidities, indicators for primary diagnosis as described in the text, and date, hospital x year, hospital x month, and hospital x day-of-week fixed effects. The underlying sample is the sample of new hospital arrivals with coronary artery disease described in Section 3.3. Standard errors are clustered at the hospital-date level. Statistical significance is denoted as follows: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

3.5 Discussion and conclusion

I find that residual variation in hospital patient census does appear to meet the conditions required to identify the effects of hospital congestion on patient care and patient outcomes. In general, a one unit change in scaled ICU census appears to reduce the probability of ICU admission by one percentage point and, at low-transfer hospitals, a one unit change in scaled ICU census appears to reduce early receipt of catheterization procedures by approximately the same amount. In both cases, these reductions in utilization appear concentrated among the patients likely to have the smallest clinical benefit. There is no evidence that the observed reductions in utilization have an adverse effect on patient health outcomes, but power is lacking.

One question in interpreting the results is why ICU census seems to have an effect on care, but non-ICU census does not. With respect to ICU admission, of course, this is a simple reflection of the ICU capacity constraint. With respect to receipt of catheterization and PCI, however, the reason for this pattern is less clear. One hypothesis might be that changes in procedure utilization are a direct result of the changes in ICU admission patterns, perhaps because the physicians managing patient care in the ICU have different views on appropriate patient care than those in the rest of the hospital. This hypothesis, however, is inconsistent with the fact that the effects on procedure utilization are larger than the effect on ICU admission, so even if every patient excluded from the ICU had his care modified in this way (which seems highly unlikely) this mechanism would not be enough to account for the results. It seems more likely that ICU census is a marker of the hospital's current complement of seriously ill patients and, thus, a marker of stress on a broad set of hospital resources, including catheterization facilities and perhaps physician time. Gaining more insight into this question is a priority for future work.

The most important open question left by these results, however, is whether congestion has a meaningful effect on patient outcomes. Whether the application of these estimates is ultimately gauging the benefits of additional capacity or testing theories about the cross-sectional pattern of hospital productivity, obtaining precise answers to this question is essential. Obtaining access to larger datasets that permit more precise estimation should therefore be the top priority for future work.

References

Abadie, Alberto, "Semiparametric Difference-in-Differences Estimators," *Review of Economic Studies*, 2005, *72* (1), 1–19.

Abowd, John M., Francis Kramarz, and David N. Margolis, "High Wage Workers and High Wage Firms," *Econometrica*, March 1999, *67* (2), 251–333.

_ **, Robert H. Creecy, and Francis Kramarz**, "Computing Person and Firm Fixed Effects Using Linked Longitudinal Employer-Employee Data," Technical Paper TP-2002-06, United States Cenus Bureau April 2002.

Almond, Douglas, Hilary W. Hoynes, and Diane Whitmore Schanzenbach, "Inside the War on Poverty: The Impact of Food Stamps on Birth Outcomes," *Review of Economics and Statistics*, 2011, *93*, 387–403.

American College of Obstetricians and Gynecologists, "Guidelines for vaginal delivery after a cesarean childbirth," ACOG Committee Statement, American College of Obstetricians and Gynecologists 1982.

_ , "Guidelines for vaginal delivery after a previous cesarean birth," ACOG Committee Statement, American College of Obstetricians and Gynecologists 1984.

_ , "Guidelines for vaginal delivery after a previous cesarean birth," ACOG Committee Opinion 64, American College of Obstetricians and Gynecologists 1988.

_ , "Vaginal delivery after a previous cesarean birth," ACOG Committee Opinion 143, American College of Obstetricians and Gynecologists 1994.

_ , "Vaginal delivery after previous cesarean birth," ACOG Practice Patterns 1, American College of Obstetricians and Gynecologists 1995.

_ , "Vaginal delivery after previous cesarean delivery," ACOG Practice Bulletin 2, American College of Obstetricians and Gynecologists 1998.

_ , "Vaginal delivery after previous cesarean delivery," ACOG Practice Bulletin 5, American College of Obstetricians and Gynecologists 1999.

_ , "Ob Gyns Issue Less Restrictive VBAC Guidelines," July 2010.

_ , "Vaginal delivery after previous cesarean delivery," ACOG Practice Bulletin 115, American College of Obstetricians and Gynecologists 2010.

Anderson, David, Bruce Golden, Wolfgang Jank, and Edward Wasil, "The impact of hospital utilization on patient readmission rate," *Health Care Management Science*, 2012, *15* (1), 29–36.

_ , Carter Price, Bruce Golden, Wolfgang Jank, and Edward Wasil, "Examining the discharge practices of surgeons at a large medical center," *Health Care Management Science*, 2011, *14* (4), 338–347.

Anderson, Jeffrey L., Cynthia D. Adams, Elliott M. Antman, Charles R. Bridges, Robert M. Califf, Donald E. Casey, William E. Chavey II, Francis M. Fesmire, Judith S. Hochman, Thomas N. Levin, A. Michael Lincoff, Eric D. Peterson, Pierre Theroux, Nanette K. Wenger, and R. Scott Wright, "2012 ACCF/AHA Focused Update Incorporated Into the ACCF/AHA 2007 Guidelines for the Management of Patients With Unstable Angina/Non-ST-Elevation Myocardial Infarction," *Journal of the American College of Cardiology*, 2013, *61* (23), e179–e347.

Angrist, Joshua D. and Jörn-Steffen Pischke, *Mostly Harmless Econometrics: An Empricist's Companion*, Princeton University Press, 2009.

Arias, Elizabeth, "United States Life Tables, 2008," National Vital Statistics Report 61(3), National Center for Health Statistics 2012.

Asper, Faith and Erin Mann, "Medicare Managed Care Enrollees and the Medicare Utilization Files," KnowledgeBase Article 114, Research Data Assistance Center 2011.

Baicker, Katherine, Kasey S. Buckles, and Amitabh Chandra, "Geographic Variation in the Appropriate Use of Cesarean Delivery," *Health Affairs*, 2006, *25*, 355–367.

Bartel, Ann P., Ciaran S. Phibbs, Nancy Beaulieu, and Patricia Stone, "Human Capital and Organizational Performance: Evidence from the Healthcare Sector," Working Paper 17474, National Bureau of Economic Research 2011.

Bavry, Anthony A., Dharam J. Kumbhani, Andrew N. Rassi, Deepak L. Bhatt, and Arman T. Askari, "Benefit of Early Invasive Therapy in Acute Coronary Syndromes: A Meta-Analysis of Contemporary Randomized Clinical Trials," *Journal of the American College of Cardiology*, 2006, *48* (7), 1319–1325.

Biondi-Zoccai, Giusepp G. L., Antonio Abbate, Pierfrancesco Agostoni, Luca Testa, Francesco Burzotta, Marzia Lotrionte, Carlo Trani, and Luigi M. Biasucci, "Long-term benefits of an early invasive management in acute coronary syndromes depend on intracoronary stenting and aggressive antiplatelet treatment: A metaregression," *American Heart Journal*, 2005, *149* (3), 504–511.

Callaghan, William M., Andrea P. MacKay, and Cynthia J. Berg, "Identification of severe maternal morbidity during delivery hospitalizations, United States, 1991-2003," *American Journal of Obstetrics and Gynecology*, 2008, *199*, 133.e1–133.e8.

Card, David, David Lee, Zhuan Pei, and Andrea Weber, "Nonlinear Policy Rules and the Identification and Estimation of Causal Effects in a Generalized Regression Kink Design," Working Paper 18564, National Bureau of Economic Research 2012.

Carneiro, Anabela, Paulo Guimarães, and Pedro Portugal, "Real Wages and the Business Cycle: Accounting for Worker, Firm, and Job Title Heterogeneity," *American Economic Journal: Macreoeconomics*, April 2012, *4* (2), 133–152.

Case, Barbara D., R. Corcoran, Norman Jeffcoate, and G.H. Randle, "Caesarean Section and its Place in Modern Obstetric Practice," *Journal of Obstetrics and Gynaecology of the British Commonwealth*, 1971, *78*, 203–214.

Center for Medicare and Medicaid Services, *Medicare Claims Processing Manual* 2013.

Center for Studying Health System Change, *Technical Publications* 2013.

Centers for Disease Control and Prevention, *Documentation of the Detail Natality Public Use File for 2004* 2006.

_ , *User Guide to the 2010 Natality Public Use File* 2012.

Chandra, Amitabh and Douglas O. Staiger, "Productivity Spillovers in Health Care: Evidence from the Treatment of Heart Attacks," *Journal of Political Economy*, 2007, *115*, 103–140.

_ **and Jonathan Skinner**, "Technology Growth and Expenditure Growth in Health Care," *Journal of Economic Literature*, 2012, *50* (3), 645–680.

Chernozhukov, Victor, Iván Fernández-Val, Jinyong Han, and Whitney Newey, "Average and Quatile Effects in Nonseparable Panel Models," November 2010. Unpublished manuscript.

Choudhry, Niteesh K., Geoffrey M. Anderson, Andreas Laupacis, Dennis Ross-Degnan, Sharon-Lise T. Normand, and Stephen B. Soumerai, "Impact of adverse events on prescribing warfarin in patients with atrial fibrillation: matched pair analysis," *British Medical Journal*, 2006, *332*, 141–143.

Congressional Budget Office, "Technological Change and the Growth of Health Care Spending," Congressional Budget Office Paper 2764, Congressional Budget Office 2008.

Cragin, Edwin B., "Conservatism in Obstetrics," *New York Medical Journal*, 1916, *104* (1), 1–3.

Cutler, David M., *Your Money or Your Life*, Oxford University Press, 2004.

_ **and Robert S. Huckman**, "Technological development and medical productivity: the diffusion of angioplasy in New York State," *Journal of Health Economics*, 2003, *22* (2), 187–217.

_ **, Mark McClellan, Jospeh P. Newhouse, and Dahlia Remler**, "Are Medical Prices Declining? Evidence from Heart Attack Treatments," *Quarterly Journal of Economics*, 1998, *113* (4), 991–1024.

_ **, Robert S. Huckman, and Jonathan T. Kolstad**, "Input Constraints and the Efficiency of Entry: Lessons from Cardiac Surgery," *American Economics Journal: Economic Policy*, 2010, *2* (1), 51–76.

Davidson, Russell and James G. MacKinnon, *Econometric Theory and Methods*, Oxford University Press, 2004.

Deutsch, Frank and Hein Hundal, "The Rate of Convergence for the Method of Alternating Projections, II," *Journal of Mathematical Analysis and Applications*, 1997, *205*, 381–405.

Dewhurst, C. J., "The Ruptured Caesarean Section Scar," *Journal of Obstetrics and Gynaecology of the British Commonwealth*, 1957, *64* (1), 113–118.

Dickstein, Michael J., "Efficient Provision of Experience Goods: Evidence from Antidepressent Choice," January 2012. Unpublished manuscript.

Dolan, Paul, "The Measurement of Health-Related Quality of Life for Use in Resource Allocation Decisions in Health Care," in Anthony J. Culyer and Joseph P. Newhouse, eds., *Handbook of Health Economics*, Vol. 1, Elsevier, 2000, pp. 1723–1760.

Doyle, Joseph J., Steven M. Ewer, and Todd H. Wagner, "Returns to physician human capital: Evidence from patients randomized to physician teams," *Journal of Health Economics*, 2010, *29* (6), 866–882.

Dranove, David and Yasutora Watanbe, "Influence and Deterrence: How Obstetricians Respond to Litigation Against Themselves and Their Colleagues," *American Law and Economics Review*, 2009, *12*, 69–94.

Elixhauser, Anne, Claudia Steiner, Robert D. Harris, and Rosanna M. Coffey, "Comorbidity Measures for Use with Administrative Data," *Medical Care*, 1998, *36* (1), 8–27.

Ellis, Randall P. and Thomas G. McGuire, "Provider Behavior Under Prospective Reimbursement," *Journal of Health Economics*, 1986, *5* (2), 129–151.

Emmett, Clare L., Alan A. Montgomery, and Deirdre J. Murphy, "Preferences for mode of delivery after previous caesarean section: what do women want, what do they get and how do they value outcomes?," *Health Expectations*, 2010, *14*, 397–404.

Epstein, Andrew J. and Sean Nicholson, "The formation and evolution of physician treatment styles: An application to cesarean sections," *Journal of Health Economics*, 2009, *28*, 1126–1140.

_ , Jonathan D. Ketcham, and Sean Nicholson, "Specialization and matching in professional services firms," *RAND Journal of Economics*, 2010, *41* (4), 811–834.

Fee, Christopher, Ellen J. Weber, Carley A. Maak, and Peter Bacchetti, "Effect of Emergency Department Crowding on Time to Antibiotics in Patients Admitted With Community-Acquired Pneumonia," *Annals of Emergency Medicine*, 2007, *50* (5), 501–509.

Fihn, Stephan D., Julius M. Gardin, Jonathan Abrams, Kathleen Berra, James C. Blankenship. Dalls, Pamela S. Douglas, Joanne M. Foody, Thomas C. Gerber, Alan L. Hinderliter, Spencer B. King III, Paul D. KIigfield, Harlan M. Krumholz, Raymond Y. K. Kwong, Michael J. Lim, Jane A. Linderbaum, Michael J. Mack, Mark A. Munger, Richard L. Prager, Joseph F. Sabik, Leslee J. Shaw, Joanna D. Sikkema Sidney C. Smith Jr., Sidney C. Smith Jr., John A. Spertus, and Sankey V. Williams, "2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS Guideline for the Diagnosis and Management of Patients With Stable Ischemic Heart Disease: A Report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines, and the American College of Physicians, American Association for Thoracic Surgery, Preventive Cardiovascular Nurses Association, Society for Cardiovascular Angiography and Interventions, and

Society of Thoracic Surgeons," *Journal of the American College of Cardiology*, 2012, *60* (24), e44–e164.

Fisher, Elliott S., David E. Wennberg, Thérèse A. Stukel, Daniel J. Gottlieb, and F. L. Lucas, "The Implications of Regional Variations in Medicare Spending. Part 1: The Content, Quality, and Accessibility of Care," *Annals of Internal Medicine*, 2003, *138* (4), 273–287.

Flamm, Bruce L., "Once a Cesarean, Always a Controversy," *Obstetrics and Gynecology*, 1997, *90* (2), 312–315.

_ , Janice R. Goings, Yunbao Liu, and Girma Wolde-Tsadik, "Elective Repeat Cesarean Delivery Versus Trial of Labor: A Prospective Multicenter Study," *Obstetrics and Gynecology*, 1994, *83* (6), 927–932.

Frank, Richard G. and Richard J. Zeckhauser, "Custom-made versus ready-to-wear treatments: Behavioral propensities in physician' choices," *Journal of Health Economics*, 2007, *26* (6), 1101–1127.

Freedman, Seth, "Capacity Utilization in Health Care: The Effect of Empty Beds on Neonatal Intensive Care Admission," December 2012. Unpublished manuscript.

Gaure, Simen, "OLS with multiple high dimensional category variables," *Computational Statistics and Data Analysis*, 2013, *66*, 8–18.

Gawande, Atul, *Complications: A Surgeon's Note on an Imperfect Science*, Picador, 2002.

Gehrki, Bernie, "A Different Kind of Diagnosis: How to Limit Potential Pitfalls in Catheterization Lab Design," *Cath Lab Digest*, 2004, *12* (7).

Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin, *Bayesian Data Analysis*, Chapman and Hall/CRC, 2004.

Greene, Michael F., "Vaginal Delivery after Cesarean Section - Is the Risk Acceptable?," *New England Journal of Medicine*, 2001, *345* (1), 54–55.

Gregory, Kimberly D., Lisa M. Korst, Jeffrey A. Gornbein, and Lawrence D. Platt, "Using Administrative Data to Identify Indications for Elective Primary Cesarean Delivery," *Health Services Research*, 2002, *37* (5), 1387–1401.

Grytten, Jostein, Irene Skau, and Rune Sørensen, "Do Mothers Decide? The Impact of Preferences on Healthcare," *Journal of Human Resources*, 2013, *48*, 142–168.

Guimarães, Paulo and Pedro Portugal, "A simple feasible procedure to fit models with high-dimensional fixed effects," *The Stata Journal*, 2010, *10* (4), 628–649.

Guise, Jeanne-Marie, Karen Eden, Cathy Emeis, Mary Anna Denman, Nicole Marshall, Rongwei Fu, Rosalind Janik, Peggy Nygren, Miranda Walker, and Marian McDonagh, "Vaginal Birth After Cesarean: New Insights," Evidence Report/Technology Assessment 191, Agency for Healthcare Research and Quality 2010.

_ , Marian S. McDonagh, Jason Hashima, Dale F. Kraemer, Karen B. Eden, Michelle Berlin, Peggy Nygren, Particia Osterweil, Kathryn Pyle Krages, and Mark Helfand, "Vaginal Birth After Cesarean," Evidence Report/Technology Assessment 71, Agency for Healthcare Research and Quality 2003.

_ , **Mary Anna Denman, Cathy Emeis, Nicole Marshall, Miranda Walker, Rongwei Fu, Rosalind Janik, Peggy Nygren, Karen B. Eden, and Marian McDonagh**, "Vaginal Birth After Cesarean: New Insights on Maternal and Neonatal Outcomes," *Obstetrics and Gynecology*, 2010, *115* (6), 1267–1278.

Hadfield, Jarrod D., "MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package," *Journal of Statistical Software*, 2010, *33* (2), 1–22.

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw, "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," *Econometrica*, 2001, *69*, 201–209.

Halperin, Israel, "The Product of Projection Operations," *Acta Scientiarum Mathematicarum*, 1962, *23*, 96–99.

Hannan, Edward L., Chuntao Wu, Gary Walford, Spencer B. King III, David R. Holmes Jr., John A. Ambrose, Samin Sharma, Stanley Katz, Luther T. Claker, and Robert H. Jones, "Volume-Outcome Relationships for Percutaneous Coronary Interventions in the Stent Era," *Circulation*, 2005, *112*, 1171–1179.

Healthcare Cost and Utilization Project, *User Guide for the 1988-2002 NIS Trends Supplemental Files* Agency for Healthcare Research and Quality 2006.

_ , *Introduction to the HCUP Nationwide Inpatient Sample (NIS) 2010* Agency for Healthcare Research and Quality 2012.

Heckman, James, Hidehiko Ichimura, Jeffrey Smith, and Petra Todd, "Characterizing Selection Bias Using Experimental Data," *Econometrica*, 1998, *66* (5), 1017–1098.

Heckman, James J., Hidehiko Ichimura, and Petra E. Todd, "Matching As An Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program," *Review of Economic Studies*, 1997, *64*, 605–654.

Henderson, Robert A., Stuart J. Pockock, Tim C. Clayton, Rosemary Knight, Keith A. A. Fox, Desmond G. Julian, and Douglas A. Chamberlain, "Seven-Year Outcome in the RITA-2 Trial: Coronary Angioplasty Versus Medical Therapy," *Journal of the American College of Cardiology*, 2003, *42* (7), 1161–1170.

Henry, Olivia A., Kimberly D. Gregory, Calvin J. Hobel, and Lawrence D. Platt, "Using ICD-9 Codes to Identify Indications for Primary and Repeat Cesearean Sections: Agreement with Clinical Records," *American Journal of Public Health*, 1995, *85* (8), 1143–1146.

Hilger, Nathaniel G., "How Does Family Income Affect College Enrollment? Evidence from the Timing of Parental Layoffs," November 2012. Unpublished manuscript.

Hoenig, Michel R., Constantine N. Aroney, and Ian A. Scott, "Early invasive versus conservative strategies for unstable angina and non-ST elevation myocardial infarction in the stent era," *Cochrane Database of Systematic Reviews*, 2010, *3.*

Holland, Paul W., "Statistics and Causal Inference," *Journal of the American Statistical Association*, 1986, *81* (396), 945–960.

Iapichino, Gaetano, Luciano Gattinoni, Danilo Radrizzani, Bruno Smini, Guido Bertolini, Luca Ferla, Giovanni Mistraletti, Francesca Porta, and Dinis R. Miranda, "Volume of activity and occupancy rate in intensive care units. Association with mortality," *Intensive Care Medicine*, 2004, *30* (2), 290–297.

Imbens, Guido and Karthik Kalyanaraman, "Optimal Bandwidth Choice for the Regression Discontinuity Estimator," *Review of Economic Studies*, 2012, *79* (3), 933–959.

Imbens, Guido W. and Jeffrey M. Wooldridge, "Recent Developments in the Econometrics of Program Evaluation," *Journal of Economic Literature*, 2009, *47* (1), 5–86.

_ **and Thomas Lemieux**, "Regression discontinuity designs: A guide to practice," *Journal of Econometrics*, 2008, *142*, 615–635.

Jacobson, Louis S., Robert J. LaLonde, and Daniel G. Sullivan, "Earnings Losses of Displaced Workers," *American Economic Review*, 1993, *83* (4), 685–709.

Kc, Diwas S. and Christian Terwiesch, "Impact of Workload on Service Time and Patient Safety: An Econometric Analysis," *Management Science*, 2009, *55* (9), 1486–1498.

Keeley, Ellen C., Judith A. Boura, and Cindy L. Grines, "Primary angioplasy versus intravenous thrombolytic therapy for acute myocardial infarction: a quantitative review of 23 randomised trials," *The Lancet*, 2003, *361* (9351), 13–20.

Kim, Song-Hee, Carri W. Chan, Marcelo Olivares, and Gabriel Escobar, "ICU Admission Control: An Empirical Study of Capacity Allocation and its Implication on Patient Outcomes," February 2013. Unpublished manuscript.

Kolm, Paul, Zugui Zhan, John Spertus, and William S. Weintraub, "Abstract 197: Multilevel Analysis Of Clinical Trial Data: Is It Necessary?," *Circulation: Cardiovascular Quality and Outcomes*, 2013, *6* (3), A197.

Korst, Lisa M., Kimberly D. Gregory, and Jeffrey A. Gombein, "Elective primary caesarean delivery: accuracy of administrative data," *Paediatric and Perinatal Epidemiology*, 2004, *18*, 112–119.

_ , _ , **Moshe Fridman, and Jeffery P. Phelan**, "Nonclinical Factors Affecting Women's Access to Trial of Labor After Cesarean Delivery," *Clinics in Perinatology*, 2011, *38*, 193–216.

Landon, Mark B., John C. Hauth, Kenneth J. Leveno, Catherine Y. Spong, Sharon Leindecker, Michael W. Varner, Atef H. Moawad, Steve N. Caritis, Margaret Harper, Ronald J. Wapner, Yoram Sorokin, Menachem Miodovnik, Mashall Carpenter, Alan M. Peaceman, Mary Jo O'Sullivan, Baha Sibai, Oded Langer, John M. Thorp, Susan M. Ramin, Brian M. Mercer, and Steven G. Gabbe, "Maternal and Perinatal Outcomes Associated with a Trial of Labor after Prior Cesarean Delivery," *New England Journal of Medicine*, 2004, *351* (25), 2581–2589.

Lavin, Justin P., Robert J. Stephens, Menachem Miodovnik, and Tom P. Barden, "Vaginal Delivery in Patients with a Prior Cesarean Section," *Obstetrics and Gynecology*, 1982, *59* (2), 135–148.

Lee, David S. and David Card, "Regression discontinuity inference with specification error," *Journal of Econometrics*, 2008, *142*, 655–674.

Levine, Glenn N., Erin R. Bates, James C. Blankenship, Steven R. Bailey, John A. Bittl, Bojan Cercek, Charles E. Chambers, Stephen G. Ellis, Robert A. Guyton, Steven M. Hollenberg, Umesh N. Khot, Richard A. Lange, Laura Mauri, Roxana Mehran, Issam D. Moussa, Debabrata Mukherjee, Bramajee K. Nallamothu, and Henry H. Ting, "2011 ACCF/AHA/SCAI Guideline for Percutaneous Coronary Intervention: A Report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines and the Society for Cardiovascular Angiography and Interventions," *Journal of the American College of Cardiology*, 2011, *58* (24), e44–122.

Lieu, Tracy A., Robert J Lundstrom, Thomas Ray, Bruce H. Fireman, R. Jan Gurley, and William M. Parmley, "Initial Cost of Primary Angioplasty for Acute Myocardial Infarction," *Journal of the American College of Card*, 1996, *28* (4), 882–889.

Lydon-Rochelle, Mona, Victoria L. Holt, Thomas R. Easterling, and Diane P. Martin, "Risk of Uterine Rupture During Labor Among Women with a Prior Cesarean Delivery," *New England Journal of Medicine*, 2001, *345*, 3–8.

MacDorman, Marian, Eugene Declerq, and Fay Menacker, "Recent Trends and Patterns in Cesarean and Vaginal Birth After Cesarean (VBAC) Deliveries in the United States," *Clinics in Perinatology*, 2011, *38*, 179–192.

Mankiw, N. Gregory and Micahel D. Whinston, "Free Entry and Social Inefficiency," *The RAND Journal of Economics*, 1986, *17* (1), 48–58.

McClellan, Mark and Joseph P. Newhouse, "The marginal cost-effectiveness of medical technology: A panel instrumental-variables approach," *Journal of Econometrics*, 1997, *77*, 39–64.

McMahon, Michael J., Edwin R. Luther, Watson A. Bowes, and Andrew F. Olshan, "Comparison of a Trial of Labor with An Elective Second Cesarean Section," *New England Journal of Medicine*, 1996, *335* (10), 689–695.

Mehta, Shamir R., Christopher P. Cannon, Keith A. A. Fox, Lars Wallentin, William E. Boden, Rudolf Spacek, Petr Widimsky, Peter A. McCullough, David Hunt, Eugene Braunwald, and Salim Yusuf, "Routine vs Selective Invasive Strategies in Patients With Acute Coronary Syndromes: A Collaborative Meta-analysis of Randomized Trials," *Journal of the American Medical Association*, 2005, *293* (23), 2908–2917.

Miquel, Ruth, "Identification of Effects of Dynamic Treatments with a Difference-in-Differences Approach," 2003. Unpublished manuscript.

Moulton, Brent R., "Random Group Effects and the Precision of Regression Estimates," *Journal of Econometrics*, 1986, *32* (3), 385–397.

Murphy, Kevin M. and Robert H. Topel, "The Value of Health and Longevity," *Journal of Political Economy*, 2006, *114* (5), 871–904.

National Conference of State Legislatures, "Certificate of Need Laws: State Health Laws and Programs," March 2012.

National Institutes of Health, "Cesarean Childbirth," *National Institutes of Health Consensus Statement Online*, 1980, *3* (6), 1–30.

_ , "National Institutes of Health Consensus Development Conference Statement: Vaginal Birth After Cesarean: New Insights," *Obstetrics and Gynecology*, 2010, *115* (6), 1279–1295.

Nevalainen, Jaakko, Somnath Datta, and Hannu Oja, "Inference on the marginal distribution of clustered data with informative cluster size," *Statistical Papers*, 2013.

Newhouse, Joseph P., "Medical Care Costs: How Much Welfare Loss?," *Journal of Economic Perspectives*, 1992, *6* (3), 3–21.

_ , *Pricing the Priceless: A Health Care Conundrum*, MIT Press, 2002.

O'Gara, Patrick T., Frederick G. Kushnber, Debora D. Ascheim, Donald E. Casey, Mina K. Chung, James A. de Lemos, Steven M. Ettinger, James C. Fang, Francis M. Fesmire, Barry A. Franklin, Christopher B. Granger, Harlan M. Krumholz, Jane A. Linderbaum, David A. Morrow, L. Kristin Newby, Joseph P. Ornato, Narith Ou, Marth J. Radford, Jacqueline E. Tamis-Holland, Jacqueline E. Tommaso, Cynthia M. Tracy, Y. Joseph Woo, and David X. Zhao, "2013 ACCF/AHA Guideline for the Management of ST-Elevation Myocardial Infarction: Executive Summary," *Circulation*, 2013, *127*, 529–555.

Orlander, Jay D., Thomas W. Barber, and B. Graeme Fincke, "The Morbidity and Mortality Conference: The Delicate Nature of Learning from Error," *Academic Medicine*, 2002, *77* (10), 1001–1006.

O'Sullivan, M.J., Fred Fumia, Kathlee Nolsinger, and Allan G. W. McLeod, "Vaginal Delivery After Cesarean Section," *Clinics in Perinatology*, 1981, *8* (1), 131–143.

Phelps, Charles E., "Information Diffusion and Best Practice Adoption," in A. J. Culyer and J. P. Newhouse, eds., *Handbook of Health Economics*, Vol. 1, Elsevier, 2000.

_ **and Cathleen Mooney**, "Variations in medical practice use: causes and consequences," in Richard J. Arnould, Robert F. Rich, and William D. White, eds., *Competitive approaches to health care reform*, The Urban Institute Press, 1993.

Pierluissi, Edgar, Melissa A. Fischer, Andre R. Campbell, and C. Seth Landefeld, "Discussion of Medical Errors in Morbidity and Mortality Conferences," *Journal of the American Medical Association*, 2003, *290* (21), 2838–2842.

Pines, Jesse M., A. Russell Localio, Judd E. Hollander, William G. Baxt, Hoi Lee, Carolyn Phillips, and Joshua P. Metlay, "The Impact of Emergency Department Crowding Measures on Time to Antibiotics for Patients With Community-Acquired Pneumonia," *Annals of Emergency Medicine*, 2007, *50* (5), 510–516.

_ **, Charles V. Pollack, Deborah B. Diercks, Anna Marie Chang, Frances S. Shofer, and Judd E. Hollander**, "The Association Between Emergency Department Crowding and Adverse Cardiovascular Outcomes in Patients with Chest Pain," *Academic Emergency Medicine*, 2009, *16* (7), 617–624.

_ **, Judd E. Hollander, A. Russell Localio, and Joshua P. Metlay**, "The Association between Emergency Department Crowding and Hospital Performance on Antibiotic Timing for Pneumonia and Percutaneous Intervention for Myocardial Infarction," *Academic Emergency Medicine*, 2006, *13* (8), 873–878.

Pitkin, Roy M., "Once a Cesarean?," *Obstetrics and Gynecology*, 1991, *77* (6), 939.

Porter, Jack, "Estimation in the Regression Discontinuity Model," 2003. Unpublished manuscript.

Price, Joseph and Kosali Simon, "Patient education and the impact of new medical research," *Journal of Health Economics*, 2009, *28* (6), 1166–1174.

Richardson, Drew B., "Increase in patient mortality at 10 days associated with emergency department overcrowding," *Medical Journal of Australia*, 2006, *184* (5), 213–216.

Robinson, James, "Hospitals Respond to Medicare Payment SHortfalls by Both Shifting Costs and Cutting Them, Based on Market Concentration," *Health Affairs*, 2011, *30* (7), 1265–1271.

Rubin, Donald B., "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology*, 1974, *66* (5), 688–701.

_ , "Assignment to Treatment Group on the Basis of a Covariate," *Journal of Educational Statistics*, 1977, *2* (1), 1–26.

Ryan, Thomas J., William B. Bauman, J. Ward Kennedy, Dean J. Kereiakes, Specer B. King III, Ben D. McCallister, Sidney C. Smith Jr., and Daniel L. Ullyot, "Guideliens for Percutaneous Translumina Coronary Angioplasty: A Report of the American College of Cardiology/American Heart Association Task Force on Assessment of Diagnostic and Therapeutic Cardiovascular Procedures (Committee on Percutaneous Coronary Angioplasty)," *Journal of the American College of Cardiology*, 1993, *22* (7), 2033–2054.

Sandler, Danielle H. and Ryan Sandler, "Multiple Event Studies in Public Finance: A Simulation Study with Applications," September 2012. Unpublished manuscript.

Santerre, Rexford E., "The Effect of the ACOG Guideline on Vaginal Births After Cesarean," *Medical Care Research and Review*, 1996, *53* (5), 315–329.

Scott, James R., "Solving the Vaginal Birth After Cesarean Dilemma," *Obstetrics and Gynecology*, 2010, *115* (6), 1112–1113.

Selker, Harry P., John L. Griffith, Fredrick J. Dorey, and Ralph B. D'Agostino, "How Do Physicians Adapt When the Coronary Care Unit Is Full? A Prospective Multicenter Study," *Journal fo the American Medical Association*, 1987, *257* (9), 1181–1185.

Skinner, Jonathan and Douglas Staiger, "Technology Diffusion and Productivity Growth in Health Care," Working Paper 14865, National Bureau of Economic Research 2009.

Smith, Donna M., "Coding for Heart Disease," *For the Record*, 2011, *23* (1), 27.

Stelfox, Henry T., Brenda R. Hemmelgarn, Sean M. Bagshaw, Song Gao, Christopher J. Doig, Cheri Nijssen-Jordan, and Braden Manns, "Intensive Care Unit Bed Availability and Outcomes for Hospitalized Patients With Sudden Clinical Deterioration," *Archives of Internal Medicine*, 2012, *172* (6), 467–474.

Steriopulos, Kathleen and David L. Brown, "Initial Coronary Stent Implanatation with Medical Therapy vs Medical Therapy Alone for Stable Coronary Artery Disease," *Archives of Internal Medicine*, 2012, *172* (4), 312–319.

Strauss, Michael J., James P. LoGerfo, James A. Yeltatzie, Nancy Temkin, and Leonard D. Hudson, "Rationing of Intensive Care Unit Services: An Everyday Occurrence," *Journal of the American Medical Association*, 1986, *255* (9), 1143–1146.

Sun, Benjamin C., Renee Y. Hsia, Robert E. Weiss, David Zingmond, Li-Jung Liang, Weijuan Han, Heather McCreath, and Steven M. Asch, "Effect of Emergency Department Crowding on Outcomes of Admitted Patients," *Annals of Emergency Medicine*, 2013, *61* (6), 605–611.

Tengs, Tammy O. and Amy Wallace, "One Thousand Health-Related Quality-of-Life Estimates," *Medical Care*, 2000, *38* (6), 583–637.

Thomas, Sabu, Rohit Gokhale, William E. Boden, and P.J. Devereaux, "A Meta-Analysis of Randomized Controlled Trials Comparing Percutaneous Coronary Intervention With Medical Therapy in Stable Angina Pectoris," *Canadian Journal of Cardiology*, 2013, *29* (4), 472–482.

Tversky, Amos and Daniel Kahneman, "Availability: A Heuristic for Judging Frequency and Probability," *Cognitive Psychology*, 1973, *5* (2), 207–232.

Weintraub, William S., John A. Spertus, Paul Kolm, David J. Maron, Zefeng Zhang, Claudine Jurkovitz, Wei Zhang, Pamela M. Hartigan, Cheryl Lewis, Emir Veledar, Jim Bowen, Sandra B. Dunbar, Christi Deaton, Stanley Kaufman, Robert A. O'Rourke, Ron Goeree, Paul G Barnett, Koon K. Teo, and William E. Boden, "Effect of PCI on Quality of Life in Patients with Stable Coronary Disease," *New England Journal of Medicine*, 2008, *359* (7), 677–687.

Weiss, J., A. Nannini, S. Fogerty, B. Sachs, F. Frigoletto, D. Roberts, S. Ringer, S. DeJoy, J. P. O'Grady, G. Kraus, J. Weber, National Center for Chronic Disease and Health Promotion, Division of Applied Public Health Training, Epidemiology Program Office Massachusetts Division of Reproductive Health, and Centers for Disease Control, "Use of Hospital Discharge Data to Monitor Uterine Rupture - Massachusetts, 1990-1997," *Morbidity and Mortality Weekly Report*, 2000, *49* (12), 245–248.

Wijeysundera, Harindra C., Brahmajee K. Nallamothu, Harlan M. Krumholz, Jack V. Tu, and Dennis T. Ko, "Meta-analysis: Effects of Percutaneous Coronary Intervention Versus Medical Therapy on Angina Relief," *Annals of Internal Medicine*, 2010, *152* (6), 370–379.

Wooldridge, Jeffrey M., *Econometric Analysis of Cross Section and Panel Data*, MIT Press, 2010.

Zinberg, Stanley, "Vaginal Delivery After Previous Cesarean Delivery: A Continuing Controversy," *Clinical Obstetrics and Gynecology*, 2001, *44* (3), 561–570.

Zingmond, David S., Zhinshen Ye, Susan L. Ettner, and Honghu Liu, "Linking hospital discharge and death records - accuracy and sources of bias," *Journal of Clinical Epidemiology*, 2004, *57*, 21–29.

Zwart, JJ, JM Richters, F Öry, JIP de Vries, KWM Bloemenkamp, and J van Roosmalen, "Uterine rupture in the Netherlands: a nationwide population-based cohort study," *British Journal of Obstetrics and Gynaecology*, 2009, *116*, 1069–1080.

Appendix A

Appendix to Chapter 1

A.1 Estimating the distribution of latent VBAC/TOLAC rates

At various points in the paper, I wish to estimate the distribution of "latent" hospital-level VBAC rates (or TOLAC rates): the probability that a randomly chosen woman delivering at that hospital would undergo VBAC (or TOLAC). Focusing on this distribution of latent rates purges cross-hospital variation in TOLAC or VBAC rates that is attributable solely to small-sample variation and permits me to focus on the permanent differences in practice style that are of interest.

I estimate this distribution of latent rates using a simple beta-binomial mixture model. For simplicity, I focus only on estimate the distribution of latent VBAC rates, but the method used to estimate the distribution of latent TOLAC rates is totally identical. I suppose that, in each year t, each hospital h has a latent VBAC rate ν_{ht} drawn from a beta distribution with parameters α_t and β_t. For each hospital, I observe its total number of deliveries by women with a history of cesarean delivery (N_{ht}), as well as the number of those women who deliver vaginally (V_{ht}). I assume that $V_{ht} \sim \text{binomial}(\nu_{ht}, N_{ht})$.

Under these assumptions, the likelihood function for the parameter vector (α_t, β_t) is given by

$$\mathcal{L}(\alpha_t, \beta_t | \mathbf{V}_t, \mathbf{N}_t) = \prod_i \binom{N_{ti}}{V_{ti}} \frac{\mathrm{B}(V_{ti} + \alpha_t, N_{ti} - V_{ti} + \beta_t)}{\mathrm{B}(\alpha_t, \beta_t)}, \tag{A.1}$$

where $\mathrm{B}(a, b)$ is the beta function, $\mathbf{V}_t$ is the vector of observations of V_{ht} for year t, and $\mathbf{N}_t$ is the vector of observations of N_{ht} for year t. I estimate (α_t, β_t) for each t using data from the Nationwide Inpatient Sample by maximizing this likelihood numerically in Stata using the default optimization

settings. When doing so, I limit the sample to hospitals with at least 20 deliveries by women with a prior cesarean delivery. The excluded hospitals account for 2 percent of all deliveries. I weight each hospital's contribution to the likelihood by its HCUP-assigned sampling weight. In practice, the results are insensitive to the inclusion or exclusion of sampling weights.

A.2 Clinical literature search terms

As described in section 1.3, I identify articles that may report TOLAC case series via a search in the Thomson Reuters World of Science database. Specifically, I include any English language article (excluding letters to the editor) that matches one of the following search expressions:

- (vagina* NEAR/3 (birth* OR born OR deliver*) NEAR/7 ((after* OR follow* OR previous* OR prior OR history) NEAR/3 (cesarean* OR caesarean*)))
- VBAC
- TOLAC
- trial of labor
- trial of labour

I also include English language articles that match both of the following search expressions:

- ((uterine or uterus) NEAR/5 (ruptur* OR tear* OR torn OR perforat* OR lacera* OR trauma* OR damag* OR injur*))
- ((after* OR follow* OR previous* OR prior OR history) NEAR/3 (cesarean* OR caesarean*))

These search expressions are adapted from Guise et al. (2010a).

A.3 Estimated distribution of OB/GYNs across practice types

Table A.1 reports the distribution of obstetricians/gynecologists across practice types based on the physician survey component of the Center for Studying Health System Change's (HSC) Community Tracking Study (CTS) and its successor, the Health Tracking Study (HTS). The CTS is available

Table A.1: Distribution of OB/GYNs across practice types by survey wave

	Practice type						
CTS/HTS Wave	Solo or 2-person	Group (≥ 3)	HMO	Medical school	Hospital-based	Other	**Total**
1996-1997	41	35	4	5	9	6	100
1998-1999	44	30	3	5	10	8	100
2000-2001	41	32	5	7	9	7	100
2004-2005	39	32	6	9	9	6	100
2008	33	45	4	8	8	2	100

Notes: This table reports the share of obstetrician/gynecologists in each practice type in the Community Tracking Study and Health Tracking Study. These data are described in detail in Appendix A.3.

in four waves 1996-1997, 1998-1999, 2000-2001, and 2004-2005, while the HTS has so far released a single wave for 2008.[1]

The CTS used a stratified sampling design that significantly oversampled 60 specific communities to enable calculation of community-specific estimates for those communities; national estimates can be obtained using the provided sampling weight. The HTS used a simpler stratified sampling scheme that also permits obtaining national estimates via a national sampling weight. HSC advises against directly comparing the 2004-2005 and 2008 waves to the earlier waves or to each other. Detailed documentation are available from the technical publications section of the HSC website (HSC, 2013).

A.4 Proofs of learning model propositions

This appendix reports the proofs of the propositions stated in the exposition of the learning model in section 1.2.

Proof of Proposition 1. We start by taking a conditional expectation with respect to the number of additional attempts of labor that occur by a time $s > t$. Because new labor attempts arrive according to the Poisson process $\lambda(t)y_h(t)$ and the probability of multiple events is negligible for

[1]These data were accessed via the data archive maintained by the Inter-university Consortium for Political and Social Research, study numbers 2597, 3267, 3820, 4584, 27202.

small differences $s - t$, we obtain

$$\begin{aligned}\mathbb{E}[\tilde{p}_h(s)|R_h(t), L_h(t), p_h] = [1 - \lambda(t)y_h(t)(s-t) - o(s-t)] &\left[\frac{\alpha(s) + R_h(t)}{\alpha(s) + \beta(s) + L_h(t)}\right] \\ &+ \lambda(t)y_h(t)(1-p_h)(s-t)\left[\frac{\alpha(s) + R_h(t)}{\alpha(s) + \beta(s) + L_h(t) + 1}\right] \\ &+ \lambda(t)y_h(t)p_h(s-t)\left[\frac{\alpha(s) + R_h(t) + 1}{\alpha(s) + \beta(s) + L_h(t) + 1}\right] + o(s-t).\end{aligned}$$

Switching into the alternative parametrization $(\mu(t), \phi(t))$, subtracting $\tilde{p}_h(t)$ from both sides, dividing both sides by $s - t$, taking the limit as $s \to t$, and then doing a modest amount of algebra yields the result. The derivation of $\frac{d}{ds}\mathbb{E}[y_h(s)|R_h(t), L_h(t), p_h]|_{s=t}$ is similar and omitted. □

Proof of Proposition 2. The assumptions on the arrival process, together with assumptions (i) and (ii) imply that the instantaneous rate at which a provider with current experience ℓ makes a new attempt of labor can be bounded below by

$$\delta_\ell = \underline{\lambda}\left[1 - F\left(\frac{\bar{\mu}\underline{c} + \ell}{\underline{c} + \ell}\right)\right] > 0.$$

Since $L_h(s)$ increases monotonically, we have

$$\mathbb{P}(L_h(t) = 0) \leq \exp[-\delta_0 t]\mathbb{P}(L_h(0) = 0),$$

which implies that $\mathbb{P}(L_h(t) = 0) \to 0$ as $t \to \infty$. Similarly, for $k > 0$, we have for any time s and t with $s \geq t$,

$$\mathbb{P}(L_h(s) \leq \ell) \leq \mathbb{P}(L_h(t) \leq \ell - 1) + \exp[-\delta_\ell(s-t)]\mathbb{P}(L_h(t) = k),$$

which, by induction, implies that $\mathbb{P}(L_h(s) \leq \ell) \to 0$ for all ℓ. This implies in turn that $L_h(t) \xrightarrow{a.s.} \infty$ as $t \to \infty$.

The time at which each provider reaches ℓ deliveries, denoted $T_h(\ell)$, is therefore well-defined with probability one. Define the discrete sub-process $R_h^*(\ell) = R_h(T_h(\ell))$, which is the number of ruptures experienced by the provider after ℓ attempts. For each provider h, $R_h(\ell)$ is simply the sum of ℓ Bernoulli trials with success probability p_h, so we must have $\ell^{-1}R_h(\ell) \xrightarrow{a.s.} p_h$ as $\ell \to \infty$. Since $R_h(t)/L_h(t) = R_h^*(L_h(t))/L_h(t)$, we have $R_h(t)/L_h(t) \xrightarrow{a.s.} p_h$ as well.

Finally, if the first part of condition (iii) holds, we write

$$\tilde{p}_h(t) = \frac{\alpha(t)/L_h(t) + R_h(t)/L_h(t)}{[\alpha(t) + \beta(t)]/L_h(t) + 1},$$

and we see that terms other that $R_h(t)/L_h(t)$ vanish since $L_h(t) \xrightarrow{a.s.} \infty$. Applying the fact that $R_h(t)/L_h(t) \xrightarrow{a.s.} p_h$ then completes the proof. If, on the other hand, the second part of condition (iii) holds, then write

$$\begin{aligned}\tilde{p}_h(t) &= \frac{p_0[\alpha(t) + \beta(t)] + \{\alpha(t) - p_0[\alpha(t) + \beta_t]\} + p_0 L_h(t) + \{R_h(t) - p_0 L_h(t)\}}{\alpha(t) + \beta(t) + L_h(t)} \\ &= p_0 + \frac{\{\alpha(t) - p_0[\alpha(t) + \beta(t)]\} + \{R_h(t) - p_0 L_h(t)\}}{\alpha(t) + \beta(t) + L_h(t)},\end{aligned}$$

from which the result follows immediately. □

A.5 Additional econometric details

In this appendix, I provide additional detail on the event study estimator used in this paper. The first subsection provides a a proof of Lemma 2. The second subsection discusses the estimator's small-sample properties.

A.5.1 Proof of Lemma 2

Before proving Lemma 2, it will be helpful to establish an additional lemma, which is essentially a law of large numbers for hierarchical settings like the current one.

Lemma 4. *Let $\{(G_i, Z_i)\}_{i\in\mathbb{N}}$ be an i.i.d. sequence of random tuples where $G_i \in \mathbb{N}$ and Z_i is a $G_i \times k$ matrix. Define $S_i(\mathcal{Z}) = \{j : Z_{ij} \in \mathcal{Z}\}$. For any set $\mathcal{Z}$ such that $0 < \mathbb{E}^*(|S_i(\mathcal{Z})|) < \infty$ and function h such that $\mathbb{E}^*[|\sum_{j\in S_i(\mathcal{Z})} h(Z_{ij})|]$ exists,*

$$\frac{1}{\sum_{i=1}^M |S_i(\mathcal{Z})|} \sum_{i=1}^M \sum_{j\in S_i(\mathcal{Z})} h(Z_{ij}) \xrightarrow{p} \mathbb{E}[h(Z_{ij}) \,|\, Z_{ij} \in \mathcal{Z}]. \tag{A.2}$$

as $M \to \infty$, where $\mathbb{E}^*[\cdot]$ denotes expectation with respect to the distribution of (G_i, Z_i) and $\mathbb{E}[\cdot]$ denotes expectation with respect to the population marginal distribution of Z_{ij}.

Proof. Since $\mathbb{E}[|S_i(\mathcal{Z})|] > 0$, the left-hand-side of (A.2) is well-defined with probability approaching

one, so we can safely focus on this case. We can re-write the left-hand-side of (A.2) as follows:

$$\frac{1}{\sum_{i=1}^{M}|S_i(\mathcal{Z})|}\sum_{i=1}^{M}\sum_{j\in S_i(\mathcal{Z})}h(Z_{ij}) = \left(\frac{1}{M}\sum_{i=1}^{M}|S_i(\mathcal{Z})|\right)^{-1}\left(\frac{1}{M}\sum_{i=1}^{M}\sum_{j\in S_i(\mathcal{Z})}h(Z_{ij})\right).$$

Applying a standard law of large numbers to each term on the right-hand-side then implies that the right-hand-side converges to $\mathbb{E}^*[|S_i(\mathcal{Z})|]^{-1}\mathbb{E}^*[\sum_{j\in S_i(\mathcal{Z})}h(Z_{ij})]$.

The population marginal distribution of Z_{ij} satisfies

$$\mathbb{P}(Z_{ij}\in A) = \frac{1}{\mathbb{E}^*[G_i]}\int_{\mathbb{N}}\left[\sum_{k=1}^{g}\mathbb{P}^*(Z_{ij}\in A\,|\,G_i=g)\right]d\mathbb{P}^*(G_i=g).$$

It follows immediately that

$$\mathbb{E}[h(Z_{ij})\,|\,Z_{ij}\in\mathcal{Z}] = \frac{\int_{z\in\mathcal{Z}}h(z)\,d\mathbb{P}(z)}{\mathbb{E}[\mathbf{1}\{Z_{ij}\in\mathcal{Z}\}]} = \frac{\mathbb{E}^*[\sum_{j\in S_i(\mathcal{Z})}h(Z_{ij})]/\mathbb{E}^*[G_i]}{\mathbb{E}^*[|S_i(\mathcal{Z})|]/\mathbb{E}^*[G_i]} = \frac{\mathbb{E}^*[\sum_{j\in S_i(\mathcal{Z})}h(Z_{ij})]}{\mathbb{E}^*[|S_i(\mathcal{Z})|]},$$

completing the proof.

□

Proof of lemma 2. By condition OO, $B(e,x)\xrightarrow{p}1$ for all $(e,x)\in\mathcal{C}$. Then, applying lemma 4 separately to each group sum in the definition of $\hat{\Delta}^q$ and pointwise to $\hat{W}_{\text{WTOT}}$ demonstrates that each converges to its population counterpart. The result follows. □

A.5.2 Estimator small sample properties

Lemma 2 establishes that the event study estimator used herein is consistent, but the proof relies on an argument that the relevant means are consistent for the desired quantities for each tuple $(e,x)\in\mathcal{C}$. In the current application, however, the number of providers with events can be small or zero for many (e,x) tuples, which raises the question of whether the asymptotic results provide a good guide to the estimator's properties in practice. In this appendix, I show that, under plausible conditions, the estimator is conditionally unbiased for a weighted average of conditional average causal effects (see Imbens and Wooldridge (2009) for a discussion of such estimands). Although this causal effect may differ from the weighted average causal effect described in Lemma 1, this result provides reassurance regarding the estimator's small-sample properties.

Formally, we are interested in the properties of the estimator for any number of sampled providers M. The operator $\mathbb{E}_M[\cdot]$ will denote expectation with respect to the joint distribution

of an M-provider random sample. I establish the properties of the small-sample properties of the estimator under the following condition:

Condition CT$'$ (Finite sample common trends)**.** For all provider sample sizes M, all event time and characteristic tuples $(e,x) \in \mathcal{C}$, and all times-since-event $q, r \in \mathcal{H}$, the following two conditions holds:

$$\begin{aligned}
&\mathbb{E}_M\left[\frac{1}{N_1^q(e,x)}\sum_{(i,j)\in S_1^q(e,x)} Y_{ij}^q(0)\,\middle|\, B(e,x)=1, \{(B(e',x'),\hat{W}(e',x'))\}_{(e',x')\in\mathcal{C}}\right] \\
&\qquad - \mathbb{E}_M\left[\frac{1}{N_1^r(e,x)}\sum_{(i,j)\in S_1^r(e,x)} Y_{ij}^r(0)\,\middle|\, B(e,x)=1, \{(B(e',x'),\hat{W}(e',x'))\}_{(e',x')\in\mathcal{C}}\right] \\
&= \mathbb{E}_M\left[\frac{1}{N_0^q(e,x)}\sum_{(i,j)\in S_0^q(e,x)} Y_{ij}^q(0)\,\middle|\, B(e,x)=1, \{(B(e',x'),\hat{W}(e',x'))\}_{(e',x')\in\mathcal{C}}\right] \\
&\qquad - \mathbb{E}_M\left[\frac{1}{N_0^r(e,x)}\sum_{(i,j)\in S_0^r(e,x)} Y_{ij}^r(0)\,\middle|\, B(e,x)=1, \{(B(e',x'),\hat{W}(e',x'))\}_{(e',x')\in\mathcal{C}}\right].
\end{aligned}$$

Condition CT$'$ replaces Condition CT. This new condition states that, conditional on the realized estimate of the weighting function and the set of cells containing a positive number of deliveries, the counterfactual trend for event units is the same as the realized trend for non-event units.

Condition CT is not sufficient to ensure that the estimator has good finite-sample properties for two reasons. The first is that the birth-level marginal distribution in any finite sample may differ from the *population* marginal distribution used to state Condition CT, so Condition CT may not directly apply in finite samples; this will occur if providers' level and time pattern of volume is predictive of the potential outcomes for their associated deliveries.[2] The second is that, in finite samples, the desired weighting function $W(e,x)$ may be estimated with error. To the extent the error in $W(e,x)$ covaries with the deviation from common trends in any particular sample, bias can result. Once again, the most plausible source of such a correlation is correlation between providers'

[2]To see why this is the case, it may be helpful to consider a simple numerical example. Consider a cluster-sampling setting with two types of units. Type A units have 1 sub-unit and mean μ_A, while type B units have 9 sub-units and mean μ_B. Assuming both types of units are equally likely to be drawn, it easy to see that the population marginal mean is $(1/10)\mu_A + (9/10)\mu_B$. Under these assumptions, however, the mean of the sub-unit marginal distribution for a consisting of one provider, however, is $(1/2)\mu_A + (1/2)\mu_B$. As the number of sampled providers increases, the marginal distribution will converge toward the population marginal distribution.

level and time pattern of volume and the potential outcomes for their associated deliveries.

With Condition CT in place, the following lemma can be proved using arguments parallel to those used to prove Lemma 1.

Lemma 5. *If conditions NPE and CT′ hold, then*

$$\mathbb{E}_M[\hat{\Delta}^q - \hat{\Delta}^r \mid \{(B(e',x'), \hat{W}(e',x'))\}_{(e',x')\in\mathcal{C}}] = \sum_{(e,x)\in\mathcal{C}:B(e,x)=1} \hat{W}(e,x) \cdot \mathbb{E}_M\left[\frac{1}{N_1^q(e,x)} \sum_{(i,j)\in S_1^q(e,x)} Y_{ij}^q(1) - Y_{ij}^q(0) \,\middle|\, B(e,x)=1, \{(B(e',x'), \hat{W}(e',x'))\}_{(e',x')\in\mathcal{C}}\right]$$

for all M and all $q, r \in \mathcal{H}$ such that $q > 0$ and $r < 0$. In addition,

$$\mathbb{E}_M[\hat{\Delta}^q \mid \{(B(e',x'), \hat{W}(e',x'))\}_{(e',x')\in\mathcal{C}}] = \mathbb{E}_M[\hat{\Delta}^r \mid \{(B(e',x'), \hat{W}(e',x'))\}_{(e',x')\in\mathcal{C}}]$$

for all $q, r \in \mathcal{H}$ with $q, r < 0$.

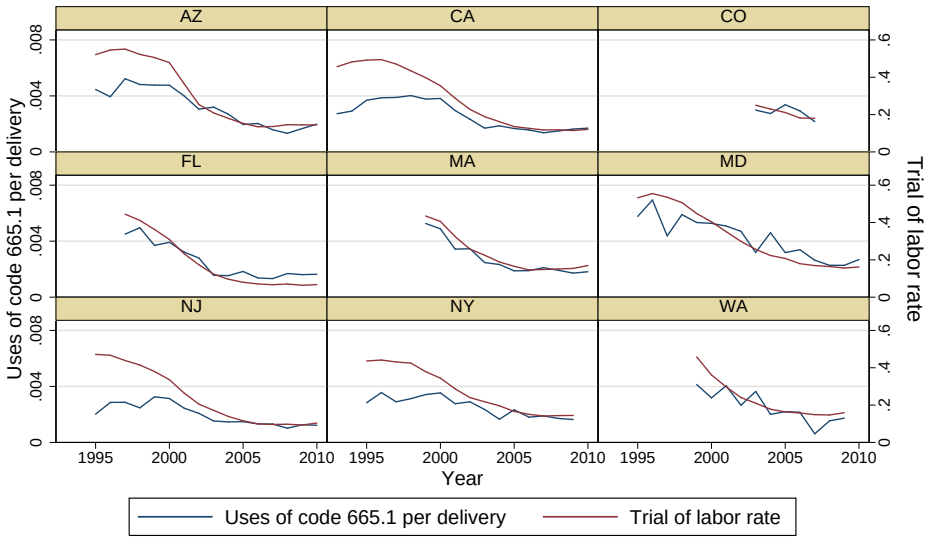
The first part of Lemma 5 demonstrates that the difference-in-difference estimator is (conditionally) unbiased for a weighted average of expected conditional sample average treatment effects on the treated.[3] This estimand is closely related to, but distinct from the estimand in Lemmas 1 and 2. In practice, the most important differences is that different (e,x) cells are weighted by the estimated version of $W(e,x)$, rather than the true version; for choices of weights like W_{WTOT}, the present estimand will therefore afford more weight to cells that happen to have a large number of "treated" births in the realized sample. The second part of Lemma 5 indicates that, as in the large-sample case, we should expect the difference between event and non-event units to be constant prior to event occurrence. Thus, as in the large-sample case, the research design permits a direct test of the required common trends assumption.

A.6 Modeling the use of ICD-9-CM code 665.1

As discussed in detail in Section 1.3, the ICD-9-CM diagnosis code 665.1 can be used either for uterine rupture (the event of interest) or for relatively minor uterine injuries that can occur during

[3]For this interpretation to be valid, it must also be the case that $\hat{W}(e,x) \geq 0$ for all $(e,x) \in \mathcal{C}$ and $\sum_{(e,x)\in\mathcal{C}:B(e,x)=1} \hat{W}(e,x) = 1$. Virtually any sensible estimator $\hat{W}(e,x)$ will satisfy these properties.

Figure A.1: Incidence of code 665.1 versus trial of labor rate by state and year



Notes: This figure plots the trial of labor rate and the rate of use of the ICD-9-CM diagnosis code 665.1 among women with a prior cesarean delivery in the SID and OSHPD discharge records. The rate of use of code 665.1 is plotted on the left y-axis. The trial of labor rate is plotted on the right y-axis.

cesarean section. Understanding the prevalence of the two uses of this code is important for interpreting the event study estimates. This appendix describes the procedure used to estimate the relative prevalence of the two uses of this code.

To do so, I exploit time series variation in trial of labor rates. Intuitively, if non-rupture uses of code 665.1 are rare, then the prevalence of code 665.1 should vary approximately proportionally with trial of labor rates since rupture is extremely rare among women who do not labor. In contrast, if non-rupture uses are common, then the prevalence of code 665.1 should fall less than proportionally when the trial of labor rate falls. Figure A.1 plots trial of labor rates and the prevalence of code 665.1 for each state included in the event study analysis; we see that the use of code 665.1 generally falls with trial of labor rates, but less than proportionally, suggesting that non-rupture uses of code 665.1 occur with some regularity.

A.6.1 Statistical model

In order to convert this qualitative time series evidence into an estimate of the share of uses of code 665.1 that are attributable to rupture and non-rupture events, I develop and estimate a statistical

model for the use of code 665.1.

For each delivery, I observe an indicator for cesarean delivery (C) and an indicator for the use of code 665.1 (R). I also observe an indicator L that is equal to one under the following two conditions: (1) the mother actually labored; or (2) the code 665.1 is used, but the mother did not labor. The indicator L reflects labor status as actually observed in my data since, as described in Section 1.3, use of code 665.1 are taken to indicate labor. Of interest but unobserved is the indicator variable R^*, which is equal to one if the code 665.1 is used for a true rupture event.

Each delivery proceeds as follows. A woman with a prior cesarean section in state s and quarter q attempts labor with probability ϕ_{sq}, and the remaining $1-\phi_{sq}$ undergo elective cesarean section. Women who attempt labor experience uterine rupture with some probability γ. The probability that an attempt of labor ends in cesarean section depends on the whether rupture has occurred and is given by $\alpha_r = \mathbb{P}(C = 1|R^* = r)$. A cesarean section before or after labor can lead to a non-rupture injury that is recorded using code 665.1, and the probability of such an injury is δ. Figure A.2 depicts the full probability tree corresponding to this model.

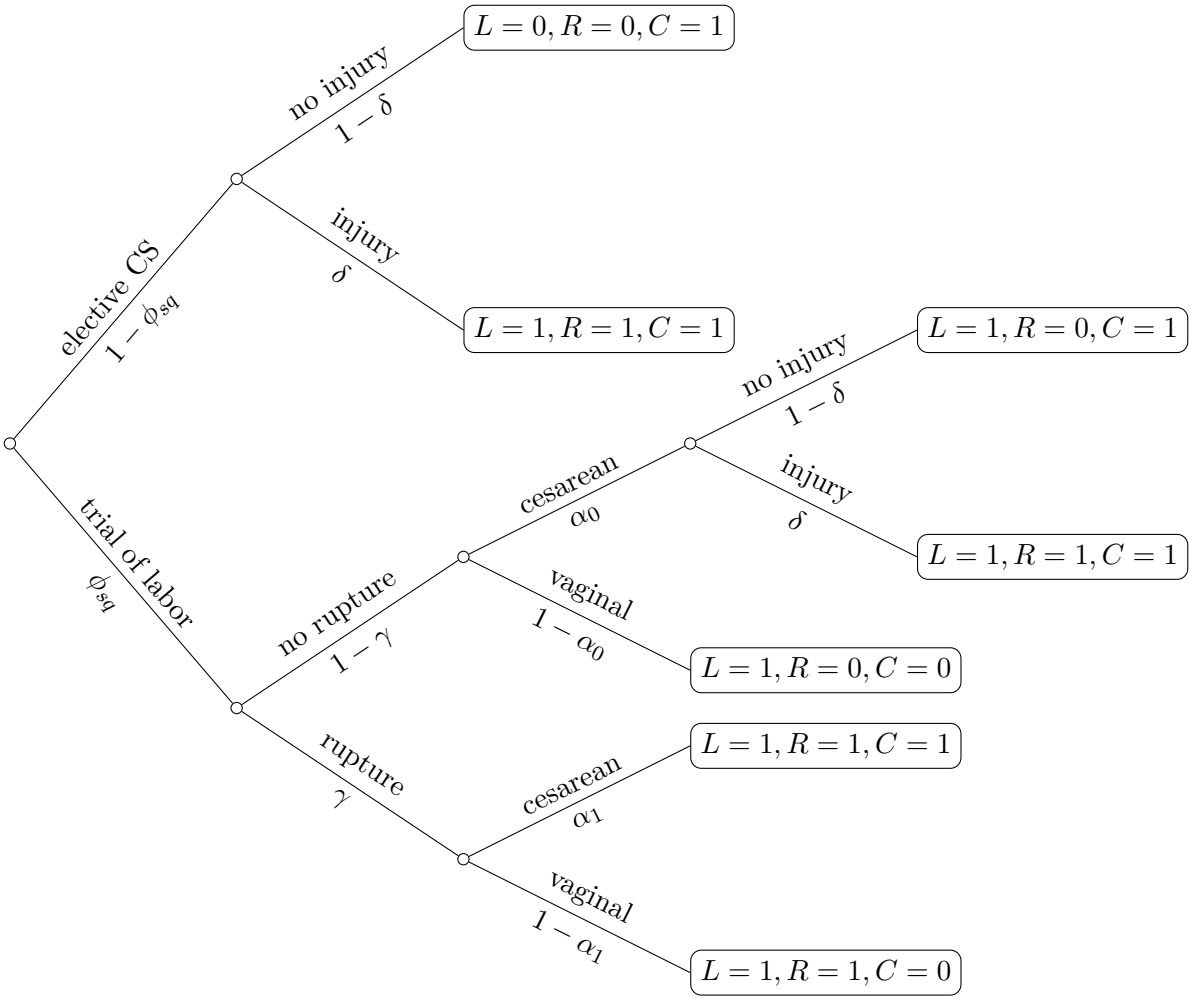
The object of interest is the "true positive rate," the share of uses of code 665.1 that are attributable to uterine rupture events. For a patient i in state s and quarter q, it is straightforward to show that the true positive rate can be written as

$$\begin{aligned} \text{TPR}_{sq} &\equiv \mathbb{P}(R^*_{isq} = 1|R_{isq} = 1) \\ &= \frac{\phi_{sq}\gamma}{\phi_{sq}\gamma + \phi_{sq}(1-\gamma)\alpha_0\delta + (1-\phi_{sq})\delta} \\ &\approx \frac{\mathbb{P}(L_{isq} = 1)\gamma}{\mathbb{P}(L_{isq} = 1)\gamma + \mathbb{P}(C_{isq} = 1)\delta}, \end{aligned} \tag{A.3}$$

where the approximate equality holds whenever γ and δ are small. As is intuitive, the true positive rate is higher when the true risk of rupture (γ) is large and when the probability of non-rupture uses of code 665.1 (δ) is small. Note that an estimate of the true positive rate for the full sample can be obtained by averaging TPR_{sq} over state-quarter cells, weighting each state-quarter cell by th enumber of uses of code 665.1.

I estimate the model by maximum likelihood. For patient i, the likelihood of the observed tuple

Figure A.2: Probability tree depicting model of use of ICD-9-CM diagnosis code 665.1



Notes: This figure depicts the model for the use of ICD-9-CM code 665.1 (uterine rupture during labor) that is described in Appendix A.6. The probability of traversing each branch is indicated underneath the branch. The leaf labels show the tuple (L, C, R) that will be observed for a delivery that reaches that leaf. The probability of reaching any particular leaf can be computed by multiplying the edge probabilities leading to that node. The probability of observing any tuple (L, C, R) can then be obtained by summing across all leaves displaying the relevant tuple.

Table A.2: Estimated parameters for model of use of ICD-9-CM code 665.1

Parameter	Estimate	SE
True uterine rupture rate (γ)	0.00711	0.00018
Probability of coded 665.1 injury given c-section (δ)	0.00096	0.00005
C-section probability given rupture (α_0)	0.905	0.005
C-section probability without rupture (α_1)	0.262	0.001
Auxiliary statistics		
Implied true positive rate for code 665.1	0.649	
N	2,982,279	

Notes: The table reports the estimated parameters from estimating the model defined in Appendix A.6 for the use of ICD-9-CM code 665.1 (uterine rupture during labor). The model is estimated by maximum likelihood, and the likelihood is defined equation (A.4). The true positive rate is defined in equation (A.3).

$(L_{isq}, C_{isq}, R_{isq})$ is then given by

$$\mathcal{L}(L_{isq}, C_{isq}, R_{isq}|\{\phi_{sq}\}, \gamma, \delta, \{\alpha_r\}) =$$
$$\begin{cases} (1-\phi_{sq})(1-\delta) & \text{if } L_{isq}=0,\ R_{isq}=0, \text{ and } C_{isq}=1 \\ \phi_{sq}\gamma\alpha_1 + \phi_{sq}(1-\gamma)\alpha_0\delta + (1-\phi_{sq})\delta & \text{if } L_{isq}=1,\ R_{isq}=1, \text{ and } C_{isq}=1 \\ \phi_{sq}(1-\gamma)\alpha_0(1-\delta) & \text{if } L_{isq}=1,\ R_{isq}=0, \text{ and } C_{isq}=1 \\ \phi_{sq}(1-\gamma)(1-\alpha_0) & \text{if } L_{isq}=1,\ R_{isq}=0, \text{ and } C_{isq}=0 \\ \phi_{sq}\gamma(1-\alpha_1) & \text{if } L_{isq}=1,\ R_{isq}=1, \text{ and } C_{isq}=0 \end{cases} . \quad \text{(A.4)}$$

Estimation is carried out in Stata using the default optimization settings. Estimation using appropriately collapsed data and frequency weights takes a few seconds.

A.6.2 Estimation results

Table A.2 reports the parameter estimates from estimating equation (A.4). The table also reports the true positive rate for the sample. Just under two thirds of cases in which code 665.1 is used appear to be true cases of uterine rupture.

The underlying parameters are precisely estimated and reasonable. The estimated risk of rupture during labor (γ) is approximately 7 in 1,000, which is well within the range of estimates obtained in Section 1.5 of this paper and within the 95 percent confidence interval of the meta-

analysis conducted by Guise et al. (2010b).

The estimated probability that cesarean delivery leads to a non-rupture use of code 665.1 (δ) is approximately 10 in 10,000. While no other estimates of this parameter exist in the literature, I do compare this estimate to alternative estimates of δ by examining the rate at which which code 665.1 appears on records for deliveries by women who: (1) are at low baseline risk of rupture; and (2) have a medical indication for elective cesarean section. Among women meeting these criteria, the risk of rupture should be virtually zero, meaning that all uses of code 665.1 will be for uterine injuries occurring during cesarean section. I examine cesarean deliveries among two such groups: (1) women with placenta previa; and (2) women delivering multiple infants. Code 665.1 appears on 7 in 10,000 records in the first group and 3 in 10,000 records in the second group. This suggests that the estimate obtained from estimating equation (A.4) modestly overstates δ and, thus, the true positive rate reported in Table A.2 may modestly understate the true positive rate.

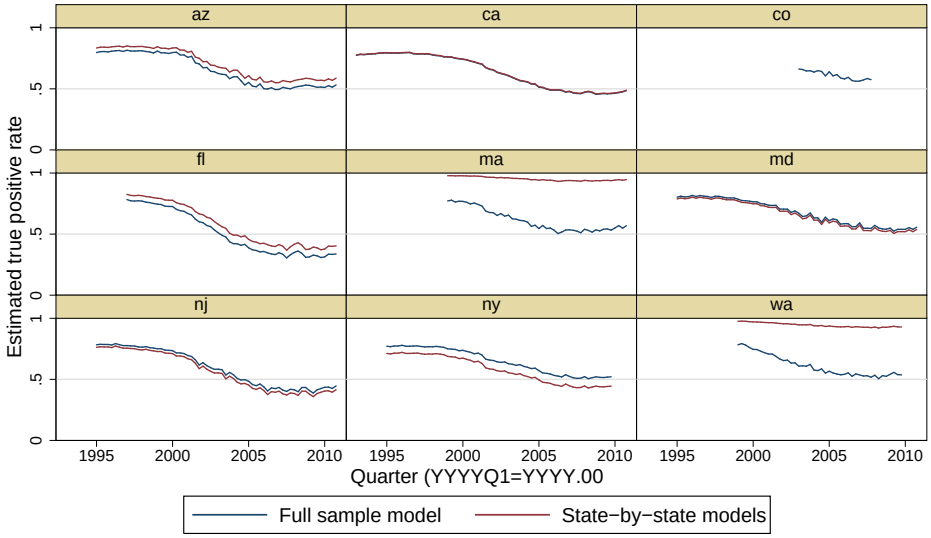
Figure A.3 plots estimated true positive rates by state and time. As expected, true positive rates fall over time as trial of labor and, hence, cases of true uterine rupture become less common. As a specification check, Figure A.3 also plots estimates obtained from fitting the model separately for each state, which allows for cross-state differences in uterine rupture rates and coding practices.[4] In general, the true positive rates estimated on state-specific samples are very similar to those estimated using the full sample. The exceptions are Massachusetts and Washington, where the available time series are relatively short and the model is correspondingly weakly identified.

A.7 Additional event study heterogeneity specifications

This appendix reports evidence on whether the size of the response appears to vary by patient characteristics (Tables A.3 and A.4) or event severity (Table A.5). For the patient characteristic analyses, I estimate equation (1.7) separately for each indicated patient population. For the event severity specifications, I categorize uterine rupture events into a "high severity" subgroup and a "low severity" subgroup and estimate equation (1.7) separately for each treatment subgroup; that is, I first estimate (1.7) excluding births associated with a "low severity" event and then estimate it excluding births associated with a "high severity" event. I define "high severity" uterine rupture

[4] I do not estimate the model for Colorado, as the available time series is too short to provide reliable estimates.

Figure A.3: Estimated true positive rates by state and time



Notes: This figure plots the true positive rates for the use of ICD-9-CM code 665.1 (uterine rupture during labor) obtained by estimating the model defined in Appendix A.6 for the full sample and then separately for each state. Additional details on estimation are provided in the text. The formula for the true positive rate appears in equation (A.3).

events as those cases in which the discharge record also reports transfusion, hysterectomy, or a maternal hospital stay in excess of five days. This definition is similar to that proposed by Callaghan et al. (2008) and encompasses approximately 17 percent of all rupture events. It is important to note, however, that because it is not possible to link maternal and neonatal records in these data, this definition does not include any information on adverse neonatal outcomes, which are typically the most serious outcomes in this context and need not coincide with serious maternal outcomes. For this reason, the event severity results should be taken with a significant grain of salt.

Table A.3: Heterogeneity by patient characteristics in the effect of a uterine rupture event on subsequent TOLAC

Column:	(1)	(2)	(3)	(4)	(5)	(6)
Subgroup category:	Mother's age		Mother's race		Mother's insurance	
Subgroup definition/ Post-event horizon	< 30	≥ 30	White	Non-white	Private	Other
+ 1 quarter	-0.0031^{+}	-0.0061***	-0.0055**	-0.0044**	-0.0039**	-0.0051*
	(0.0016)	(0.0018)	(0.0018)	(0.0017)	(0.0015)	(0.0021)
+ 2 quarters	-0.0018	-0.0041*	-0.0009	-0.0054**	-0.0032*	-0.0033
	(0.0014)	(0.0021)	(0.0019)	(0.0020)	(0.0015)	(0.0022)
+ 3 quarters	-0.0041**	-0.0077***	-0.0053*	-0.0065**	-0.0042*	-0.0080***
	(0.0016)	(0.0020)	(0.0021)	(0.0021)	(0.0016)	(0.0024)
+ 4 quarters	-0.0057**	-0.0080***	-0.0049*	-0.0091***	-0.0065***	-0.0082**
	(0.0019)	(0.0024)	(0.0025)	(0.0024)	(0.0018)	(0.0029)
+ 5 quarters	-0.0068***	-0.0067**	-0.0077**	-0.0074**	-0.0063***	-0.0087**
	(0.0020)	(0.0026)	(0.0028)	(0.0026)	(0.0019)	(0.0031)
+ 6 quarters	-0.0053*	-0.0091**	-0.0062^{+}	-0.0085***	-0.0053*	-0.0090**
	(0.0022)	(0.0028)	(0.0032)	(0.0024)	(0.0023)	(0.0029)
Pooled	-0.0045**	-0.0069***	-0.0051*	-0.0069***	-0.0049***	-0.0070**
	(0.0015)	(0.0019)	(0.0020)	(0.0019)	(0.0014)	(0.0023)
Auxiliary information						
Common trends p-value	0.170	0.464	0.093	0.500	0.526	0.420
Unique births	1,476,617	1,431,712	1,184,888	1,662,965	1,612,493	1,337,662
Nominal N	14,624,309	13,582,653	11,090,757	16,642,425	15,647,515	12,909,960

Notes: This table reports event study estimates of the effect of a uterine rupture event on hospitals' subsequent management of deliveries by women with a prior cesarean delivery. The reported estimates are obtained by estimating equation (1.7) separately for each listed patient group. Events are defined at the hospital level, and the outcome is trial of labor. Estimation uses HCUP State Inpatient Databases for the states and years described in the text and the OSHPD Patient Discharge Data for 1993-2010. The point estimate reported in the row labeled "$+q$ quarters" is defined as $\hat{\Delta}^q - \frac{1}{6}\sum_{r=-6}^{-1}\hat{\Delta}^r$; the results in the text show that this may be interpreted as the causal effect of an event q quarters later. The "pooled" estimate is defined as $\frac{1}{6}\sum_{q=1}^{6}\hat{\Delta}^q - \frac{1}{6}\sum_{q=-6}^{-1}\hat{\Delta}^q$; this quantity may be interpreted as the average causal effect over the first six post-event quarters. Standard errors are obtained via a block bootstrap at the hospital level using 200 replications and displayed in parentheses. The common trends p-value is obtained from a standard χ^2 test of the of the hypothesis that $\Delta^q = \Delta^r$ for all $q, r < 0$. Statistical significance is denoted as follows: $^{+}$ $p < .1$, * $p < .05$, ** $p < .01$, *** $p < .001$.

Table A.4: Heterogeneity by patient characteristics in the effect of a uterine rupture event on subsequent VBAC

Column:	(1)	(2)	(3)	(4)	(5)	(6)
Subgroup category:	Mother's age		Mother's race		Mother's insurance	
Subgroup definition/ Post-event horizon	< 30	≥ 30	White	Non-white	Private	Other
+ 1 quarter	-0.0020	-0.0063***	-0.0042*	-0.0044**	-0.0029*	-0.0057**
	(0.0014)	(0.0018)	(0.0019)	(0.0017)	(0.0014)	(0.0019)
+ 2 quarters	-0.0022	-0.0036^{+}	-0.0014	-0.0049*	-0.0024^{+}	-0.0041^{+}
	(0.0014)	(0.0020)	(0.0018)	(0.0020)	(0.0013)	(0.0022)
+ 3 quarters	-0.0040**	-0.0067**	-0.0050*	-0.0059**	-0.0042**	-0.0071**
	(0.0016)	(0.0021)	(0.0021)	(0.0021)	(0.0015)	(0.0025)
+ 4 quarters	-0.0046*	-0.0080**	-0.0041^{+}	-0.0083***	-0.0052**	-0.0083**
	(0.0018)	(0.0025)	(0.0024)	(0.0024)	(0.0016)	(0.0028)
+ 5 quarters	-0.0069***	-0.0087**	-0.0093**	-0.0078**	-0.0071***	-0.0098**
	(0.0021)	(0.0026)	(0.0028)	(0.0025)	(0.0019)	(0.0032)
+ 6 quarters	-0.0058**	-0.0105***	-0.0075*	-0.0088***	-0.0065**	-0.0097***
	(0.0021)	(0.0030)	(0.0034)	(0.0025)	(0.0023)	(0.0029)
Pooled	-0.0043**	-0.0073***	-0.0053**	-0.0067***	-0.0047***	-0.0074**
	(0.0015)	(0.0021)	(0.0020)	(0.0020)	(0.0014)	(0.0023)
Auxiliary information						
Common trends p-value	0.042	0.373	0.069	0.846	0.317	0.628
Unique births	1,476,617	1,431,712	1,184,888	1,662,965	1,612,493	1,337,662
Nominal N	14,624,309	13,582,653	11,090,757	16,642,425	15,647,515	12,909,960

Notes: This table reports event study estimates of the effect of a uterine rupture event on hospitals' subsequent management of deliveries by women with a prior cesarean delivery. The reported estimates are obtained by estimating equation (1.7) separately for each listed patient group. Events are defined at the hospital level, and the outcome is vaginal delivery. Estimation uses HCUP State Inpatient Databases for the states and years described in the text and the OSHPD Patient Discharge Data for 1993-2010. The point estimate reported in the row labeled "$+q$ quarters" is defined as $\hat{\Delta}^q - \frac{1}{6}\sum_{r=-6}^{-1}\hat{\Delta}^r$; the results in the text show that this may be interpreted as the causal effect of an event q quarters later. The "pooled" estimate is defined as $\frac{1}{6}\sum_{q=1}^{6}\hat{\Delta}^q - \frac{1}{6}\sum_{q=-6}^{-1}\hat{\Delta}^q$; this quantity may be interpreted as the average causal effect over the first six post-event quarters. Standard errors are obtained via a block bootstrap at the hospital level using 200 replications and displayed in parentheses. The common trends p-value is obtained from a standard χ^2 test of the of the hypothesis that $\Delta^q = \Delta^r$ for all $q, r < 0$. Statistical significance is denoted as follows: $^{+}$ $p < .1$, * $p < .05$, ** $p < .01$, *** $p < .001$.

Table A.5: Heterogeneity by event severity in the effect of a uterine rupture event on management of subsequent patients

Column:	(1)	(2)	(3)	(4)
Outcome:	Trial of labor		Vaginal delivery	
Event severity/ Post-event horizon	Low	High	Low	High
+ 1 quarter	-0.0042**	-0.0061*	-0.0041**	-0.0040
	(0.0014)	(0.0025)	(0.0015)	(0.0025)
+ 2 quarters	-0.0031*	-0.0032	-0.0032*	-0.0024
	(0.0016)	(0.0029)	(0.0016)	(0.0026)
+ 3 quarters	-0.0055***	-0.0068*	-0.0053**	-0.0055*
	(0.0016)	(0.0028)	(0.0017)	(0.0026)
+ 4 quarters	-0.0073***	-0.0057^{+}	-0.0068**	-0.0047^{+}
	(0.0019)	(0.0029)	(0.0022)	(0.0025)
+ 5 quarters	-0.0073***	-0.0059	-0.0081***	-0.0072*
	(0.0022)	(0.0037)	(0.0023)	(0.0033)
+ 6 quarters	-0.0079***	-0.0041	-0.0086***	-0.0052
	(0.0023)	(0.0036)	(0.0025)	(0.0032)
Pooled	-0.0059***	-0.0053*	-0.0060***	-0.0048*
	(0.0016)	(0.0026)	(0.0018)	(0.0024)
Auxiliary information				
Common trends p-value	0.611	0.012	0.935	0.026
Unique births	2,944,522	2,720,625	2,944,522	2,720,625
Nominal N	26,846,042	16,390,236	26,846,042	16,390,236

Notes: This table reports event study estimates of the effect of a uterine rupture event on hospitals' subsequent management of deliveries by women with a prior cesarean delivery. The reported estimates are obtained by estimating equation (1.7) separately for each event severity group, where the groups are as defined in Appendix A.7. Events are defined at the hospital level. Estimation uses HCUP State Inpatient Databases for the states and years described in the text and the OSHPD Patient Discharge Data for 1993-2010. The point estimate reported in the row labeled "$+q$ quarters" is defined as $\hat{\Delta}^q - \frac{1}{6}\sum_{r=-6}^{-1}\hat{\Delta}^r$; the results in the text show that this may be interpreted as the causal effect of an event q quarters later. The "pooled" estimate is defined as $\frac{1}{6}\sum_{q=1}^{6}\hat{\Delta}^q - \frac{1}{6}\sum_{q=-6}^{-1}\hat{\Delta}^q$; this quantity may be interpreted as the average causal effect over the first six post-event quarters. Standard errors are obtained via a block bootstrap at the hospital level using 200 replications and displayed in parentheses. The common trends p-value is obtained from a standard χ^2 test of the of the hypothesis that $\Delta^q = \Delta^r$ for all $q, r < 0$. Statistical significance is denoted as follows: $^{+}$ $p < .1$, * $p < .05$, ** $p < .01$, *** $p < .001$.

Appendix B

Appendix to Chapter 2

B.1 Data appendix

This appendix provides additional details on the processing of the MedPAR data used in this analysis. The first subsection of this appendix describes construction of the hospital universe, while the second subsection of this appendix describes the codes we used to identify particular diagnoses and procedures of interest.

B.1.1 Hospital universe

In order to construct the hospital universe for this analysis, we first limited the MedPAR file to acute care admissions to short-term hospitals. We identify short-term hospitals by matching the MedPAR file to the CMS Provider of Services (POS) file using the reported CMS provider numbers, which reports a variety of facility characteristics for registered Medicare institutional providers. We retain facilities identified in the POS file as short-term hospitals, critical access hospitals, or children's hospitals; we retain the last of these because many hospitals that are technically considered children's hospitals by Medicare serve a non-trivial number of adult patients.[1] We also exclude any admissions to specialized units of included facilities (e.g. long-term care units and psychiatric units) and "swing bed" skilled-nursing facility stays that occur in a critical access

[1]We also retain CMS provider numbers in the range 271225-271299. These provider numbers correspond to Montana medical assistance facilities, which were predecessors to critical access hospitals created as part of an early 1990s demonstration project. We exclude a small number of reported admissions to hospitals that do not participate in Medicare (typically military hospital and foreign hospitals).

hospital bed.[2]

After this initial processing step, we address the fact that the raw CMS provider numbers reported on the MedPAR file are not a full satisfactory panel identifier. Over the period observed, many hospitals convert to critical access hospital (CAH) status and consequently experience a change of provider number. In addition, many multi-campus hospitals switch from reporting under multiple provider numbers to report under a single provider number; mergers of truly unrelated hospitals create similar problems. In order to ensure that each "unit" in the panel analysis corresponds to a common set of physical facilities over time, we group provider numbers that are "related" in one of these ways under a single consolidated identifier. We identify related provider numbers using the provider number transition matrix created by the Dartmouth Institute for Health Policy and Clinical Practice.[3]

After consolidating hospital identifiers in this fashion, we exclude facilities that average less than 50 discharges per year during the quarters in which the hospital is active; the 406 facilities excluded by this restriction account for 0.03 percent of the total admissions reported in the MedPAR file. The resulting hospital universe still includes some facilities that, notwithstanding their categorization in the POS file, are primarily specialty hospitals (e.g. cancer hospitals and specialty surgical hospitals) or long-term care hospitals and thus should not be included in the present analysis. To exclude hospitals that do not primarily provide acute care, we exclude any hospital in which fewer than 5 percent of admissions report a primary diagnosis falling in a specified set of acute cardiac diagnoses, which is described in greater detail in the next subsection of this appendix; this restriction excludes 280 hospitals that account for just approximately 0.6 percent of all admissions. We also exclude any hospital with a length of stay at least 7 days above the sample average after adjusting for the quarters in which the hospital operates; this restriction excludes an additional 94 hospitals that account for 0.01 percent of all admissions.[4] We are left with 5,156 hospitals, which

[2]Admission to one of these specialized units is indicated by a one-character "special unit" code reported on the MedPAR file.

[3]The transition matrix is available for download at http://www.intensity.dartmouth.edu/?q=node/129.

[4]The cross-hospital distribution of both acute cardiac share and length of stay consists of two largely-isolated modes, with the larger mode containing the desired hospitals and the smaller mode containing various types of specialty hospitals. The 5 percent threshold and the 7 day threshold were chosen so as to cleanly separate the two modes in each case.

is reduced further to the 5,094 hospitals reported in Table 2.1 because 62 hospitals start operation after 2005Q2.

B.1.2 Diagnosis and procedure coding

We identify AMI, CABG, and PCI discharges using raw ICD-9-CM procedure codes. The codes and criteria used are as follows:

- **AMI discharges:** The primary diagnosis code has the form 410.XX.
- **PCI discharges:** Any procedure code field contains 36.01, 36.02, 36.05, 36.09, 36.00, 36.0, or 00.66.
- **CABG discharges:** Any procedure code field contains a code of the form 36.1X.

We identify other categories of discharges by mapping the raw ICD-9-CM codes to the multi-level Clinical Classification Software codes defined by the Healthcare Cost and Utilization Project. The codes and criteria used are as follows:

- **Acute cardiovascular discharges:** The primary diagnosis code falls in CCS categories 7.1, 7.2.3, 7.2.5, 7.2.6, 7.2.9, 7.2.11, 7.3.1, or 7.3.4. These particular codes were chosen on the basis of a preliminary analysis in Nationwide Inpatient Sample data indicating that patients with these diagnoses were very likely to have been admitted on an emergency basis. As described in the prior subsection of this appendix, this category is used in defining the hospital universe.
- **Cardiovascular discharges:** The primary diagnosis code falls in CCS category 7. Tallies of discharges falling in this category are reported in Table 2.1.

Appendix C

Appendix to Chapter 3

C.1 Estimation with high-dimensional fixed effects

As described in the main text, I estimate equation (3.2), which contains multiple high-dimensional fixed effects, by applying an algorithm due to Gaure (2013). This appendix provides a brief discussion of why naive approaches fail and then a brief exposition of the algorithm, portions of which are adapted from Gaure (2013). For a full treatment, see Gaure (2013).

Consider a general fixed-effects regression model written in dummy variable form:

$$Y = X\beta + \sum_{g=1}^{G} D_g \gamma_g + \epsilon. \tag{C.1}$$

The vector Y contains the values of the dependent variable for each observation, X is the matrix of covariates, and $D_1, D_2, \ldots, D_G$ are matrices coding the categorical dummy variables corresponding to each set of fixed effects. The coefficient vector β is the estimand of interest. In the present context, the fixed effects themselves γ_g are nuisance parameters that are not of independent interest, although this need not always be the case (e.g. Abowd et al. (1999) and Carneiro et al. (2012)).

Direct approaches to computing the least squares estimates of β are infeasible in this context. To see why, write $D = [D_1\ D_2\ \cdots D_G]$ and $\gamma' = [\gamma_1'\ \gamma_2'\ \cdots \gamma_G']$. The standard approach to obtaining the full least squares estimates is to numerically solve the so-called normal equations

$$\left([X\ D]'[X\ D]\right) \begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} = [X\ D]'Y \tag{C.2}$$

for the least squares estimates $(\hat{\beta}' \ \hat{\gamma}')$.[1] This system has dimension $m = k + \sum_g d_g$, where k is the number of covariates included in X and d_g is the number of categories in the gth set of fixed effects. In the present context, m is greater than 3500. Directly solving a linear system of this size using general-purpose solvers is prohibitively expensive.

To reduce the dimensionality of the problem, it is common to instead consider the projected regression

$$M_D Y = M_D X\beta + M_D\epsilon, \tag{C.3}$$

where $M_D = I - D(D'D)^{-1}D'$ is the matrix that projects vectors onto the space orthogonal to the column space of D.[2] By the Frisch-Waugh-Lovell theorem, the projected regression in (C.3) yields the same estimates of β and the same residuals as the full regression in (C.1). Similarly, after a degrees-of-freedom correction that accounts for the estimation of the fixed effects (discussed in more detail below), the regression generates correct standard errors. Furthermore, provided that $M_D Y$ and $M_D X$ are computationally inexpensive to obtain, this approach can be very fast, since the projected regression is only k-dimensional and, thus, easy to compute.

The ease with which one can obtain $M_D Y$ and $M_D X$ depends upon the number of sets of fixed effects G. When there is only one set of fixed effects ($G = 1$), then M_D is just the familiar "de-meaning" operator that subtracts the group mean from each observation, so computing $M_D Y$ and $M_D X$ is trivial; this is the standard "within estimator" for unobserved effects models (see, for example, Wooldridge (2010)). When $G > 1$, however, M_D does not have a simple form, and obtaining $M_D Y$ and $M_D X$ is more difficult.

To solve this problem, Gaure (2013) turns to a result of Halperin (1962) that shows that as $n \to \infty$

$$(M_{D_G} M_{D_{G-1}} \cdots M_{D_2} M_{D_1})^n \to M_D,$$

where $M_{D_g} = I - D_g(D_g'D_g)^{-1}D_g'$ is the matrix that projects onto the space orthogonal to D_g;

[1] In order to focus on the matter at hand, I am glossing over the fact that the matrix D does not, in general, have full rank and, thus, there is not a unique vector $\hat{\gamma}$ that satisfies equation (C.2). This can be remedied by normalizing a suitable number of fixed effects to zero and deleting the corresponding columns from D. In most applications, these normalizations will not appreciably reduce the dimension of the system. Note also that, regardless, there is a unique vector $\hat{\beta}$ that satisfies the system provided that $X'M_D X$ has full rank, where M_D is the matrix that projects onto the orthogonal complement of D.

[2] Again, I am glossing over the fact that, when $G > 1$, D does not have full rank. In this case, $M_D = I - D(D'D)^{+}D'$, where A^{+} denotes the Moore-Penrose pseudoinverse of A.

convergence here is strong convergence. Gaure cites results of Deutsch and Hundal (1997) that demonstrate that this convergence is uniform, occurs at a geometric rate, and will be particularly fast when the column spaces of the matrices D_g are close to orthogonal.

Based on these results and the fact that each matrix M_{D_g} is just the demeaning operator corresponding to the gth set of fixed effects, Gaure suggests a simple algorithm for computing $M_D X$ and $M_D Y$. Specifically, starting with a vector v, we demean with respect to the first set of fixed effects, then demean the result with respect to the second set of fixed effects, and continue in this fashion through the Gth set. At this point, we compare the current vector in memory to the vector in memory after the last full round of demeaning. If the change is acceptably small, we stop and use the current vector as our approximation of $M_D v$. Otherwise, we undertake another round of demeaning.

This algorithm has the key advantage of being exceptionally simple to code. Furthermore, when implemented in Stata and applied to the present problem, it is modestly faster than both the accelerated Gauss-Seidel algorithm proposed by Guimarães and Portugal (2010) and an adapted version of the method proposed by Abowd et al. (2002).[3] An algorithm based on Abowd et al. (2002) might be faster than the present algorithm if implemented in a language with facilities for handling sparse matrices (which Stata lacks), but I have not explored this possibility.

One implementation detail remains to be addressed. As noted above, obtaining correct standard errors from estimating equation (C.3) requires adjusting the regression degrees of freedom to account for estimation of the fixed effects. The appropriate adjustment reduces the degrees of freedom by the rank of D. When $G = 1$, D has full rank, so determining the appropriate adjustment is trivial. However, as noted in footnotes 1 and 2, when $G > 1$, D does not have full rank, so determining the appropriate adjustment is harder. For the case $G = 2$, Abowd et al. (2002) have developed a general algorithm for computing the appropriate adjustment, but, to my knowledge, no general algorithm exists for the case $G > 2$.

In the present application, however, the structure of the fixed effects makes it easy to directly

[3]The original algorithm proposed by Abowd et al. (2002) use a conjugate gradient algorithm that exploits the sparseness of the matrix $[X\ D]'[X\ D]$ in order to efficiently solve equation (C.2). This approach does not provide an easy way of obtaining standard errors for β, so it is not suitable for the present problem. However, one can instead use a similar conjugate gradient algorithm to obtain $M_D X$ and $M_D Y$ and then estimate equation (C.3) as described above.

calculate the rank of D. Recall that the main empirical results were obtained from a regression of the following form:

$$Y_{iht} = \alpha + \tilde{N}_{ht}^{\text{ICU}}\beta_{\text{ICU}} + \tilde{N}_{ht}^{\text{OIP}}\beta_{\text{OIP}} + X_{iht}\gamma + \psi_t + \sum_{\tau \in S} \phi^{\tau}_{h \times \tau(t)} + \epsilon_{iht},$$

where the notation is as defined in the main text.

To calculate the rank of D, let W be the space spanned by the dummies corresponding to the hospital-time effects, and let V be the space spanned by the dummies corresponding the calendar-date effects. We wish to find $\dim(W + V)$. It is a standard result in linear algebra that

$$\dim(W + V) = \dim(W) + \dim(V) - \dim(W \cap V).$$

We compute each quantity in turn:

- $\dim(W)$: For any given hospital h, it is easy to see the we can obtain a linearly independent set from the dummies for that hospital by dropping one column for each set of hospital-time effects after the first. Letting H denote the total number of hospitals, we therefore conclude that $\dim(W)$ is the total number of hospital-time effects minus $H(|S| - 1)$.

- $\dim(V)$: The date dummies are mutually orthogonal, so this space has dimension equal to the total number dates for which observations exist in the dataset.

- $\dim(W \cap V)$: The space $W \cap V$ is spanned by a set that contains the full set of time dummies corresponding to each $\tau \in S$. We obtain a linearly independent set from these dummies by removing one dummy corresponding to each element of S beyond the first. Thus, we see that this space has dimension $\sum_{\tau \in S} |\{\tau(t) : t \in T\}| - (|S| - 1)$, where T is the set of time indexes appearing in the dataset.

To summarize, the rank of D is the width of D minus

$$H(|S| - 1) + [\sum_{\tau \in S} |\{\tau(t) : t \in T\}| - (|S| - 1)] = \sum_{\tau \in S} |\{\tau(t) : t \in T\}| + (H - 1)(|S| - 1).$$

C.2 Coding appendix

C.2.1 Diagnosis categories for analysis sample definition

The ICD-9-CM diagnosis codes I use to define the diagnosis categories of interest are as follows:

- **Acute myocardial infarction (AMI):** Patients are classified as having a diagnosis of AMI if the primary diagnosis code has the form 410.xx. The subgroup of AMI patients who are experiencing an ST-elevation MI (STEMI) are identified by presence of a digit other "7" in the fourth position of the code.

- **Unstable angina:** Patients are classified as having a diagnosis of unstable angina if either of the following holds: (1) the primary diagnosis code has the form 411.1x or 411.8x; or (2) the primary diagnosis code has the form 414.xx and the record reports a secondary diagnosis code of the form 411.1x or 411.8x.

- **Other acute coronary artery disease:** Patients are classified as having a diagnosis of other acute coronary artery disease if either: (1) the patient has a primary diagnosis code of the form 414.xx and does not have unstable angina as defined above; or (2) the patient has a primary diagnosis code of the form 413.xx (stable angina).

Records reporting a primary diagnosis code of 411.1x, 411.8x, or 413.xx are technically coded incorrectly, as these codes correspond to symptoms rather than underlying medical conditions and, thus, should not be used as primary diagnoses (Smith, 2011). Use of such codes as primary diagnoses seems reasonably common in practice, however, so excluding such records would fail to capture a large number of records that should be included.

ICD-9-CM codes of the form 414.xx are frequently reported as the primary diagnosis for scheduled hospital admissions for PCI or bypass surgery. For obvious reasons, I do not wish to include claims of this type in my cohort of arriving patients. Thus, patients who are classified as having unstable angina or other acute coronary artery disease on the basis of a primary diagnosis code of the form 414.xx are only included in the analysis cohort if the patient's record also demonstrates interaction with the hospital's emergency department.

I classify a record as demonstrating emergency department interaction if one of the following applies: (1) it is an emergency department record; (2) it is an inpatient record and reports a billing

code for emergency services (see the subsection below on billing codes); or (3) it is an outpatient observation record and the "ED flag" provided by the hospital is set. I impose an exception to criterion (2) for one small hospital that does not appear to consistently report billing codes for emergency department services; in this case, I use the hospital-reported "ED flag" to identify emergency department interaction.

C.2.2 Main procedure outcomes

I identify the main procedures of interest, cardiac catheterization and PCI, using both ICD-9-CM procedures codes and, on records that report them, CPT codes. The ICD-9-CM codes used are as follows:

- **Cardiac catheterization:** Patients are classified as having received cardiac catheterization if at least one of the codes 37.22, 37.23, 88.55, 88.56, or 88.57 is present in any procedure code field.
- **PCI:** Patients are classified as having received PCI if at least one of the code 36.01, 36.02, 36.05, 36.06, 36.07, or 00.66 is present in any procedure code field.

I identify the list of CPT codes associated with each of these procedures from existing data sources. Specifically, I start with the Clinical Classification Software (CCS) groupings of CPT codes available from the Healthcare Cost and Utilization Project.[4] CPT codes falling in groups 46 and 47 correspond to cardiac catheterization, while CPT codes falling in group 45 correspond to PCI. Unfortunately, the CCS file only reports ranges of codes, rather than individual codes. I convert the code ranges to lists of individual codes by matching the ranges to the Medicare Physician Fee Schedule for each year 2002-2009, which provides an exhaustive list of the CPT codes that are active in any particular year. The resulting list is available upon request.

Any patient undergoing PCI is coded as undergoing cardiac catheterization as well.

[4]The CCS groupings are available in machine-readable form from the HCUP website at http://www.hcup-us.ahrq.gov/toolssoftware/ccs_svcsproc/ccssvcproc.jsp.

C.2.3 Processing UB-92/UB-04 codes

I use UB-92/UB-04 codes to identify inpatient stays that include an ICU admission or that included emergency department services. I classify an inpatient stay as including an ICU stay if it reports a UB-92/UB-04 revenue code indicating a stay in a medical/surgical ICU (0200), a burn unit (0207), a coronary care unit (0210), a pulmonary coronary care unit (0212), or an ICU bed of type "other" (0209). I classify an inpatient stay as including emergency department services if it reports the revenue code "0450."

C.2.4 CCS categories used in balance tests

For the balance tests in Section 3.4, I use diagnosis and procedure categories defined using the Healthcare Cost and Utilization Project's multi-level Clinical Classification Software for ICD-9-CM.[5]

The definitions of the diagnosis categories used are as follows: diabetes (CCS code 3.2 or 3.3); hypertension (CCS code 7.1); chronic obstructive pulmonary disease (CCS code 8.2). Because these categories are used for comorbidity balance checks, the indicator corresponding to each of these categories is equal to one if any secondary diagnosis field contains an ICD-9-CM code falling in the relevant CCS category.

The definitions of the procedure categories used are as follows: spinal procedures (CCS code 1.3 or 14.11); cholecystectomy (CCS code 9.16.1 or 9.16.2); hysterectomy (12.5); and hip or knee procedure (CCS code 14.7.1 or 14.7.2). An inpatient stay is included in these procedure categories if any procedure field contains an ICD-9-CM code falling into the relevant CCS category.

[5]The CCS definitions are available in machine-readable form from the HCUP website at http://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp.