

# ESTIMATING HETEROGENEITY IN THE BENEFITS OF MEDICAL TREATMENT INTENSITY

William N. Evans and Craig Garthwaite\*

*Abstract*—We exploit increases in postpartum length of stay generated by legislative changes in the late 1990s to identify the impact of greater hospital care on the health of newborns. Using all births in California over the 1995–2000 period, two-stage least-square estimates show that increased treatment intensity had a modest impact on readmission probabilities for the average newborn. Allowing the treatment effect to vary by two objective measures of medical need demonstrates that the law had large impacts for those with the greatest likelihood of a readmission. The results suggest that the returns to average and marginal patients vary considerably in this context.

## I. Introduction

IT is not difficult to generate a case that the marginal product of medical care spending in the United States is very low. In 2006, per capita spending on health in the United States was \$6,714, more than twice the median value for OECD countries, about twice the value of Canada, and nearly 2.5 times the value of the United Kingdom. Despite this spending, in 2005, the United States ranked twenty-fifth of 29 countries in average life expectancy and had the fourth highest infant mortality rate of 28 reporting countries.<sup>1</sup> The Dartmouth Atlas Project<sup>2</sup> shows that per capita Medicare reimbursements across hospital referral regions vary by a factor of three (Wennberg et al., 2008), but there is little evidence that the differences in spending lead to better quality of care (Baicker & Chandra, 2004) or better mortality outcomes (Fisher et al., 2003). In some instances, changes in insurance status, which are associated with large changes in care, typically have little impact on mortality. Finkelstein and McKnight (2008) found that the introduction of Medicare dramatically increased health insurance coverage and health care spending of the elderly, but it had no discernable impact on elderly mortality through the first ten years of the program. Data from the RAND Health Insurance Experiment showed that a reduction in copayments increases health care use but has no detectable impact on health outcomes (Newhouse & the Insurance Experiment Group, 1993).<sup>3</sup>

Received for publication July 23, 2009. Revision accepted for publication January 6, 2011.

\* Evans: Department of Economics, University of Notre Dame; Garthwaite: Department of Management and Strategy, Kellogg School of Management, Northwestern University.

We thank the employees at the California Office of Statewide Health Planning and Development for their assistance with the data, Jonah Gelbach and Amitabh Chandra for a number of helpful discussions, and seminar participants at the University of Chicago and the University of Wisconsin for helpful comments. The work on this project was supported by a grant from the National Institutes of Health, R01HD05289.

<sup>1</sup> All data are from the OECD's frequently requested data series, [http://www.irdes.fr/EcoSante/Download/OECDHealthData\\_FrequentlyRequestedData.xls](http://www.irdes.fr/EcoSante/Download/OECDHealthData_FrequentlyRequestedData.xls).

<sup>2</sup> <http://www.dartmouthatlas.org/>.

<sup>3</sup> In contrast to these results, Card, Dobkin, and Maestas (2009) demonstrate that at age 65, insurance rates increase dramatically because of the Medicare program. For patients admitted to the hospital through the emergency room, Medicare enrollment is estimated to reduce death rates for these severely ill patients by 20%.

The weak relationship between aggregate spending and health outcomes is in stark contrast to evidence showing pronounced medical benefits for the use of specific medical devices, procedures, or pharmaceuticals. Advances in the treatment of heart attacks reduced the one-year mortality rate for these patients by 5 percentage points between 1984 and 1991 (Cutler et al., 1998). The use of antiretroviral drugs among HIV/AIDS patients is associated with approximately a 70% drop in mortality (Hammer et al., 1997; Schwarcz et al., 2000; Duggan & Evans, 2008), while the use of statin drugs to reduce cholesterol is associated with a 12% reduction in all-cause mortality (Law, Wald, & Rudnicka, 2003; Cholesterol Treatment Trialists' Collaborative, 2005). Between 1950 and 1990, treatment options for low-birthweight infants expanded greatly, generating a 72% reduction in infant mortality rates over the same period (Cutler & Meara, 2000).

The disparity in the two sets of results we have outlined may not be all that surprising. Studies relying on aggregate data suggest that for the average patient, the marginal productivity of additional care might be low, but as the patient's condition becomes more sharply defined, results may not be so average. This difference could be driven by underlying heterogeneity in the benefits of greater medical treatment intensity. Medical researchers have long discussed the potential effects of this treatment heterogeneity. Kravitz, Duan, and Braslow (2004) note that "when HTE [heterogeneity of treatment effects] is present, the modest benefit ascribed to many treatments in clinical trials can be misleading because modest average effects may reflect a mixture of substantial benefits for some, little benefit for many, and harm for a few." The heterogeneity in benefits of treatment intensity has led to a greater emphasis on reducing waste and improving the quality of clinical decisions as cornerstones of any health care reform initiative. For example, \$1.1 billion was earmarked for comparative-effectiveness research as part of the American Recovery and Reinvestment Act, signed into law on February 19, 2009.<sup>4</sup> This research is designed to establish treatment protocols for specific conditions with the goal of reducing medical spending on cases where the marginal benefits of treatment are thought to be low.

In this paper, we use an exogenous increase in hospital care for newborns generated by legislation to examine the heterogeneity in the benefits of greater medical care use. In response to growing concerns about the declining length of hospital care following childbirth, 43 states and the federal government passed laws during the mid- to late 1990s

<sup>4</sup> PL 111-5, <http://www.gpo.gov/fdsys/pkg/PLAW-111publ5/pdf/PLAW-111publ5.pdf>.

requiring insurance carriers to provide minimum length of postpartum stays (Eaton, 2001). Previous research has documented that these laws greatly increased the average length of a postpartum hospital stay (Udom & Betley, 1998; Liu, Dow, & Norton, 2004; Madlon-Kay & DeFor, 2005; Evans, Garthwaite, & Wei, 2008), but there is conflicting evidence about whether these laws improved infant health (Madden et al., 2004; Meara et al., 2004; Datar & Sood, 2006; Evans, Garthwaite, & Wei, 2008; Almond & Doyle, 2011). The question of interest is whether newborns discharged early from a postpartum stay (defined as staying fewer than the recommended number of nights in the hospital) have improved infant health (as measured by hospital readmissions).

Initially we estimate a two-stage least-square (2SLS) model for all births using the legislation-induced increase in postpartum stays as an exogenous decline in the fraction of children discharged early. At first glance, this appears to be the ideal statistical environment for a 2SLS procedure. The laws reduce early discharges by about 25 percentage points, and our analysis sample has nearly 2.5 million observations. However, the local average treatment effects (LATE) suggest a small and marginally statistically significant impact of early discharges on readmission rates. But the benefits are not uniform across newborns, and therefore the LATE for the entire population may not be the relevant statistic.

Therefore, we estimate how the benefits of longer hospital stays vary with the observed characteristics of the newborn. Initially we use data on reported labor and delivery complications to estimate heterogeneity in the benefits of longer hospital stays. We find that individuals with any of these complications benefited from increased length of stay, and those with uncomplicated deliveries received no benefit. Unfortunately, the incidence rates for particular complications are very low in some cases, making power an issue in isolating who benefits from greater treatment.

As a second method to measure treatment heterogeneity, we use a procedure suggested by Baicker, Buckles, and Chandra (2006) and Chandra and Staiger (2007) and expanded on in detail here. Using prelaw data on mother's demographic characteristics, baby's birth weight and gestation, delivery complications, and other observable medical conditions, we estimate the propensity score for a baby being released "early" from the hospital. We next use these estimates to predict the propensity score in the post-federal law period. Given the large sample sizes and the detailed characteristics of the mother, pregnancy, and delivery, the propensity score of an early discharge spans the entire  $[0,1]$  interval. The distribution of the propensity score in the pre- and postlaw period is also nearly identical. More importantly, the propensity score of an early discharge provides substantial evidence of the medical risk of a readmission. The propensity score is monotonically increasing in birth weight and gestation and monotonically declining in the average number of complications. The readmission rate declines nearly monotonically, with the propensity score

suggesting that from a medical standpoint, those most likely to have a longer stay have the greatest risk of a readmission. This evidence suggests that the propensity score can be thought of as an index of the medical appropriateness of a longer length of stay in the hospital.

We then estimate the basic 2SLS model by thirds of the propensity score. These results show that those newborns with the lowest probability of being released early—often those with particular conditions or complications—received the largest benefit from the increased length of stay. The vast majority of individuals, however, received no benefit from the increased length of stay generated by the law.

Our results are instructive. Identifying and understanding the source of heterogeneity in the benefits of medical treatments is essential for developing effective health policy. Rather than relying on the estimated marginal benefit for the average patient, effective and targeted policy should focus on the benefits to patients for whom the treatment or intervention is most appropriate. Failing to consider this heterogeneity in benefits could lead to erroneous conclusions about the efficacy of different policy initiatives. In our concrete example, the estimates suggest that the same medical benefits of early discharge laws can be achieved by applying the law only to those most in need of care rather than applying the legislation to all births.

## II. Minimum Postpartum Stay Legislation and Previous Literature

Beginning in the 1970s and continuing until the mid-1990s, managed care cost-cutting efforts and attempts to "demedicalize" the childbirth process led to dramatic reduction in postpartum hospital stays for mothers and newborns (Braveman et al., 1995; Eaton, 2001). The average length of postpartum stay for vaginal births fell from 3.9 days in 1970 to 1.8 days in 1996. Over the same time period, the length of stay for cesarean section (c-section) births fell from 7.8 to 3.5 days. By 1995, one-day postpartum stays accounted for 47% of all vaginal deliveries compared to just 7.6% in 1980 (Hyman, 1999).

This change in the provision of medical services eventually drew popular press attention, with numerous articles using the phrases "drive-through" and "drive-by" deliveries to describe short postpartum hospital stays. News stories and articles in the medical literature detailed tragic stories of preventable conditions afflicting mothers and newborns who were swiftly discharged from the hospital following the delivery. State lawmakers responded to these concerns with a series of laws designed to increase the length of postpartum hospital stays. Between 1995 and 1997, 43 states mandated some length of minimum postpartum stay (Eaton, 2001). California, the state of interest in this paper, passed a minimum postpartum law on August 11, 1997, and the law went into effect the same day.

Despite nearly every state passing such mandates, the preemption clause of the Employment Retirement Income

Security Act (ERISA) left a large number of women (primarily those covered by multistate and self-funded employer-provided insurance plans) unaffected by these state-level mandates. As a result, in a bipartisan effort epitomized by a 98–0 vote in the U.S. Senate, federal lawmakers passed the Newborns and Mothers Health Protection Act of 1996. This law, which went into effect on January 1, 1998, required a minimum length of stay of 48 hours following a vaginal delivery and 96 hours following a c-section. Discharges for a stay of shorter duration required the consent of both doctor and patient.

While the federal law increased the number of women covered by these mandates, it still did not apply to all women. The federal statute and most state laws exempted most Medicaid patients. In California, patients in private Medicaid managed care plans, a rapidly growing component of the Medicaid population, were covered by both the California and federal minimum-stay laws.<sup>5</sup> Following implementation of the federal and state statutes, lawmakers in California moved to end the unequal treatment of Medicaid patients. On August 26, 1998, the legislature passed AB 1397, which extended the California minimum postpartum stay law to all Medicaid recipients regardless of the source of insurance. The effective day for this law was January 1, 1999.

As a result of these changes, new mothers in California were subject to four different legislative environments. Prior to August 1997, there were no restrictions on what insurance companies were required to reimburse for postpartum hospital stays. Between August 1997 and the end of that year, women with private insurance and Medicaid patients in private managed care plans were subject to the state law, which was superseded by the federal statute on January 1, 1998, and expanded to cover insurance plans regulated by ERISA. Finally, the state law was expanded to include all Medicaid patients on January 1, 1999. At this point, all insured births in California were covered by a mandated minimum length of stay.

It is clear that these laws achieved their goal of increasing the average length of postpartum stays. Udom and Betley (1998), Liu et al. (2004), and Madlon-Kay and DeFor (2005) found that the passage of the laws lengthened stays and increased medical spending. The effect of these laws on health outcomes, however, has been less clear. Madden et al. (2002, 2004) found little impact of different length of stays on patients in Massachusetts. Using data for 150,000 newborns in Ohio, Meara et al. (2004) focused on the effect of that state's mandated postpartum stay law on Medicaid patients, who were explicitly covered by the state-level mandate. They found that the law generated a statistically significant decrease in the number of early dis-

charges and large but statistically insignificant decreases in the number of readmissions.<sup>6</sup>

Both Datar and Sood (2006) and Evans et al. (2008) used data on the universe of hospital births in California from 1995 to 2000 to analyze the effect of mandated increases in postpartum stays on infant health. Datar and Sood (2006) found large and statistically significant decreases in the probability of readmission for infants as a result of the mandates, but Evans et al. (2008) document this finding as largely a result of data limitations in the publicly available data. Using a richer data set than that employed by Datar and Sood, this later study found smaller and statistically insignificant results. Evans et al. found, however, that there was heterogeneity in the estimated impacts of the law. The law had little impact on newborns from uncomplicated vaginal deliveries, but the increased hospital stays generated statistically significant reductions in newborn readmission for privately insured c-section patients and for Medicaid-insured vaginal deliveries.

Almond and Doyle (2011) used the exact timing of birth as the basis of a regression discontinuity identification strategy to estimate the effect of longer postpartum stays on infant health in California. The authors exploit a billing rule that causes babies born just after midnight to receive another day of reimbursable care compared to those born just before midnight. Furthermore, the authors used the fact that the California mandate further increased the length of stay for infants born just after midnight (in this case, from two days of care for those just before midnight to three days of care for those after midnight). Using this strategy, the authors find no effect of longer hospital stays on infant health for uncomplicated deliveries.

Despite the large number of studies, there is little consensus on the benefits to infants of longer hospital stays. One reason for this may be that most studies in this literature eliminate at least some categories of complicated births under the assumption that these mothers and newborns would likely be unaffected by the mandated stay laws. This assumption rests on the belief that these individuals were not discharged early either before or after the law. In reality, however, a large fraction of these babies with complications are released early. For example, from January 1, 1995, to December 31, 1996, nearly 41% of infants below 2,500 grams in California were discharged early. After passage of the federal mandate, this rate dropped to 27%. Failing to consider the effect of the law on complicated cases may be the source of conflicting estimates of the efficacy of longer lengths of postpartum stay.

<sup>5</sup> Evans, Garthwaite, and Wei (2008) contains a detailed discussion of the structure of the California Medicaid system and its relation to early discharge laws.

<sup>6</sup> The authors, however, did find that a follow-up visit within three days caused a statistically significant decrease in the probability of a hospital readmission within ten days of birth. Importantly for the purposes of this analysis, the authors focused on "normal" newborns, eliminating those below 2,000 grams and with a gestational age of less than 37 weeks or without the diagnosis-related group (DRG) code for "normal newborn." It is possible that the largest impact of Ohio's mandate may have been for these more complicated cases.

Overall, understanding the dynamics of maternity care is an interesting question in its own right. Pregnancy and delivery is one of the largest sources of health expenditures in the United States. Approximately 23% of all hospital discharges are either new mothers or newborns. In total, expenditures related to these procedures account for 27% of Medicaid charges and 15% of private insurance charges (Sakala & Corry, 2008).

### III. Data

The data for this project are from a restricted-use administrative record database of all mothers and newborns discharged from nonfederal hospitals in California from January 1, 1995, to December 31, 2000. The data set is created by the State of California Office of Statewide Health Planning and Development (OSHPD) by linking patient discharge data sets with birth, death, and fetal death certificate information. The database contains the original discharge record for a new mother and her newborn, plus all other hospital discharges over the next year for these patients. Each discharge record contains basic demographic information (such as the age, race, and sex of the patient), information about the admission (such as the length of stay, procedures used, diagnoses codes, hospital charges, the type of insurance, and whether the patient died in the hospital), plus a unique hospital ID. A scrambled social security number (SSN) available on the discharge record and coded by OSHPD allows records of the mother to be linked to that of her newborn.

Linking the discharge record to vital statistics files provides the exact date and time of birth (and therefore the newborn's admission to the hospital) and the postal code of residence. The birth record of the child also contains detailed information about the mother's pregnancy (an estimate of gestation, a list of complications during pregnancy, previous births, whether multiple births are present), some additional information about the mother (such as years of education), plus detailed information about the birth (such as birth weight and whether the baby was delivered by c-section, the sex of the infant, and whether there were delivery complications). The infant death record allows us to examine when and where a newborn died for up to one year after discharge. Our six-year database has data on approximately 3 million births, with almost 1.7 million births occurring after the passage of the California law.

The question we address is whether greater treatment intensity generated by the federal and state laws improved the health of newborns. Treatment intensity is defined by the dummy variable *Discharged Early*, which equals 1 if the newborn was released before the recommended time of 2 days for a vaginal delivery and 4 days for a c-section.<sup>7</sup>

<sup>7</sup> Although the state and federal postpartum discharge law specifies the required time in hours, we do not have the exact hour of discharge and cannot calculate the hours in the hospital. Therefore, we approximate the intent of the law by basing this key covariate on the number of nights in the hospital.

The health of the newborn is measured by the variable *28-Day Readmission*, an indicator that equals 1 if the newborn was readmitted to the hospital within 28 days.<sup>8</sup> Although the early discharge laws apply to the postpartum stays for both mothers and their newborns, in this analysis we focus exclusively on the outcomes of infants because readmission rates for mothers are so low that there is little power to analyze this outcome.

Many studies examining the effect of mandated maternity stay laws exclude large categories of newborns with complications under the assumption that these individuals were likely unaffected by the mandates (Datar & Sood, 2006; Meara et al., 2004). Examining the prelaw means of *Discharged Early*, however, reveals that a large percentage of newborns with a wide variety of complications, low birth weights, and short gestation periods are discharged early from the hospital. Furthermore, as the results will show, many of these complicated births received significant benefits from the increased length of stay. In constructing the sample for this analysis, we attempted to include more complicated deliveries while excluding categories of individuals whose length of stay was essentially unaffected by the law's passage.

In the prelaw data, the percentage of individuals released early is monotonically increasing in birth weight. For this analysis, we include those whose birth weight was 68 ounces (4.25 pounds or 1927 grams) or greater—the first birth weight with an early discharge rate greater than 10% in the prelaw period. We also exclude individuals reporting a birth weight of greater than 180 ounces (11.25 pounds), a category accounting for only 0.1 percent of the data set.

While newborns from multiple births are often excluded from these analyses, a large fraction of both twins (60%) and triplets (40%) were discharged early in the prelaw period. Therefore, we include and control for individuals in these categories while excluding quadruplets or higher—individuals who were rarely discharged early in the prelaw period.

In the data, gestation is measured in days and is calculated by taking the difference between the last normal menses and the date of birth, which is a noisy measure of the length of the pregnancy; hence, there are some extremely large and small values. Therefore, we exclude individuals whose reported gestation was fewer than 182 days (26 weeks) or greater than 315 days (45 weeks). We also exclude all births not covered by insurance, which in California is a small fraction of births.

The linked Hospital Discharge Data/Vital Statistics birth files contain a wealth of information on complications and concurrent illnesses affecting the mother or the infant. These data come from the Certificate of Live Birth—Medical Data Supplemental Worksheet. This worksheet contains

<sup>8</sup> We can also measure neonatal mortality and infant mortality, but the incidence rates for these outcomes are low, making power a challenge even for high-risk groups (Evans et al., 2008).

TABLE 1.—SAMPLE MEANS, PRIVATELY INSURED, AND MEDICAID BIRTHS IN CALIFORNIA, 1995–2000

Variable	Prelaw period, January 1995–August 1997	Postlaw period, January 1998–December 2000
<i>Discharged Early</i> (1 = yes, 0 = no)	0.788	0.511
<i>28-Day Readmission</i> (1 = yes, 0 = no)	0.034	0.032
Medicaid birth (1 = yes, 0 = no)	0.480	0.418
C-section (1 = yes, 0 = no)	0.207	0.225
Birth weight (in ounces)	119.5	119.6
Gestation (in weeks)	39.55	39.46
No pregnancy/deliver problems (1 = yes, 0 = no)	0.798	0.797
Mother's education (1 = yes, 0 = no)		
< High school	0.331	0.296
High school graduate	0.479	0.480
College graduate	0.180	0.211
Education missing	0.001	0.013
Mother's race (1 = yes, 0 = no)		
White	0.813	0.803
Black	0.069	0.065
Other	0.118	0.132
Mother's age (1 = yes, 0 = no)		
<20	0.163	0.149
≥20, <25	0.244	0.234
≥25, <30	0.271	0.272
≥30, <35	0.215	0.222
≥35, <40	0.091	0.103
≥40	0.015	0.018
Mother Is Hispanic (1 = yes, 0 = no)	0.431	0.441
Observations	1,176,797	1,322,859

30 possible complications or concurrent illnesses affecting pregnancy (medical providers can select up to 16) and 31 possible complications affecting the labor and delivery (medical providers can select up to 9). We generate indicator variables for the presence of conditions or complications that may affect the length of stay for a mother or infant. In doing so, we created a composite category for preeclampsia, eclampsia, and herpes, which were included in both the pregnancy and the delivery complications section of the worksheet. We also exclude measures that are related to medical procedures, such as electronic fetal monitoring or ultrasound, as opposed to underlying medical conditions. This results in 45 dummy variables that indicate complications.<sup>9</sup>

In the top two rows of table 1, we report the sample means for two periods. The first is in the period before any

early discharge law was in effect in California: January 1995 through August 1997. The second period is after the federal law went into effect, which provides coverage to all privately insured births and Medicaid managed care births that are part of private plans. This period covers January 1998 through the end of 2000. There are 1.18 million births in the first sample and 1.32 million births in the other. In the first two rows, we present results for the two key outcomes in the analysis: *Discharged Early* and *28-Day Readmission*. Notice the large drop in early discharges of 27 percentage points, yet there is only two-tenths of a percentage point change in readmission rates, suggesting that at best, the law had modest impact on average readmission rates.

In the remaining rows of the table, we present means of some important characteristics of the mother and the delivery. There is no change in average birth weights between the two periods and a slight decrease in the fraction of births with any of the 45 problems identified in the birth record data set. Across the two periods, there is a 10% increase in the number of c-section deliveries and a 13% decline in the fraction of births paid for by Medicaid. The drop in Medicaid births is common to a number of other states over this period as welfare reform was implemented during the late 1990s and national welfare rolls fell considerably.<sup>10</sup> There is a slight decline in mothers without a high school education, a noticeable drop in births to black

<sup>9</sup> The complications dummies are defined for the following conditions: (1) preeclampsia (pregnancy-induced hypertension); (2) eclampsia (convulsion or coma); (3) hypertension, chronic; (4) renal disease; (5) pyelonephritis; (6) anemia; (7) cardiac disease; (8) lung disease, acute or chronic; (9) diabetes; (10) Rh sensitivity; (11) uterine bleeding before labor; (12) hemoglobinopathy; (13) transport of mother from another facility prior to delivery; (14) polyhydramnios; (15) incompetent cervix; (16) premature labor; (17) genital herpes; (18) other STD; (19) hepatitis B; (20) rubella; (21) smoking; (22) birth weight more than 4,000 grams; (23) birth weight less than 2,500 grams; (24) cervical cerclage; (25) less than 37 weeks gestation; (26) chronic villus sampling; (27) cord prolapse; (28) fetal distress; (29) seizures during delivery; (30) maternal blood transfusion; (31) fetopelvic disproportion; (32) shoulder dystocia; (33) breech presentation; (34) precipitous delivery (less than 3 hours); (35) prolonged delivery (more than 20 hours); (36) unsuccessful attempt at vaginal birth after c-section; (37) other dysfunctional delivery; (38) premature rupture of membrane (over 12 hours); (39) abruptio placenta; (40) placenta previa; (41) other excessive bleeding; (42) amnionitis/sepsis; (43) febrile (over 100°F); (44) presence of meconium, moderate/heavy; (45) anesthetic complications.

<sup>10</sup> For example, data from the Centers for Disease Control's Pregnancy Risk Assessment Monitoring System show large declines in the fraction of births covered by Medicaid during the 1995–2000 time period in a number of larger states in their sampling frame, including Florida, New York, Illinois, and Washington. <http://www.cdc.gov/PRAMS/2002/PRAMSSurvReport/MultiStateExhibits/Multistates9.htm>.

mothers, a slight decline in births to younger women, and a slight increase in the fraction of births to Hispanic mothers. Given the extremely large sample sizes, all differences in the two columns except for the gestation variable are statistically significant.

#### IV. Empirical Model and Baseline Results

In theory, we could estimate the impact of an early discharge on the 28-day readmission rate using a single-equation OLS model where we control for a detailed set of covariates describing the mother, the pregnancy, and the delivery. Let  $y_{it}$  represent readmission rates for person  $i$  in time period  $t$ . The equation of interest related how an early discharge ( $x_{it}$ ) alters readmission rates, and this can be captured by the equation

$$y_{it} = \beta_0 + x_{it}\beta_1 + w_{it}\beta_2 + \varepsilon_{it}, \quad (1)$$

where  $w_{it}$  represents a vector of exogenous covariates,  $\varepsilon_{it}$  is a random error, and we anticipate that  $\beta_1 > 0$ . In practice, however, this model is likely to be subject to an omitted variables bias, and we anticipate that  $\text{cov}(x_{it}, \varepsilon_{it}) < 0$ . Evans et al. (2008) demonstrate that factors that tend to increase the probability of 28-Day Readmission also tend to reduce Discharged Early. For example, Discharged Early rates are lowest for first births, multiple births, younger women, women with insurance, and those with complications such as low birth weight, breech presentation, preeclampsia, or eclampsia. These same characteristics are also the factors that predict greater readmission rates among newborns. This suggests that  $\text{cov}(x_{it}, \varepsilon_{it}) < 0$  and single equation estimates of  $\beta_1$  will be biased toward 0.

OLS estimates of equation (1) using only data from the prelaw period (January 1995–August 1997) are provided at the top of table 2. In this equation, we control for a number of covariates, including dummies for all 45 prenatal and delivery complications; a complete set of dummies that describe the age, education, race, marital status, twins, triplets, and Hispanic origin of the mother; a dummy for the sex of the newborn; dummies that measure the month, year, weekday of birth, and birth hour; plus dummies that measure the ownership of the hospital, the size of the hospital, and the hospital service region within California. Because readmission rates vary considerably based on type of delivery (vaginal or c-section) and insurance status (private or Medicaid), we include a set of dummies for privately insured c-sections, Medicaid-insured vaginal deliveries, and Medicaid-insured c-sections with privately insured vaginal births as the reference group. We also include quadratic terms in birth weight and gestational age (measured in ounces and weeks, respectively). Finally, we include a linear trend in the months since January 1995, plus three trends that vary based on insurance status and delivery method dummies outlined above. In all, there are 160 covariates in this model and 1,176,797 observations. In this and subsequent models, we estimate the variance-covariance

TABLE 2.—OLS AND 2SLS ESTIMATES OF 28-DAY READMISSION EQUATION, PRIVATELY INSURED AND MEDICAID BIRTHS IN CALIFORNIA, JANUARY 1995–AUGUST 1997, JANUARY 1998–DECEMBER 2000

OLS Estimates of 28-Day Readmission equation in prelaw period	
Discharged Early	0.0010 (0.0007)
$R^2$	0.0034
Sample means, prelaw period	
Discharged Early	0.788
28-Day Readmissions	0.034
Observations	1,176,797
First stage estimates, Discharged Early equation	
Private insurance $\times$ Vaginal Delivery $\times$ Federal Law	-0.2958*** (0.0018)
Private insurance $\times$ C-Section $\times$ Federal Law	-0.1374*** (0.0033)
Medicaid $\times$ Vaginal Delivery $\times$ Federal Law	-0.1367*** (0.0020)
Medicaid $\times$ C-Section $\times$ Federal Law	-0.0311*** (0.0039)
Medicaid $\times$ Vaginal Delivery $\times$ State Expansion	-0.0667*** (0.0017)
Medicaid $\times$ C-Section $\times$ State Expansion	-0.0522*** (0.0033)
F-test, $H_0$ : all instruments = 0	5916.3
(p-value)	(0.000)
Observations	2,499,656
2SLS Estimates, 28-Day Readmission equation	
Discharged Early	0.0047* (0.0028)
P-value, test of overidentifying restrictions	0.059
Observations	2,499,656

Standard errors (in parentheses) were calculated allowing for an arbitrary correlation in errors within hospitals over time. Other covariates in the model include quadratic terms in birth weight and gestation; fixed effects for mother's age, race, ethnicity, education, previous births, and multiple births; the month, weekday, and hour of the day; the size, location, and ownership status of the hospital; and the baby's sex. The models also include dummies for the interaction of delivery type (c-section versus vaginal) and insurance status (private versus Medicaid), plus a unique monthly time trend for each of these unique combinations.  $p$ -value \* < 0.10. \*\* < 0.05. \*\*\* < 0.001.

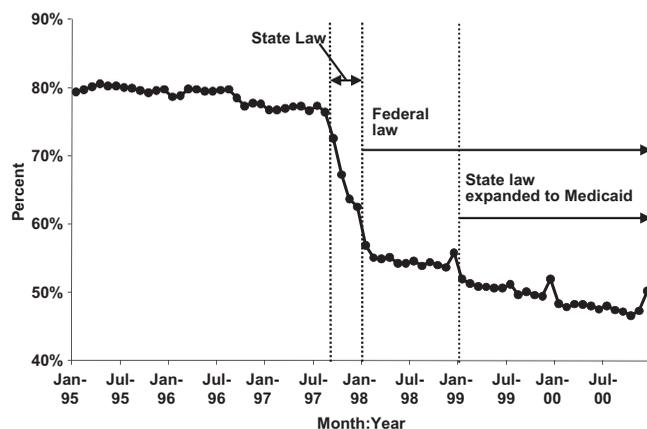
matrix, allowing for arbitrary correlation in errors within a particular hospital.

In this instance, the estimated impact of an early discharge on readmission rates is rather modest. The coefficient in the first line of table 2 suggests that an early discharge raises the probability of a readmission by a statistically insignificant one-tenth of a percentage point or 3% of the sample mean.

Given a vector of instruments  $z_{it}$  that predict early discharge rates but are uncorrelated with unexplained portion of equation (1), we can use 2SLS to provide a consistent estimate of  $\beta_1$ . In this case, the federal and state early-discharge laws provide useful variation that can be exploited as instruments. As we demonstrate, the laws dramatically decreased the fraction of newborns discharged early. If the greater intensity of care improves newborn health, we should see a noticeable drop in readmission rates after passage of the various state and federal laws.

A graphical treatment of the first-stage and reduced-form models associated with the 2SLS estimation are presented in figures 1 and 2 respectively. In figure 1, we present the monthly averages for the percentage discharge rate early from January 1995 through the end of 2000. On the graph, we indicate the periods when only the state law was in effect (which covered only some private and Medicaid

FIGURE 1.—PERCENTAGE OF NEWBORNS IN CALIFORNIA DISCHARGED EARLY, PRIVATELY INSURED AND MEDICAID BIRTHS, 1995–2000

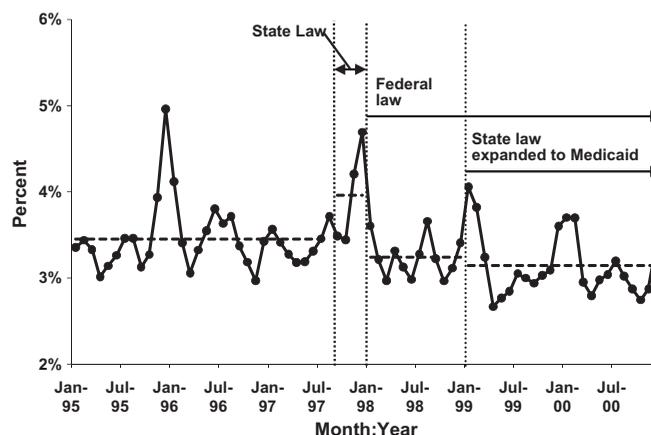


patients), when the federal law went into effect (which covered all privately insured births and Medicaid patients in privately provided Medicaid managed care plans), and when the state law was expanded to include all Medicaid patients. Note that early discharges were declining slowly prior to the passage of the state law, but early discharge rates did not drop sharply until the state law was passed and went into effect the same day. Insurance carriers did not appear to be adjusting to the federal law in advance of its effective date. Second, the change in early discharge rates produced by the various statutes is very large. In July 1997, early discharge rates were 77%. By January 1998, this number had fallen to 57% and by January 1999, the rate was down to 52%.

In figure 2, we graph the time series of the 28-day readmission rates for the same period. The dotted lines represent the mean readmission rates during the prelaw, state law, federal law, and expanded state law period, respectively. Three results are apparent in these graphs. First, the month-to-month variation in readmission rates is large, with readmission rates highest during the winter months. Second, during the four months the state law was in effect, there was a large spike in readmissions, generated by a particularly heavy flu season. Any analysis of the benefits of greater intensity of care using data from this period will generate spurious results showing more care produces worse outcomes since the intensity of care increased (early discharge rates fell) and readmission soared. For this reason, in our subsequent regression work, we will delete data from the four-month period when only the state law was in effect. Third, there is a small but noticeable drop in readmission rates when the federal law and expanded state law went into effect. Between the prelaw and federal law period, readmission rates fell by about two-tenths of a percentage point, and rates fell by another one-tenth of a percentage point in the expanded state law period.

We present a set of 2SLS estimates in the bottom of table 2. In this case, we estimate a model similar to the OLS model but include data from January 1995 through December

FIGURE 2.—PERCENTAGE OF NEWBORNS IN CALIFORNIA WITH A 28-DAY READMISSION, PRIVATELY INSURED AND MEDICAID BIRTHS, 1995–2000



2000, excluding the four months the state law was the only law in effect (September–December 1997). We include the same set of covariates as in the OLS model. As instruments for *Discharged Early*, we exploit the timing of the laws and the fact that the laws affected patients based on their insurance status and method of delivery. We use six instruments. We allow the federal law to vary based on insurance status (private versus Medicaid) and delivery method (vaginal and c-section), and we allow the expansion of the state law to all Medicaid patients to effect vaginal and c-section deliveries for this group as well.

In the middle of the table, we present the first-stage estimates for the instruments in the *Discharged Early* equation. All of the instruments are statistically significant and suggest that the various early discharge laws worked as anticipated, albeit with differing levels of success. For those with private insurance, the federal law reduced early discharges by 30 percentage points among vaginal births but by only 13.7 percentage points among c-sections. The estimates for vaginal births represent a larger change in absolute and relative terms. Among those with private insurance in the pre-federal law period, the fraction discharged early averaged 83% for vaginal births and 88% for c-sections. Among Medicaid patients, the federal law again had its largest impact on vaginal deliveries. The expansion of the state law to include Medicaid patients reduced early discharges by another 5 to 6 percentage points for both vaginal and c-section deliveries. The *F*-test that the instruments are jointly 0 indicates that this hypothesis can be easily rejected, and the size of the test statistic indicates no concern associated with finite sample bias (Bound, Jaeger, & Baker, 1995; Staiger & Stock, 1997).

In the final portion of table 2, we report the 2SLS estimates of the readmission equation. As expected, the 2SLS estimate on *Discharged Early* is larger than the OLS value. The estimated coefficient of 0.0047 is statistically significant at the 10% level and represents a change of about 14% of the sample mean for the dependent variable in the prelaw period. The *p*-value on the test of overidentifying restric-

tions is greater than 0.05, meaning we cannot reject the null the model is properly specified.

### V. Heterogeneity in the Returns to Greater Medical Care Use

Angrist, Imbens, and Rubin (1996) demonstrate that 2SLS results can be interpreted as a local average treatment effect (LATE) where the 2SLS model estimates identify the impact of the endogenous variable of interest for people whose behavior was altered as a result of receiving the instrument. In this context, the 2SLS model identifies the impact of staying longer in the hospital for the 25% of newborns who had longer stays as a result of the federal and state early discharge laws. The sample means in table 1, the graphical treatment in figure 2, and now the 2SLS estimates in table 2 all tell a similar story: the benefits of longer stays for the average newborn required by the law are modest at best. This result is consistent with the flat-of-the-curve story that is typically used to describe medical care in the United States. Given third-party reimbursements, the marginal cost to insured consumers of using additional medical care is low. As a result, medical care will be consumed to the point where marginal benefits equal marginal cost. Hence, the marginal benefits of additional treatments are low.

Not all newborns are subject to the same risk of readmission, and as a result, one would not expect the benefits of longer hospital stays to be the same across different groups. The early-discharge laws were a rather blunt policy instrument: all newborns, regardless of risk, were eligible for a minimum stay of two days after a vaginal birth and four days after a c-section. From a policy standpoint, if one could a priori identify those at the greatest risk of readmission and apply the law only to those newborns, the law might be able to generate similar medical outcomes at a much lower cost.

From an econometric standpoint, identifying heterogeneity in the returns to longer stays across different groups is hampered by two factors. First, the incidence rate for the outcome of interest (readmission within 28 days) is low, so identifying precise 2SLS estimates is a challenge even with our large sample sizes. From table 2, note that with over 2 million observations, a first stage that changes treatment use by 25 percentage points, and a treatment effect that is 15% of the sample mean, the results are statistically significant only at a  $p$ -value of 0.10. While we can generate large subsamples of people with easily identifiable characteristics (say, those born by c-section or Medicaid patients), we anticipate that even within these large, relatively heterogeneous groups, the treatment effect will likely vary. Second, we anticipate that longer hospital stays should be most beneficial to those most in medical need. However, those with higher risks of readmission are already staying longer in the hospital, meaning that the first-stage relationship should be smaller for this group. Since the precision of the second-stage model is related to the size of the first-stage effect, the

declining first stage for those most in need of treatment should lessen our ability to detect a precise 2SLS estimate in the groups most likely to benefit.

In this analysis, we attempt to isolate heterogeneity in the returns to additional medical care use through two procedures. Initially we restrict the sample to groups with identifiable demographic and medical characteristics that are indicative of medical needs and examine the treatment heterogeneity across those subsamples. This method produces some predictable results. For example, in each case, the 2SLS coefficient is greater than the OLS estimate, suggesting that the OLS models systematically understate the benefits of longer stays. However, there are few statistically significant 2SLS coefficients, and it is hard to identify persistent patterns across the subgroups. Given these results, we suggest a procedure based on the propensity score of early discharge to aggregate the data about the risks of readmission to newborns.

#### A. Heterogeneity in Results by Specific Characteristics

In table 3, we report 2SLS estimates for certain subgroups of newborns. In the first row of the table, we reproduce 2SLS estimates for the full sample from table 2. For each model, we report the observations, the prelaw sample means for *Discharged Early* and *28-Day Readmission*, the OLS and 2SLS estimates on the *Discharged Early* variable, the  $p$ -value on the test of overidentifying restrictions, and the first-stage  $f$ -statistic for the test of the null that all instruments have a zero coefficient. In the next two rows, we provide estimates for subgroups based on type of delivery (c-sections and vaginal deliveries). The 2SLS estimates on *Discharged Early* in the vaginal delivery sample is small and statistically insignificant. In contrast, the coefficient on the same variable in the c-section models is a statistically significant 1.9 percentage points—nearly 70% of the sample mean. Treating the 2SLS coefficients on *Discharged Early* in the c-section and vaginal deliveries samples as independent random variables, we can reject the null that the coefficients are the same at a  $p$ -value of 0.10.

Next, we exploit the detailed data available on the birth and hospital discharge records to consider the impact of longer stays for newborns with certain sets of risk factors. As we noted, our data allow us to construct 45 dummy variables that measure complications associated with the pregnancy or delivery. In an initial probit equation using data from the prelaw period, we modeled the probability of an early discharge as a function of the list of covariates used in table 2, which includes these 45 conditions. We then identified the problems that were estimated to have a statistically significant decrease in the probability of an early discharge. These conditions can be thought of as those characteristics that a priori were considered by providers and the newborn's parents as requiring additional treatment; therefore, these can be considered indicators of enhanced medical risk or need for additional care. For lack of a better term, we call

TABLE 3.—OLS AND 2SLS ESTIMATES OF 28-DAY READMISSION EQUATION, PRIVATELY INSURED AND MEDICAID BIRTHS IN CALIFORNIA, JANUARY 1995–AUGUST 1997, JANUARY 1998–DECEMBER 2000

Sample	Observations	Means in Prelaw period		Coefficient (standard error) on Discharged Early		P-Value Test of Over-id.	Number of Instruments	F-test (P-value) First-Stage Instruments
		Discharged Early	28-Day Readmission	OLS	2SLS			
Full sample	2,499,656	0.788	0.034	0.001 (0.0007)	0.0047* (0.0028)	0.059	6	5,916.3 (0.000)
C-sections	541,042	0.863	0.028	0.0027** (0.0013)	0.0189** (0.0087)	0.141	3	785.3 (0.000)
Vaginal deliveries	1,958,614	0.768	0.036	0.0006 (0.0008)	0.0036 (0.0029)	0.070	3	10,821.3 (0.000)
No severe problems	1,992,783	0.815	0.034	-0.0007 (0.0008)	0.0024 (0.0029)	0.230	6	5,086 (0.000)
1 or more severe problems	506,873	0.680	0.038	0.0058*** (0.0013)	0.0158** (0.0066)	0.450	6	891.2 (0.000)
2 or more severe problems	115,441	0.552	0.042	0.0142*** (0.002)	0.0318* (0.0174)	0.533	6	132.6 (0.000)
3 or more severe problems	27,102	0.430	0.041	0.0211*** (0.0044)	0.0490 (0.0349)	0.147	6	21.9 (0.000)

Standard errors (in parentheses) were calculated allowing for an arbitrary correlation in errors within hospitals over time. Other covariates in the model include quadratic terms in birth weight and gestation; fixed effects for mother's age, race, ethnicity, education, previous births, and multiple births; the month, weekday, and hour of the day; the size, location, and ownership status of the hospital; and the baby's sex. The models also include dummies for the interaction of delivery type (c-section versus vaginal) and insurance status (private versus Medicaid), plus a unique monthly trend for each of these unique combinations. *p*-value: \* < 0.10, \*\* < 0.05, \*\*\* < 0.001.

these “severe problems,” and we estimate 2SLS models of sample of newborns with severe problems ranging from none to more than three. These results are reported in the final four rows of table 3.

Note that as we move from no problems to three or more, sample size shrinks considerably, the early discharge rate in the prelaw period falls precipitously, and the readmission rate increases nearly monotonically. The 2SLS coefficient on the *Discharged Early* variable in the “no problems” sample is the smallest estimate in that column, with a statistically insignificant value of 0.0024, a number that is only 7% of the sample mean. In contrast, the same coefficient in the sample for one or more problems is a statistically significant 1.6 percentage points (approximately 40 percent of the sample mean). Treating the 2SLS coefficients on *Discharged Early* in the no problems and one or more problems samples as independent random variables, we can reject the null that the coefficients are the same at a *p*-value of 0.06. Looking at the sample of people with two or more problems, the coefficient doubles to 3.2 percentage points (approximately 75% of the sample mean) with a *p*-value of 0.07. In the sample of three or more problems, the estimated 2SLS coefficient is even larger, at 4.9 percentage points, but the drop in the sample size to 27,102 observations results in a statistically insignificant result (*p*-value of 0.161).

We have three additional general comments about the results in table 3. First, in all cases, the 2SLS estimates of the *Discharged Early* coefficient are larger than the OLS estimates, suggesting that the OLS results are biased downward. This is no surprise: those with the greatest risk of a readmission have longer hospital stays, so the OLS coefficient on *Discharged Early* will be biased down. Second, in all cases, the first-stage *f*-statistic testing the null that the coefficients on the instruments are all 0 is large, dispelling any concern about finite sample bias in the 2SLS models.

Third, on average, the first-stage *f*-statistic is positively related with the prelaw *Discharged Early* rates (the correlation coefficient is 0.68). Again, this is no surprise: only those who would otherwise be discharged early can be affected by the law, so the law should have a smaller impact on those who are already staying longer.

In order to better understand the underlying conditions driving the pattern of results in table 3, we select five of the largest prenatal and delivery problems and estimate separate models for newborns who present with these specific conditions. In table 4 we present the results based on the incidence rate of the condition, from largest to smallest. Three of the five 2SLS estimates on the *Discharged Early* variables are statistically significant (in the fetal distress, diabetes, and preeclampsia/eclampsia samples). These point estimates are large in magnitude, ranging from 3.5 percentage points (83% of the sample mean) for the diabetes sample to nearly 5.4 percentage points (135% of the sample mean) for the preeclampsia and eclampsia sample. In the other two other cases (the meconium delivery and other dysfunctional delivery samples), the estimated LATE is fairly large—45% to 60% of the sample mean—but these results are not statistically significant.

The inconsistent pattern of results by condition status displays a need for a more systematic method of identifying the pattern of heterogeneity in the benefit of longer hospital stays. One natural candidate for this systematic method is birth weight. A baby's birth weight has been a focal point of researchers for establishing increased medical need. The last two rows of table 4 contain the results for sample based on a baby being above or below 5 pounds. While neither 2SLS estimate is statistically significant, they suggest an underlying pattern of heterogeneity. The estimated effect for babies with a birth weight less than 5 pounds is twelve times larger than the point estimate for those weighing

TABLE 4.—OLS AND 2SLS ESTIMATES OF 28-DAY READMISSION EQUATION, PRIVATELY INSURED AND MEDICAID BIRTHS IN CALIFORNIA, JANUARY 1995–AUGUST 1997, JANUARY 1998–DECEMBER 2000

Sample	Observations	Means in Prelaw Period		Coefficient (Standard Error) on Discharged Early		P-Value Test of Over-id.	Number of Instruments	F-test (P-Value) First-Stage Instruments
		Discharged Early	28-Day Readmission	OLS	2SLS			
Full sample	2,499,656	0.788	0.034	0.0010 (0.0007)	0.0047* (0.0028)	0.059	6	5,916.3 (0.000)
By specific complications and conditions								
Meconium delivery	114,416	0.739	0.032	-0.0041** (0.0020)	0.0188 (0.0138)	0.258	6	245.3 (0.000)
Fetal distress	70,005	0.718	0.034	0.0055* (0.0032)	0.0443** (0.0212)	0.061	6	95.4 (0.000)
Preeclampsia/eclampsia	54,042	0.605	0.040	0.0080** (0.003)	0.0543** (0.022)	0.736	6	77.2 (0.000)
Diabetes	51,848	0.719	0.042	0.0064** (0.0033)	0.0347* (0.0204)	0.881	6	102.4 (0.000)
Other dysfunctional delivery	44,722	0.830	0.032	-0.0027 (0.0035)	0.0140 (0.0263)	0.264	6	54.2 (0.000)
By birth weight								
Birth weight < 5 lbs	36,964	0.288	0.042	0.0272*** (0.0050)	0.0536 (0.0464)	0.561	6	20.2 (0.000)
Birth weight ≥ 5 lbs	2,462,692	0.795	0.034	0.0003 (0.0007)	0.0045 (0.0028)	0.083	6	5,724.7 (0.000)

Standard errors (in parentheses) were calculated allowing for an arbitrary correlation in errors within hospitals over time. Other covariates in the model include quadratic terms in birth weight and gestation; fixed effects for mother's age, race, ethnicity, education, previous births, and multiple births, the month, weekday, and hour of the day; the size, location, and ownership status of the hospital; and the baby's sex. The models also include dummies for the interaction of delivery type (c-section versus vaginal) and insurance status (private versus Medicaid), plus a unique monthly trend for each of these unique combinations. *p*-value: \* < 0.10, \*\* < 0.05, \*\*\* < 0.001.

more than 5 pounds. Individuals weighing less than 5 pounds clearly have high medical need, as seen by the *Discharged Early* rate of only 28.8% and a 28-day readmission rate of 4.2%. In contrast, approximately 80% of individuals above 5 pounds are discharged early from the hospital, and they have a 28-day readmission rate of 3.4%. The difficulty is that low birth weight individuals are exceptionally low in number (fewer than 40,000 babies even in this large data set are smaller than 5 pounds at birth), making it difficult to produce statistically significant estimates.

Overall, the results in tables 3 and 4 suggest several factors that may be useful in identifying heterogeneity in the treatment effect. First, our measure of problems provides an indicator of medical need that is reflected in lower early discharges and higher readmission rates in the prelaw period. Second, as these medical problems accumulate, the benefits of avoiding an early discharge increase considerably. Third, for the vast majority of patients in the sample, there is no estimated medical benefit to longer hospital stays.

These estimates also show, however, that even with our large sample sizes, we struggle to attain statistical precision when we restrict the sample to specific conditions. One method of potentially leveraging the size of data set is to estimate a model on the entire sample that allows the coefficient on *Discharged Early* to vary by the specific conditions. Because we have 161 coefficients in the 2SLS model, it would be easiest to implement such a model that does not allow the other coefficients to vary by condition. As an illustration, we estimated a 2SLS model using the same six instruments as before but allowing the *Discharged Early* coefficient to vary for c-section and vaginal births. The coefficients (standard error allowing for within-hospital cor-

relation in residuals) on *Csection* × *Discharged Early* and *Vaginal* × *Discharge Early* are 0.0157 (0.0082) and 0.0041 (0.0028), respectively. The estimates in this case are similar but slightly lower in value than the estimates for *Discharge Early* in table 4 where we estimate separate models by method of delivery. Note that in these new results, there is a slight improvement in precision in the treatment effect variable for c-section births in the pooled model but no change in precision for the same coefficient for vaginal births. It is, however, not clear that we should be pooling the models in this fashion. This model makes the strong assumption that the coefficients on the other covariates are the same for vaginal and c-section births. In a basic OLS model where we use the prelaw period data only and regress the 28-day readmission indicator on a set of covariates plus the *Early Discharge* indicator, we can easily reject null that the coefficients on the covariates are the same for c-section and vaginal births.<sup>11</sup>

#### B. Using the Propensity Score to Identify Medical Need

Identifying a pattern in the results using either individual conditions or birth weight is hampered by the relatively small sample sizes and the need for a systematic method for grouping people based on their relative levels of medical need. In order to overcome this problem, we use the rich medical and demographic data on individuals during the

<sup>11</sup> In the OLS model, table 3, there are 155 coefficients that are not interacted with type of delivery, so the *F*-test has 155 degrees of freedom and functionally infinite degrees of freedom in the numerator. The test statistic is 5.8, and under the null that the coefficients are the same, we can reject the null with a *p*-value of 5E-105.

prelaw period to identify patients who are most likely, in the opinion of medical practitioners, to require longer hospital stays. Specifically, we estimate the following probit equation on a sample of individuals during the time period before the federal or state mandate:

$$\text{Prob}(\text{Discharged Early}_{it} = 1) = P_{it} = \Phi[\gamma_0 + w_{it}\gamma_2], \quad (2)$$

where  $\Phi$  is the standard normal CDF and all other variables are defined as in equation (1). The predicted probability from this equation is the propensity score, and it provides valuable information about the clinical opinion of the appropriateness of additional medical care. Higher levels of this propensity score indicate newborns who are more likely to be discharged early from the hospital. Conversely, lower values are newborns whose medical conditions required longer hospital stays. We estimate the parameters from this model with prelaw data only, because the decisions are based on medical parameters before the physician's and patient's decision was altered by legislation. As the numbers in figure 1 indicate, there is little, if any, trend in the dependent variable from January 1995 through August 1997, so it does not appear that the underlying factors used to decide whether to discharge a newborn early were altered much over that period.

In figure 3, we present a histogram of the distribution of the propensity score in the prelaw and post-federal law period. The vertical axis is the propensity score rounded to the nearest 0.01, and the lines represent the fraction of the sample with a propensity score with that value. The dark lines on the left-hand side represent the prelaw period and the light gray lines represent the post-federal law period. Within the graph, two results are of note. First, given the large number of observations and the number of detailed covariates in the model, it should be no surprise that the propensity score of an early discharge spans all values from 0.00 to 1.00. Second, the pre- and postlaw periods have virtually identical distributions. Therefore, aggregating data within ranges of the propensity score and comparing outcomes pre- and postlaw, we are comparing similar populations in size and distribution of propensity scores.

The information contained in the propensity score about medical need appears to be quite high as well. We break the pre- and postlaw propensity scores into 20 even-sized groups (vigintiles) and graph the mean value of newborn birth weight (figure 4), the mean value of gestation in days (figure 5), and number of complications (figure 6) against the vigintiles of the propensity score. In figures 4 to 6, two trends are evident. First, the propensity score does an excellent job of reflecting risk to the newborn. Moving from the lowest to highest vigintile in the propensity score, the average birth weight and gestational age increase, and the mean number of complications declines. In these graphs, the relationships are monotonic, and the curves for the pre- and postlaw period lie virtually on top of each other, suggesting

that the propensity score in both periods is providing similar information about these two groups.

For similarly defined people, we anticipate that the early discharge laws have very different effects in the postlaw period on discharges and readmissions. In figure 7, we graph the early discharge rates for the pre- and postperiods against the vigintiles of the propensity score. Again, there appears to be tremendous information about the underlying health of the newborn. For both periods, early discharge rates are nearly monotonically increasing in the vigintiles of the propensity score. We should expect large gaps between the post- and prelaw graphs because the laws decrease the probability of an early discharge for similarly defined people. However, the gap is not uniform within the sample. Between the fifth and the fifteenth vigintiles, the average change in early discharge rates is nearly 34 percentage points. In contrast, the average for individuals between the first and fifth and the fifteenth and twentieth vigintiles is only 22 percentage points. This shows that early discharge laws have less of an impact on the length of stay for newborns who are always discharged early (high propensity scores) or those we anticipate are in most need of more intensive treatment (low propensity scores).

Finally, in figure 8, we graph the 28-day readmission rate for both groups against the vigintiles of the propensity score. The information content of the propensity score about the underlying health of the infant is illustrated by the fact that for both groups, readmission rates are nearly monotonically declining as we increase the propensity score. The gap between the two curves is somewhat uniform across the vigintiles of the propensity score, but this graph in some respects represents a reduced-form relationship: What is the impact of the early discharge laws on similarly defined people? Recall from figure 6 that for people in the middle vigintiles of propensity scores, the change in early discharge rates was much greater than at the tails, so the implied local average treatment suggested by the two graphs will be very different from the reduced-form model suggested by figure 8.

Table 5 contains the 2SLS estimates for groupings of individuals based on their respective propensity scores. For each model, we present the same estimates as in tables 3 and 4. Choosing the appropriate group size for each model involves balancing the risk of type 2 errors caused by small sample sizes with the risk of watering down the estimated treatment effect by including too many individuals who may differ in the likely benefit generated from a longer hospital stay. As can be seen in figure 3, while the propensity score is identified across the (0,1) support, the vast majority of observations have high propensity scores. Therefore, the number of individuals with low scores—those who are most likely to stay longer in the hospital—is relatively small.

Given these trade-offs, we estimate the 2SLS models by breaking the sample into thirds of the predicted *Discharged Early* propensity score and estimate the 2SLS model outlined in equation (1) for each of these subgroups. Examin-

FIGURE 3.—HISTOGRAM OF DISTRIBUTION OF PROPENSITY SCORES, PRIVATELY INSURED AND MEDICAID BIRTHS, 1995–2000

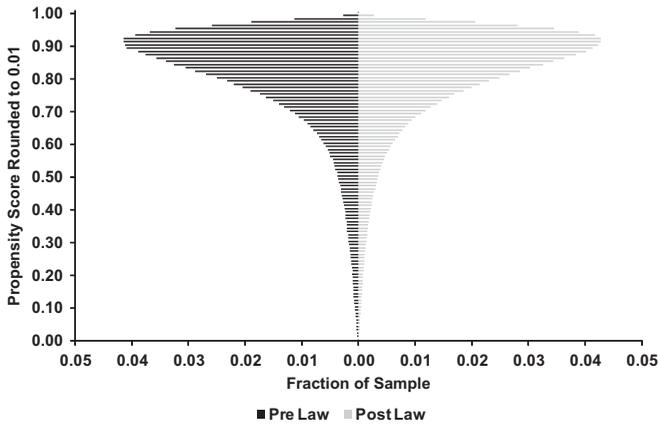


FIGURE 6.—MEAN COMPLICATIONS BY VIGINTILES OF THE PROPENSITY SCORE, PRIVATELY INSURED AND MEDICAID BIRTHS IN CALIFORNIA, 1995–2000

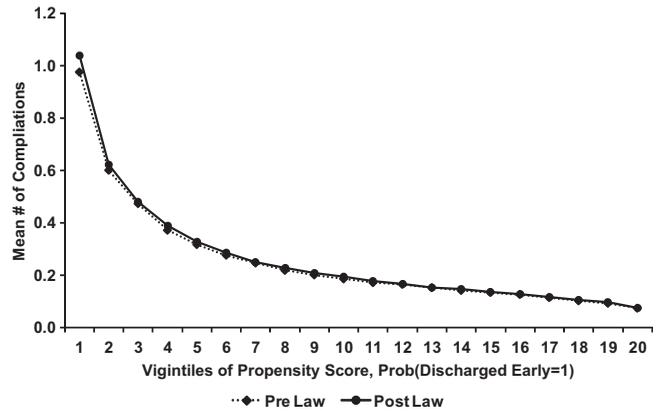


FIGURE 4.—MEAN BIRTH WEIGHT IN OUNCES BY VIGINTILES OF THE PROPENSITY SCORE, PRIVATELY INSURED AND MEDICAID BIRTHS IN CALIFORNIA, 1995–2000

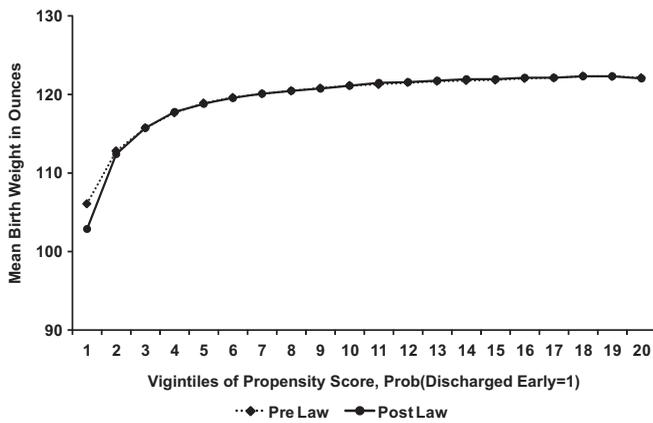


FIGURE 7.—PERCENTAGE DISCHARGED EARLY BY VIGINTILES OF THE PROPENSITY SCORE, PRIVATELY INSURED AND MEDICAID BIRTHS IN CALIFORNIA, 1995–2000

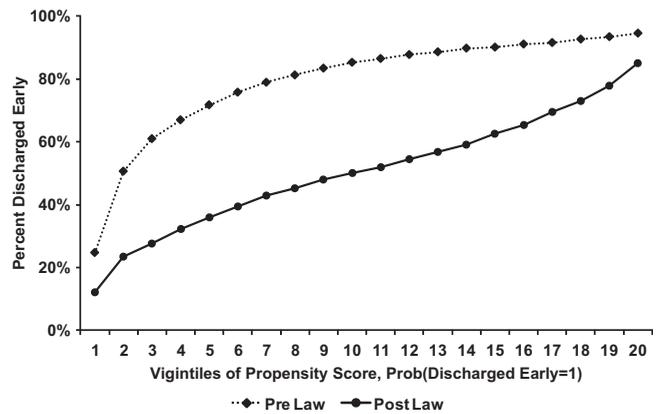


FIGURE 5.—MEAN DAYS GESTATION BY VIGINTILES OF THE PROPENSITY SCORE, PRIVATELY INSURED AND MEDICAID BIRTHS IN CALIFORNIA, 1995–2000

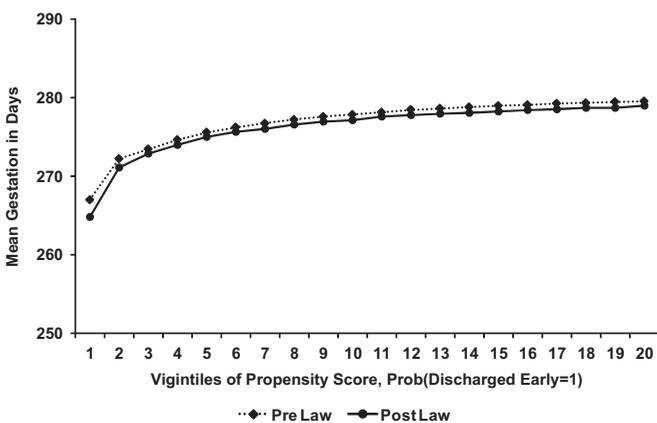
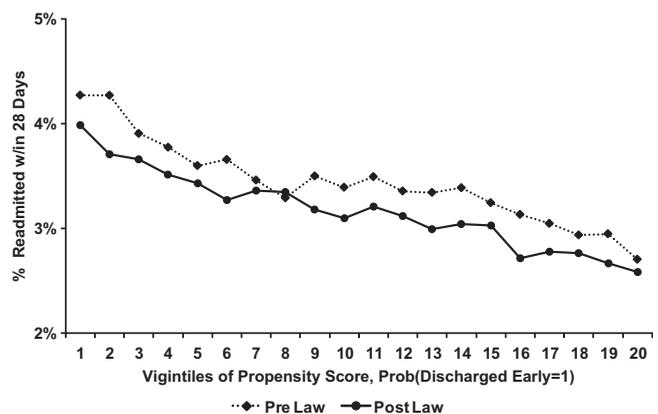


FIGURE 8.—PERCENTAGE READMITTED WITHIN 28 DAYS BY VIGINTILES OF THE PROPENSITY SCORE, PRIVATELY INSURED AND MEDICAID BIRTHS, CALIFORNIA, 1995–2000



ing the results in the first third of table 5 reveals several consistent trends. First, as was graphically seen in figure 5, the readmission rate is monotonically declining in the propensity score. Second, as with previous tables, the 2SLS estimates are always larger than the OLS estimates, again

providing suggestive evidence that the OLS model is biased down because those with the greatest risk of a readmission are the least likely to be discharged early. Finally, the  $F$ -tests for the null hypothesis that first-stage instruments are 0 are all uniformly large.

TABLE 5.—OLS AND 2SLS ESTIMATES OF 28-DAY READMISSION EQUATION, PRIVATELY INSURED AND MEDICAID BIRTHS IN CALIFORNIA, JANUARY 1995–AUGUST 1997, JANUARY 1998–DECEMBER 2000

Sample	Observations	Means in Prelaw Period		Coefficient (Standard error) on <i>Discharged Early</i>		P-Value Test of Over-id.	Number of Instruments	F-Test (P-Value) First-Stage Instruments
		<i>Discharged Early</i>	<i>28-Day Readmission</i>	OLS	2SLS			
By thirds of the propensity score of <i>Discharged Early</i>								
Bottom third	833,219	0.599	0.039	0.0027** (0.0010)	0.0115** (0.0054)	0.390	6	1,658.8 (0.000)
Middle third	833,219	0.853	0.034	-0.0010 (0.0010)	-0.0003 (0.0034)	0.150	6	3,028.4 (0.000)
Top third	833,218	0.920	0.030	-0.0015* (0.0012)	0.0061 (0.0049)	0.137	6	1,477.5 (0.000)
By thirds of the propensity score of <i>Discharged Early</i> for c-section births								
Bottom third	180,348	0.741	0.032	0.0061** (0.0019)	0.0355** (0.0137)	0.222	3	270.5 (0.000)
Middle third	180,347	0.901	0.027	-0.0022 (0.0021)	0.0192 (0.0136)	0.421	3	256.4 (0.000)
Top third	180,347	0.947	0.024	-0.0003 (0.0023)	-0.0029 (0.0152)	0.224	3	291.5 (0.000)
By thirds of the propensity score of <i>Discharged Early</i> for vaginal births								
Bottom third	652,872	0.566	0.039	0.0017 (0.0011)	0.0088 (0.0057)	0.319	3	2,714.4 (0.000)
Middle third	652,871	0.836	0.035	-0.0006 (0.0011)	0.0005 (0.0036)	0.200	3	5,204.2 (0.000)
Top third	652,871	0.911	0.035	-0.0016 (0.0013)	0.0032 (0.0043)	0.649	3	3,402.9 (0.000)

Standard errors (in parentheses) were calculated allowing for an arbitrary correlation in errors within hospitals over time. Other covariates in the model include quadratic terms in birth weight and gestation; fixed effects for mother's age, race, ethnicity, education, previous births, and multiple births; the month, weekday, and hour of the day; the size, location, and ownership status of the hospital; and the baby's sex. The models also include dummies for the interaction of delivery type (c-section versus vaginal) and insurance status (private versus Medicaid), plus a unique monthly trend for each of these unique combinations. *p*-value: \* < 0.10, \*\* < 0.05, \*\*\* < 0.001.

In the first third of the table, we present results for groupings of third of the propensity score for all delivery types. When we break the propensity score up into thirds, the bottom third of the distribution has a prelaw period readmission rate that is 30% higher than the rate for those in the top third. Classifying individuals into thirds based on their propensity score shows large differences in the estimated mean effect of being discharged early. Individuals in the bottom third of the propensity score distribution who are discharged early have an estimated increase of the chance of a readmission of 1.15 percentage points, which is approximately 28% of the sample mean. This estimate is statistically significant at the 5% level and dwarfs that estimate for those in the middle third of the distribution. For these newborns, at the mean, being discharged early decreases the chance of a readmission by 0.03 percentage points, which is less than 1% of the sample mean. The standard error for this estimate, however, is approximately ten times the parameter value. To compare the magnitudes, the estimated effect for those in the lowest third of the distribution is nearly forty times larger than those in the middle of the distribution.

Given that the results in table 3 show a great deal of heterogeneity based on the delivery method, it is important to verify that the pattern of heterogeneity in the estimates above is not simply reflecting differing distributions of women receiving c-sections versus vaginal deliveries. Therefore, the bottom two-thirds of table 5 contains separate estimates by thirds of the propensity score based on delivery type. The middle third of the table contains these estimates for children born by c-section. The 2SLS esti-

mates for these individuals are monotonically declining in groupings of the propensity score. The 2SLS estimate for the lowest third of the propensity score is large in magnitude—approximately 110% of the sample mean—and statistically significant at the 5% level. The parameter estimate for those in the top third of the propensity scores is negative, but the standard error is five times larger than the estimate. In this c-section sample, we can reject the null that the 2SLS coefficients on *Discharged Early* in the lowest and top-third propensity score samples are the same at a *p*-value of 0.06.

The bottom section of the table contains the estimates by thirds of the propensity score for children born by vaginal delivery. Although there are no statistically significant estimates, the pattern of results is similar to those for the full sample and for children born by c-section. For example, the parameter estimate for those in the bottom third is relatively large in magnitude (approximately 22% of the sample mean) and has a *p*-value of 0.12. This estimate is seventeen times larger than estimate for those in the middle third grouping. Perhaps as a result of the large standard errors, we cannot reject the null hypothesis that the 2SLS coefficients on *Discharged Early* in the lowest and top-third propensity score samples are the same at any conventional level.

Angrist et al. (1996) demonstrated that the instrumental variable estimates can be considered a local average treatment effect (LATE) where the 2SLS coefficient is the average effect of “treatment” for those whose behavior was altered as a result of receiving the instrument. It is therefore

no surprise that taking a weighted average of coefficients across the mutually exclusive subsamples discussed in tables 3 through 5 produces estimates similar to the full sample estimate in table 3. For example, roughly 80% of the sample has no problems at birth, and 20% have one or more problems. A weighted average (standard error) of the 2SLS effects across these two samples is 0.0051 (0.0027). The same numbers averaged across less than 5 pounds at birth and those who were 5 or more pounds is 0.0057 (0.0030). The average across thirds of the propensity score is 0.0057 (0.0027). The one exception is that a weighted average of treatment effects across c-section and vaginal births is 0.0072 (0.0031), which is much higher than the pooled sample estimate of 0.0047 (0.0028).

## VI. Conclusion

In recent years, a growing number of economic researchers have begun to worry about estimating heterogeneity in treatment effects. This literature has been spawned by an attempt to identify, in the words of Bitler, Gelbach, and Hoynes (2006), “what mean impacts miss.” Work by Heckman, Smith, and Clements (1997) and Abadie, Angrist, and Imbens (2002) on job training; Bitler, Gelbach, and Hoynes (2006) on welfare reform; Chandra and Staiger (2007) on heart attack treatments; and Basu et al. (2007) on breast cancer treatment all demonstrated that the average treatment effect estimated by standard multivariate models masks substantial heterogeneity. The apex of this literature is a series of papers by Heckman and Vytlačil and coauthors (Heckman & Vytlačil, 1999, 2005; Heckman, Tobias, & Vytlačil, 2003; Heckman, Urzua, & Vytlačil, 2006) that developed an econometric technique, local instrumental variables, designed to provide the entire distribution of marginal treatment effects.

In this paper, we join this literature and exploit the passage of state and federal early-discharge laws that increase the time newborns were in the hospital to examine the heterogeneity in the benefits of increased medical treatment intensity. As others have documented and the results in this paper further illustrate, the early-discharge laws have achieved an intermediate goal of increasing the length of postpartum stays. However, the ultimate goal of the laws was to improve patient health, and on this score, there is a mixed set of results. On average, being discharged early is estimated to have a small and marginally statistically significant impact on readmission rates. However, this small estimate masks considerable heterogeneity in the benefits of the more intensive treatment. For those in the lower third of the probability of being discharged early, a group with a priori the greatest medical need for care, the law is estimated to reduce the chance of a readmission by almost 30%. In contrast, there is little, if any, evidence of a benefit to greater treatment intensity for those with less medical need. The narrow conclusion is that an alternative law that applied to only those most in need of additional care can

generate the same health care outcomes with substantially reduced costs.

The broader conclusion is that more care must be taken when estimating the efficacy of medical intervention. The results here highlight the need to consider heterogeneity in expected benefits when estimating the efficacy of new treatments or policies. Dramatic increases in health expenditures have caused health systems around the globe to consider the efficacy of medications before including them in their coverage. For example, the National Institute for Health and Clinical Excellence (NICE) in the United Kingdom was established in 1999 to provide guidance to the National Health Service about which treatments are cost-effective and should be covered. Effectively implementing a system such as NICE, or the proposed comparative effectiveness research contained in the recent U.S. health system reform efforts, requires researchers to effectively confront and estimate the underlying heterogeneity in the benefits of different medical treatments. To the degree that heterogeneity is based on observable medical characteristics (as opposed to unidentifiable patient-level information), policymakers can develop cost-effective policies regarding the allocation of medical services. This should allow the achievement of the same or better health outcomes at a reduced cost.

## REFERENCES

- Abadie, Alberto, Joshua Angrist, and Guido Imbens, “Instrumental Variable Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings,” *Econometrica* 70:1 (2002), 91–117.
- Almond, Douglas, and Joseph J. Doyle, “After Midnight: A Regression Discontinuity Design in Length of Postpartum Hospital Stays,” *American Economic Journal: Economic Policy* 3 (2011), 1–34.
- Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin, “Identification of Causal Effects Using Instrumental Variables,” *Journal of the American Statistical Association* 91:434 (1996), 444–455.
- Baicker, Katherine, Kasey S. Buckles, and Amitabh Chandra, “Geographic Variation in Appropriate Use of Cesarean Section,” *Health Affairs* 25 (2006), w355–367.
- Baicker, Katherine, and Amitabh Chandra, “Medicare Spending, the Physician Workforce and Beneficiaries’ Quality of Care,” *Health Affairs* (2004), w184–197.
- Basu, Amrban, James J. Heckman, Salvador Navarro-Lozano, and Sergio Urzua, “Use of Instrumental Variables in the Presence of Heterogeneity and Self-Selection: An Application to Treatments of Breast Cancer Patients,” *Health Economics* 16:11 (2007), 1133–1157.
- Bitler, Marianne P., Jonah B. Gelbach, and Hilary W. Hoynes, “What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments,” *American Economic Review* 96:4 (2006), 989–1012.
- Bound, John, David A. Jaeger, and Regina M. Baker, “Problems with Instrumental Variables Estimation When the Correlation between the Instruments and the Endogenous Explanatory Variables Is Weak,” *Journal of the American Statistical Association* 90:430 (1995), 443–450.
- Braveman, Paula, Susan Egerter, Michelle Pearl, Kristen Marchi, and Carol Miller, “Early Discharge of Newborns and Mothers: A Critical Review of Literature,” *Pediatrics* 96:4 (1995), 716–726.
- Card, David, Carlos Dobkin, and Nicole Maestas, “Does Medicare Save Lives?” *Quarterly Journal of Economics* 124:2 (2009), 597–636.
- Chandra, Amitabh, and Douglas O. Staiger, “Productivity Spillovers in Health Care: Evidence from the Treatment of Heart Attacks,” *Journal of Political Economy* 115:1 (2007), 103–140.
- Cholesterol Treatment Trialists’ Collaborative, “Efficacy and Safety of Cholesterol-Lowering Treatment: Prospective Meta-Analysis of

- Data from 90,056 Participants in 14 Randomized Trials of Statins," *Lancet* 366:9493 (2005), 1267–1278.
- Cutler, David M., Mark McClellan, Joseph P. Newhouse, and Dahlia Remler, "Are Medical Prices Declining? Evidence from Heart Attack Treatments," *Quarterly Journal of Economics* 113:4 (1998), 991–1024.
- Cutler, David M., and Ellen Meara, "The Technology of Birth: Is It Worth It?" *Forum for Health Economics and Policy* 3:3 (2000), 1–35.
- Datar, Ashlesha, and Neeraj Sood, "Impact of Postpartum Hospital-Stay Legislation on Newborn Length of Stay, Readmission, and Mortality in California," *Pediatrics* 118:1 (2006), 63–72.
- Duggan, Mark G., and William N. Evans, "Estimating the Impact of Medical Innovation: A Case Study of HIV Antiretroviral Treatments," *Forum for Health Economics and Policy* 11:2 (2008), 1–37.
- Eaton, Antoinette Parisi, "Early Postpartum Discharge: Recommendations from a Preliminary Report to Congress," *Pediatrics* 107:2 (2001), 400–403.
- Evans, William N., Craig Garthwaite, and Heng Wei, "The Impact of Early Discharge Laws on the Health of Newborns," *Journal of Health Economics* 27:4 (2008), 843–870.
- Finkelstein, Amy, and Robin McKnight, "What Did Medicare Do? The Initial Impact of Medicare on Mortality and Out of Pocket Medical Spending," *Journal of Public Economics*, 92:7 (2008), 1644–1669.
- Fisher, Elliott S., David E. Wennberg, Therese A. Stukel, Daniel J. Gottlieb, F. L. Lucas, and Etoile L. Pinder, "The Implications of Regional Variation in Medicare Spending. Part 2: Health Outcomes and Satisfaction with Care," *Annals of Internal Medicine* 138:4 (2003), 288–298.
- Hammer, Scott M., Kathleen E. Squires, Michael D. Hughes, Janet Grimes, Lisa M. Demeter, Judith S. Currier, et al., "A Controlled Trial of Two Nucleoside Analogues Plus Indinavir in Persons with Human Immunodeficiency Virus Infection and CD4 Cell Counts of 200 per Cubic Millimeter or Less," *New England Journal of Medicine* 337:11 (1997), 725–733.
- Heckman, James J., Jeffrey Smith, and Nancy Clements, "Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts," *Review of Economic Studies* 64:4 (1997), 487–535.
- Heckman, James J., Justin L. Tobias, and Edward J. Vytlačil, "Simple Estimators for Treatment Parameters in a Latent Variable Framework," this REVIEW 85:3 (2003), 748–755.
- Heckman, James J., Sergio Urzua, and Edward J. Vytlačil, "Understanding Instrumental Variables in Models with Essential Heterogeneity," this REVIEW 88:3 (2006), 389–432.
- Heckman, James J., and Edward J. Vytlačil, "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects," *Proceedings of the National Academy of Sciences* 96:8 (1999), 4730–4734.
- "Structural Equations, Treatment Effects, and Econometric Policy Evaluation," *Econometrica* 73:3 (2005), 669–738.
- Hyman, David A., "Drive-Through Deliveries: Is Consumer Protection Just What the Doctor Ordered?" *North Carolina Law Review* 78:1 (1999), 5–100.
- Kravitz, Richard, Naihua Duan, and Joel Braslow, "Evidence-Based Medicine, Heterogeneity of Treatment Effects, and the Trouble with Averages," *Milbank Quarterly* 82:4 (2004), 661–687.
- Law, Malcolm R., Nicholas L. Wald, and A. R. Rudnicka, "Quantifying Effect of Statins on Low Density Lipoprotein Cholesterol, Ischemic Heart Disease, and Stroke: Systematic Review and Meta-Analysis," *British Medical Journal* 326:7404 (2003), 1423–1429.
- Liu, Zhimei, William H. Dow, and Edward C. Norton, "Effect of Drive-Through Delivery Laws on Postpartum Length of Stay and Hospital Charges," *Journal of Health Economics* 23:1 (2004), 129–155.
- Madden, Jeanne M., Stephen B. Soumerai, Tracy A. Lieu, Kenneth D. Mandl, Fang Zhang, and Dennis Ross-Degnan, "Effect of a Law against Early Postpartum Discharge on Newborn Follow-Up, Adverse Events, and HMO Expenditures," *New England Journal of Medicine* 347:25 (2002), 2031–2038.
- "Length-of-Stay Policies and Ascertainment of Postdischarge Problems in Newborns," *Pediatrics* 113:1 (2004), 42–49.
- Madlon-Kay, Diane J., and Terese A. DeFor, "Maternal Postpartum Health Care Utilization and the Effect of Minnesota Early Discharge Legislation," *Journal of the American Board of Family Practice* 18:4 (2005), 307–311.
- Meara, Ellen, Uma R. Kotagal, Harry D. Atherton, and Tracy A. Lieu, "Impact of Early Newborn Discharge Legislation and Early Follow-up Visits on Infant Outcomes in a State Medicaid Population," *Pediatrics* 113:6 (2004), 1619–1627.
- Newhouse, Joseph P., and the Insurance Experiment Group, *Free for All? Lessons from the RAND Health Insurance Experiment* (Cambridge, MA: Harvard University Press, 1993).
- Sakala, Carol, and Maureen Corry, "Evidence Based Maternity Care: What It Is and What It Can Achieve" (New York: Milbank Memorial Fund, 2008).
- Schwarzc, Sandra K., Ling Chin Hsu, Eric Vittinghoff, and Michael H. Katz, "Impact of Protease Inhibitors and Other Antiretroviral Treatments on Acquired Immunodeficiency Syndrome Survival in San Francisco, California, 1987–1996," *American Journal of Epidemiology* 152:2 (2000), 178–185.
- Staiger, Douglas, and James H. Stock, "Instrumental Variables with Weak Instruments," *Econometrica* 63:3 (1997), 557–586.
- Udom, Nduka, and Charles L. Betley, "Effects of Maternity-Stay Legislation on 'Drive-Through Deliveries,'" *Health Affairs* 17:5 (1998), 208–215.
- Wennberg, John E., Elliott S. Fisher, David C. Goodman, and Jonathan S. Skinner, *Tracking the Care of Patients with Severe Chronic Illness: The Dartmouth Atlas of Health Care 2008* (Lebanon, NH: Dartmouth Institute for Health Policy and Clinical Practice, Center for Health Policy Research, 2008).