
**The Australian National University**

# Centre for Economic Policy Research

## *DISCUSSION PAPER*

# Do Health Care Report Cards Cause Providers to Select Patients and Raise Quality of Care?

Yijuan Chen and Juergen Meinecke

DISCUSSION PAPER NO. 657

January 18, 2012

**ISSN: 1442-8636**
**ISBN: 978-1-921693-39-7**

# DO HEALTH CARE REPORT CARDS CAUSE PROVIDERS TO SELECT PATIENTS AND RAISE QUALITY OF CARE?

YIJUAN CHEN AND JUERGEN MEINECKE*

ABSTRACT. We exploit a brief period of asymmetric information during the implementation of Pennsylvania's "report card" scheme for coronary artery bypass graft surgery to test for improvements in quality of care and selection of patients by health care providers. During the first three years of the 1990s, providers in Pennsylvania had an incentive to bias report cards by selecting patients strategically, with patients having no access to the report cards. This dichotomy enables us to separate providers' selection of patients from patients' selection of providers.

Using data from the Nationwide Inpatient Sample, we estimate a non–linear difference–in–differences model and derive asymptotic standard errors. The mortality rate for bypass patients decreases by only 0.05 percentage points due to the report cards, which we interpret as evidence that quality of bypass surgery did not improve (at least in the short–term) nor did patient selection by providers occur. Our timing, estimation, and asymptotics are readily applicable to many other report card schemes.

*Date*: January 18, 2012

* Corresponding author.

We are grateful to two anonymous referees for their comments and suggestions. We thank Mark Pauly, Jongsay Yong, Pedro Gomis Porqueras, Terence Cheng, and Peter Sivey for helpful suggestions. We also appreciate the comments from participants at the 2010 Labour Econometrics Workshop at Deakin University, the Second Australasian Workshop on Econometrics and Health Economics 2011, and the micro brownbag at the Australian National University. All remaining errors are our own.

*Keywords:* health care report cards; provider moral hazard; quality improvement; difference–in–differences estimation.

*JEL classification:* C23, D82, I18.

## 1. INTRODUCTION

Quality "report cards" for health care providers are pervasive in health policy. Two common and open questions regarding the intended and unintended effects of report cards are: Do report cards lead to genuine improvements in quality of care and thus patient outcomes? Do report cards induce providers to select patients? It is difficult in empirical analysis to address these questions separately from each other. We propose to answer them jointly. Perhaps the most–studied report cards are those for providers of coronary artery bypass graft (CABG) surgery in the United States (U.S.). New York and Pennsylvania were the first states to publish results in the early 1990s. If providers of CABG surgery, in response to the introduction of report cards, indeed improved quality of care while simultaneously engaging in strategic patient selection then we would expect, keeping patient characteristics fixed, a noticeable reduction in mortality outcomes for bypass patients.[1] If however, as we find here, the change in mortality is sufficiently close to zero then we conclude that neither quality improvements nor patient selection occurred.

Our empirical analysis contributes to the literature in two ways. First, report cards can induce selection behavior not only by providers but also by patients (which is a major reason for their existence). Research to date has ignored the possibility that once report cards are published, both providers and patients could engage in mutual selection. For example, in the most influential study to date, Dranove et al. (2003) use the publication date of CABG report cards for New York and Pennsylvania as the cut–off point in a difference–in–differences analysis. Using a comprehensive longitudinal Medicare claims data set, combined with data from the American Hospital Association, they show that the average illness severity of bypass patients decreases by 3.47%–5.30% due to the introduction of CABG report cards, concluding that providers shift treatment from sicker patients to healthier ones. However, it is not clear to what extent the decrease in illness severity is due to selection of patients by providers or selection of providers by patients.

We propose a simple remedy that isolates selection of patients by providers. Figure 1 presents the particular timing of CABG report cards as implemented in Pennsylvania in the early 1990s. Before 1990 (period 0), there were no CABG report cards in Pennsylvania. Between January 1990 and November 1992 (period 1), the Pennsylvania Health Care Cost Containment Council (PHC4) collected data on mortality outcomes and the characteristics of bypass patients. During that period, no report cards were published, making patient selection based on report cards impossible. After November 1992 (period 2), the PHC4 published the report cards and patients were able to access the results.

Our remedy in order to isolate provider selection of patients from patient selection of providers, consists of ignoring period 2 and focusing on a comparison of period 1 and period 0 only.[2] It is during period 1 that providers may have an incentive to select patients strategically, while at

[1] We use the terms "CABG" and "bypass" synonymously. Two other terms that we use interchangeably are "provider" and "hospital".

[2] The timing suggested here is distinct from what is commonly referred to as Ashenfelter's (1978) dip, which refers to a change in individuals' behavior in anticipation of a treatment. Here, in period 1, the treatment (collection of report card data) is already in progress. Health care providers have a real incentive to adjust their behavior during this period.

the same time the report card cannot have influenced patients' choice of providers.[3] Providers in Pennsylvania were aware that data was being collected during period 1. The PHC4 mandated that hospitals use an automated system, called MedisGroups, for collecting and analyzing clinical data and, as a result, hospitals submitted data to the PHC4 every quarter. At the same time, hospitals also knew, at every step of the process, the plans and intentions of the PHC4. The council consisted of 21 members with representatives from business, labor, hospitals, insurers, medicine, health maintenance organizations, and consumer groups. The PHC4 took an altogether consultative and collaborating approach in working with the medical community (see Sessa and Bentley 1992 and Sirio et al. 1996).

[Figure 1 about here.]

A second innovation of our paper concerns the estimation. We estimate a non–linear difference–in–differences model under an asymptotic framework in which the number of groups is small but group sizes are large (few large groups). The non–linear framework is necessary because the dependent variable is a mortality dummy (equal to 1 if a patient dies during hospitalization and 0 otherwise) and the mortality incidence for bypass patients is a low probability event—less than 6% of patients who undergo CABG surgery die. Therefore, fitting a non–linear model avoids the problem of predictions outside the unit interval that linear probability models face. Further, the report cards are constructed based on logistic models, lending a non–linear estimation framework further validity.

An asymptotic framework of few large groups is required because the policy variable of interest—whether or not a report card policy is in place—does not vary at the individual patient level, but at the state/year level and we only observe few such groups (altogether 52 combinations of states and years). Moulton (1990) highlights the threat to inference when explanatory variables—particularly the policy variable of interest—are constant among members of a group (for example, a state during a certain year). Donald and Lang (2007) reiterate this point and propose a simple two–step procedure that achieves correct inference when the number of groups is small and the group size (that is, the number of patients within each state/year group) is large. Wooldridge (2006) proposes a two–stage minimum distance estimator, which, under few large group asymptotics, is expected to yield stronger inference than the Donald and Lang approach. For our estimation, Wooldridge's estimation strategy is preferable for two further reasons: it is adaptable to non–linear models and it makes inference of average partial effects possible—an important parameter in non–linear models. We derive the asymptotic properties of the efficient non–linear two–stage minimum distance estimator and the average partial effects separately in the Appendix.

How can providers select patients? They can pursue two broad strategies: shifting bad–risk patients out of CABG surgery and/or shifting good–risk patients into CABG surgery. In both cases, the observed mortality rate for CABG surgery will decrease as a result. As in Dranove et al. (2003), we assume that there exists a non–degenerate subset of patients admitted

[3]We do not rule out the possibility that patients choose providers (for example, through word–of–mouth or through recommendation by general practitioners). This type of selection is allowed to be present during all periods: 0, 1, and 2. The only assumption we need to make is that this type of selection is not affected by the report cards.

to each hospital with heart problems qualifying for both CABG surgery and a different procedure (for example, percutaneous coronary angioplasty). Therefore, providers have some degree of freedom in choosing the clinical procedure for a strict and non–empty subset of all heart patients. As such, the introduction of the report card policy affects providers' decisions monotonically: the likelihood that a bad–risk patient receives a substitute treatment rather than CABG is weakly increasing (or non–decreasing) due to the report cards; the likelihood for a good–risk patient is weakly decreasing (or non–increasing).

Providers cannot effectively select patients into or out of CABG surgery based on their observable characteristics because report cards are risk–adjusted. For example, providers that shift CABG surgery from older to younger patients will not bias report cards in their favor because age is controlled for in the construction of the report cards. Instead, rational providers have to use information that is not known by the report card authority. For example, a provider observes that a patient suffers from hypertension and that the patient has a family history of hypertension. While the hypertension itself is not a reason for avoiding the patient (because the report card authority risk–adjusts for hypertension), the fact that there is a known family history could convince the provider to avoid the patient (and offer a substitute treatment instead).

This selection of patients by providers could lead to econometric complications. As selection could change the composition of the patient pool for CABG surgery, we need to assume that observable patient characteristics do not change due to the report cards. Technically, the private (and unobserved by the report card authority) information that providers use for the selection of patients cannot be correlated with the observable characteristics of the patients. We provide empirical evidence that this indeed is the case. Patient characteristics are mostly unaffected by the report cards, and in the few cases in which they are affected, the resulting bias works against us, affording our estimation results a conservative interpretation.

Using the Nationwide Inpatient Sample (NIS) data from the U.S. Agency for Health Care Policy and Research and comparing period 0 to period 1, we do not find any evidence of improvement in the quality of care for or provider selection of bypass patients. Average illness severity of patients in Pennsylvania decreases slightly due to the report cards, but not significantly. Translated into mortality rates, we find the probability that Pennsylvanian patients undergoing CABG surgery will die decreases by at most only 0.05 percentage points under the report card scheme. We also explore whether hospital or patient heterogeneity played a role in explaining selection. In both cases, we find small effects that are statistically indistinguishable from zero.

Providers may not select patients for at least three reasons. First, the report cards trade off mortality outcomes and patient volume measures. Chen (2011), using a signaling model, shows that low–quality providers cannot successfully imitate high–quality providers and as a result do not participate in patient selection. Second, providers may not have access to valuable private information that would enable them to select patients effectively. The report card authority uses a rich data set that may function as a sufficient statistic for any private information that providers possess. Third, providers may be altruistic. Theoretical models of impure competition predict that non–profit hospitals compete for public goodwill by supplying charity care. Empirical evidence for Pennsylvania supports this view (see Rosko 2004).

The lack of quality improvements in bypass surgery may be explained by time. Given the short time frame available for our analysis, it seems possible that any steps undertaken by hospitals to improve quality of care had not manifested in lower mortality rates. Further, it may not be optimal for all hospitals to invest in technologies that improve the quality of bypass surgery. If the quality difference between hospitals is sufficiently large then a low–quality hospital does not have an incentive to improve the quality of bypass surgery because it will never be able to imitate a high–quality hospital (see Gravelle and Sivey 2010).

The rest of the paper is organized as follows. Section 2 briefly explains the background of Pennsylvania's hospital report card scheme and describes and summarizes the NIS data. Section 3 lays out the estimation framework, while Section 4 contains the results, and Section 5 provides a discussion. Section 6 concludes. The Appendix gives the details of the two–stage estimation and derives asymptotic results for estimators, including the average partial effects.

## 2. Background and Data

In 1986, the state of Pennsylvania passed a law (under Act 89) that established the Pennsylvania Health Care Cost Containment Council (PHC4), with one of its main responsibilities being "to collect, analyze and make available to the public data about the cost and quality of health care in Pennsylvania." In November 1992, the PHC4 published their first report on CABG surgery. The report provides risk–adjusted information on the health outcomes (in particular mortality) of bypass patients at hospital and surgeon level. One motivation of collecting health data and issuing report cards, as stated in PHC4 (1998), is the incentive for providers "to take steps to improve the overall quality of bypass surgery."

The report cards are supposed to evaluate the quality of a provider (or surgeon). Quality is measured by the mortality rate of a provider's bypass patients controlling for risk factors. The PHC4 research panel suggests running a logistic regression, in which the dependent variable is a dummy equal to 1 if a patient dies during his hospital stay, on a set of covariates that fall into two categories: demographic and clinical. Demographic variables include age and gender, while clinical variables include diabetes or if a person had a cardiogenic shock (for further details see PHC4 1998).

Given the results from the non–linear estimation, the report cards summarize, for each provider, whether the actual mortality rate falls below or above the predicted mortality rate (adjusted for standard errors). Therefore, a provider can either be below, above or in line with the expected mortality rate. This quality evaluation is also performed for individual surgeons.[4]

We use the U.S. Agency for Health Care Policy and Research's NIS Release 1 data, which span from 1988 to 1992. Designed to approximate a 20 percent sample of U.S. community hospitals each year, the NIS is a stratified probability sample of hospitals in the participating states, with sampling probabilities proportional to the number of community hospitals in each stratum.[5]

[4]Epstein (2006) provides an extensive discussion of the background.

[5]As defined by the American Hospital Association, community hospitals are "all nonfederal, short-term, general and other specialty hospitals, excluding hospital units of institutions". The strata are constructed based on five hospital characteristics: ownership/control, bed size, teaching status, urban/rural location, and U.S. region.

Table 1 lists the combinations of states and years for which we have data. We observe eleven states—eight states for the years 1988–1992 and three states (including Pennsylvania) for the years 1989–1992. For the year 1988, we have information on 759 hospitals, with an overall number of in–patient stays of more than five million. For each year between 1989 and 1992 we have information on more than 850 hospitals with an overall number of in–patient stays of more than six million.

We narrow the sample to patients aged 65–99. This restriction has the advantage of a majority of patients being covered by Medicare, rather than by a private insurance or health maintenance organization (HMO), which may restrict the flexibility of providers in assigning treatment (due to pre–negotiated agreements). This in turn would reduce providers' opportunities for selection of patients. In the subset of Pennsylvanian bypass patients aged 65-99, more than 92% are covered by Medicare during the years 1990–1992.[6]

[Table 1 about here.]

We further narrow the sample to patients who received one of the following nine principal diagnoses: acute myocardial infarction, coronary atherosclerosis, non–specific chest pain, pulmonary heart disease, other and ill-defined heart disease, conduction disorders, cardiac dysrhythmias, cardiac arrest and ventricular fibrillation, or congestive heart failure (non–hypertensive). These nine principal diagnoses cover almost 98% of all bypass patients within that age range. Overall, our sample contains 952,200 heart patients.

Table 2 contains descriptive statistics of patient characteristics for the control states. About 52% of patients are male and 85% are white, while the average age is 75 years. Almost half of the sample entered the hospital as an emergency admission, a third as urgent, and 15% as elective.

[Table 2 about here.]

The table also lists descriptive statistics on three principal clinical procedures that we use in the estimation: CABG, percutaneous coronary angioplasty (PTCA), and diagnostic cardiac catheterization (CATH). Only those patients who received CABG as the principal procedure are the subject of the report cards. Following Dranove et al. (2003), we regard the PTCA procedure as the primary substitute treatment for CABG and the CATH procedure as a diagnostic procedure that complements CABG.

Table 2 shows that the incidence of CABG surgery in the control states starts out at 11.2% in 1988 and slowly rises to 12.9% in 1992. The probability of receiving PTCA increases at a markedly steeper rate than for CABG, equaling 7.5% in 1988 and growing to 12.5% by 1992. The CATH procedure is more common, at 16.7% in 1988, increasing slowly to 18.0% in 1992.

Looking at mortality rates, the table reveals that heart patients are less likely to die over time across all three principal procedures. The mortality rates of patients who receive CABG surgery drop from 5.8% in 1988 to 4.5% by 1992. For PTCA and CATH the numbers are similar, both starting out at 2.1% and declining to 1.6% and 1.8%.

[6]Ideally, we would like to restrict the sample to only Medicare patients. However, the NIS data for Pennsylvania does not report payor information for the year 1989—the only pre–treatment year available for that state. By restricting the sample to over 65–year–olds, we minimize the contamination that may result from private insurance or HMO considerations.

Table 3 contains the descriptive statistics for Pennsylvania. Demographically, Pennsylvania is similar to the control states: about half of patients are male and the average age is around 74 years (information on race is not available). At about 59%, Pennsylvania does have a higher number of emergency admissions than the control states. This holds for before and after the treatment (with collection of report card data starting in 1990). Conversely, the number of urgent admissions is lower at around 20%–22%. Comparing the sum of emergency and urgent admissions between the control states and Pennsylvania, the difference is small. The discrepancy in emergency and urgent admissions when looked at individually may be due to different reporting requirements. The percentage of elective admissions in Pennsylvania grows slowly over time, from 15.3% in 1989 to 17.6% by 1992.

[Table 3 about here.]

Regarding the clinical variables, the levels and trends in Pennsylvania are arguably similar to the control states. Table 3 shows that the incidence of CABG surgery in Pennsylvania begins at 10.8% in 1989 (almost exactly at the level of the control states) and rises to 13.7% by 1992 (0.8 percentage points above the control states). The probability of receiving PTCA increases from 7.1% in 1989, growing to 11.2% by 1992. The probability of receiving CATH rises from 15.4% in 1989 to 18.1% by 1992.

Looking at mortality rates, the difference between Pennsylvania and the control states is, again, not large. CABG mortality in Pennsylvania begins slightly higher at 5.8% in 1989 (compared to 5.4%) and decreases to 4.7% in 1992 (compared to 4.5%). For PTCA, there is a drop from 1.7% to 1.0% between 1989 and 1990 and an increase back to 1.7% by 1992. For CATH there is a slight decrease from 1.8% to 1.5% between 1989 and 1992.

Contrasting Tables 2 and 3 offers a preview of our main estimation results. We do not find, when looking at aggregate numbers at the state level, that there is a significant drop in the mortality rates of bypass patients in Pennsylvania during the years 1990 to 1992 relative to the control states. Finding such a decline would be consistent with both improvements in patient care and hospital selection of patients. Instead we find that both control states and Pennsylvania follow similar downward trends in mortality for CABG, but also for PTCA and CATH. We conclude that there is no sufficient evidence for quality improvement or provider selection of patients. The estimation results below will formally confirm this view.

## 3. ESTIMATION

3.1. **Patient Illness Severity and Mortality.** We estimate the effect of a hospital report card policy on the illness severity and mortality of bypass patients. While Dranove et al. (2003) use total inpatient hospital expenditures for the year prior to admission as a proxy measure for illness severity, we follow the direction proposed by Elixhauser et al. (1998) who construct patient illness severity from administrative data. As the measure of mortality, we use actual in–hospital mortality (that is, whether or not a patient died during hospitalization). The PHC4 used the same dependent variable in the construction of its report cards.

Consider the following four–level hierarchical model for the conditional mortality probability

$$\Pr(m_{ihst} = 1 | c_{hst}, z_{ihst}) = F\left(\mu + a_s + b_t + p_{st}\beta + c'_{hst}\phi_s + z'_{ihst}\gamma_s\right), \tag{3.1}$$

where $c_{hst}$ is a vector of covariates that vary at the hospital level, $z_{ihst}$ is a vector of covariates that vary at the individual level, $m_{ihst}$ is a mortality dummy variable that equals 1 if a patient died during hospitalization and 0 otherwise, and $p_{st}$ is a policy dummy variable that equals 1 if the state corresponds to Pennsylvania during 1990–1992 and 0 otherwise. The parameter $\mu$ is a constant, while $a_s$ and $b_t$ are fixed effects at the state and year levels. The coefficient $\beta$ represents the effect of the hospital report card policy on the dependent variable. Our objective is to estimate $\beta$. We restrict the sample to all patients admitted to hospital between January 1988 and October 1992 (report cards in Pennsylvania were published in November 1992) who received CABG surgery.

The estimation needs to reflect that the policy variable of interest—whether or not report cards are in place—does not vary at the individual level but only across states over time. At the same time we observe only eleven states—eight of them for the years 1988–1992 and three of them for the years 1989–1992. This gives 52 state/year groups altogether. A difference–in–differences analysis based on only 52 observations would result in estimates with large standard errors. However, within each state/year group there are a large number of inpatient stays. Therefore, we apply an asymptotic framework of few large groups.

Moulton (1990) was the first to point out the threat to inference when explanatory variables (particularly the policy variable of interest) are constant among members of a group (for example, in a state during a certain year). Donald and Lang (2007) reiterate this point and propose a simple two–step procedure that achieves correct inference when the number of groups is small and the group sizes (that is, the number of patients within each state/year group) are large. Wooldridge (2006) proposes a two–stage minimum distance estimator, which, under few large group asymptotics, is expected to yield stronger inference than the Donald and Lang approach. For our estimation, Wooldridge's estimation strategy is preferable for two further reasons: it can be applied in a straightforward manner to non–linear models and it has the advantage that inference on average partial effects, a common parameter of interest in non–linear models, can be derived readily.[7]

For the function $F(\cdot)$ we consider both the identity function (which yields the linear probability model) and the cumulative distribution function (cdf) of the logistic distribution. We include the linear probability model because it lends the two–stage estimation procedure an intuitive interpretation. Our main results hold for both the linear and non–linear specifications. However, we believe that the logistic specification is preferable because it guarantees conditional probabilities within the unit interval and because the actual construction of the report cards by the PHC4 was based on the logit model.

*First–Stage Estimation.* The estimation proceeds in two stages (all technical details on estimation and inference are deferred to the Appendix). First, define

$$\delta_{st} := \mu + a_s + b_t + p_{st}\beta, \tag{3.2}$$

[7]Hansen (2007a) derives properties of the standard covariance matrix estimator in panel data models. Hansen (2007b) presents generalized least squares estimation of a multilevel panel data model with autocorrelation. Both papers cover linear models only and do not extend to asymptotics in which the number of groups is small and the size of each group is large.

and rewrite equation (3.1) more compactly as

$$\Pr(m_{ihst} = 1 | c_{hst}, z_{ihst}) = F\left(\delta_{st} + c'_{hst}\phi_s + z'_{ihst}\gamma_s\right). \tag{3.3}$$

The last equation summarizes the strategy for the first–stage estimation: pooling data across time for each state, we obtain within–estimates of $\delta_{st}$ (along with $\phi_s$ and $\gamma_s$). For each state/year group we observe a rich set of demographic and clinical characteristics. Table 4 lists the variables included in $c_{hst}$ and $z_{ihst}$.

Regarding the clinical characteristics included in $z_{ihst}$, we translated a large array of ICD-9-CM procedure codes into meaningful comorbidities. We borrow from Elixhauser et al. (1998), who, based on typical clinical discharge abstract data, provide a mapping from ICD-9-CM procedure codes to comorbidities. For example, ICD-9-CM procedure code 196.0 can be translated to the comorbidity metastatic cancer.

[Table 4 about here.]

The mapping from ICD-9-CM procedure codes to comorbidities is not injective. Different ICD-9-CM procedure codes can map to the same comorbidity. Altogether, Elixhauser et al. (1998) define 30 different comorbidities, most of which rarely occur in bypass patients. We focus on the five comorbidities that occur most frequently in bypass patients: peripheral vascular disorders, hypertension, chronic pulmonary disease, diabetes (uncomplicated), and fluid and electrolyte disorders.[8] In addition, we include a dummy variable if a patient received a diagnosis of acute myocardial infarction (AMI or heart attack) and a dummy variable if complications occurred during surgery (as defined in Elixhauser et al. 1998).

*Second–Stage Estimation.* The parameter $\delta_{st}$ has an intuitive interpretation when $F(\cdot)$ is the identity function. It gives the expected mortality rate for bypass patients in state $s$ during year $t$ when $c_{hst} = 0$ and $z_{ihst} = 0$. We can therefore regard $\delta_{st}$ as a measure of average illness severity for bypass patients in state $s$ during year $t$. This interpretation of $\delta_{st}$ as a measure of average illness severity also applies when $F(\cdot)$ is the cdf of the logistic distribution (because of the monotonicity of the cdf). However, in this case we cannot interpret $\delta_{st}$ as a probability because the model is non–linear. Consequently, we alternatively derive average partial effects below.

With estimates of $\delta_{st}$ in hand, we use efficient minimum distance estimation in the second stage to obtain $\hat{\beta}$ based on equation (3.2) (see Appendix). For the linear model, we interpret the estimate of $\beta$ as the causal effect of the report card policy on the average illness severity of Pennsylvanian bypass patients. For the non–linear model, we can give a tangible interpretation to the causal effect of hospital report cards by calculating the average partial effect at the patient level. Our two–stage estimation has the advantage that we can plug the estimate $\hat{\beta}$ back into equation (3.1) to calculate the effect of hospital report cards on the average mortality rates of patients. Specifically, we answer the following two questions:

(A) For Pennsylvanian bypass patients who were admitted to a hospital in 1989 (before report cards), by how much would their average mortality rate have changed had they been admitted during the years 1990–1992 instead?

[8] Results do not change qualitatively and only slightly quantitatively when we include additional comorbidities.

(B) For Pennsylvanian bypass patients who were admitted to a hospital during the years 1990–1992 (during report card data collection), by how much would their average mortality rate have changed had they been admitted in 1989 instead?

Formally, both questions are addressed by the following objects:

$$\begin{aligned}\text{APE}_A &:= \text{AMR}_{01} - \text{AMR}_{00}\\ \text{APE}_B &:= \text{AMR}_{11} - \text{AMR}_{10},\end{aligned}$$

where the average mortality rates (AMR) are defined as

$$\begin{aligned}\text{AMR}_{00} &:= E\left[F\left(\mu + a_s + b_t + c_{hst}'\phi_s + z_{ihst}'\gamma_s\right)|s = \text{PA}, t < 1990\right],\\ \text{AMR}_{01} &:= E\left[F\left(\mu + a_s + b_t + c_{hst}'\phi_s + \beta + z_{ihst}'\gamma_s\right)|s = \text{PA}, t < 1990\right],\\ \text{AMR}_{10} &:= E\left[F\left(\mu + a_s + b_t + c_{hst}'\phi_s + z_{ihst}'\gamma_s\right)|s = \text{PA}, t \geq 1990\right],\\ \text{AMR}_{11} &:= E\left[F\left(\mu + a_s + b_t + c_{hst}'\phi_s + \beta + z_{ihst}'\gamma_s\right)|s = \text{PA}, t \geq 1990\right].\end{aligned} \tag{3.4}$$

The effects $\text{APE}_A$ and $\text{APE}_B$ correspond to questions (A) and (B) above and are the average partial effects of the report card policy on mortality rates. We estimate $\text{APE}_A$ and $\text{APE}_B$ not only for the subset of patients who received CABG surgery, but also for the subset of patients who received PTCA or CATH.

*Derivation of Standard Errors.* For all estimators in the first and second stage we derive analytical standard errors via the delta method in the Appendix. In particular, we provide standard errors for $\hat{\delta}_{st}$ for all $s$ and $t$ obtained from the first–stage logit estimation of equation (3.3). For the second–stage minimum distance estimation we adjust the standard errors because the dependent variable, based on equation (3.2), is constructed from the estimates $\hat{\delta}_{st}$. Further, to conduct inference on the average partial effects, $\text{APE}_A$ and $\text{APE}_B$, we derive the asymptotic expansions of all $\widehat{\text{AMR}}_{pq}$ for $p, q \in \{0,1\}$, which enables us to calculate the exact asymptotic distributions for the average partial effects.

3.2. **Patient Characteristics.** The first–stage or within–estimations based on equation (3.3) are valid only if no unobserved heterogeneity feeds through to patient characteristics. To bias report cards in their favor, providers cannot select patients on observable characteristics: rational providers know that report cards will be "risk–adjusted". Providers can select patients only based on characteristics that are unobserved by the report card authority (and the econometrician). If the unobserved selection rule that providers apply affects observable characteristics, then the regressions based on equation (3.3) will yield inconsistent estimates.

We test whether observable patient characteristics change in response to the report cards by adapting the estimation framework from the previous subsection. We regress each independent variable contained in the vector of patient characteristics $z_{ihst}$ on a constant, a state dummy, a time dummy, and the policy dummy. Formally, let $z_{ihst}^r$ be the $r$-th element of the vector $z_{ihst}$. For example, using Table 4, the element $z_{ihst}^1$ corresponds to the first age dummy. Then, for each

independent variable (that is, for each $r$), we estimate

$$\Pr(z^r_{ihst} = 1) = \mu + a_s + b_t + p_{st}\beta = \delta_{st}, \tag{3.5}$$

where all definitions are as before. The estimation of equation (3.5) proceeds as before. In the first stage, we estimate $\delta_{st} := \mu + a_s + b_t + p_{st}\beta$ non–parametrically, using a linear probability model and a constant term as the only explanatory variable. In the second stage, we estimate $\beta$ via efficient minimum distance estimation. If providers select patients on unobservables and if these unobservables are correlated with patient characteristics then we would expect estimates of $\beta$ in equation (3.5) that are significantly different from zero.

## 4. RESULTS

*Main Findings.* Table 5 shows the results for the patient characteristics regressions from the preceding subsection. Patient characteristics are unaffected by the report cards in 10 out of 14 cases. Only for the characteristics 'urgent', 'complications', and for the last comorbidity does the 95% confidence interval not include zero. However, in all of these three cases the effect of the observable characteristics biases the results for the conditional mortality rates against us.

[Table 5 about here.]

For example, the report card policy is associated with a decrease in clinical complications. This could result from genuine improvements in quality of care or provider selection (or both). In any case, average illness severity and mortality should decline as consequence. However, we do not find a significant decline.

We now present the results of the mortality regressions from equations (3.2) and (3.3). First, we focus only on patients who received CABG surgery. We compute $\hat{\beta}$, the average mortality rates $\widehat{\text{AMR}}_{pq}$ for $p,q \in \{0,1\}$ and the average partial effects $\text{APE}_A$ and $\text{APE}_B$ based on that subsample of patients. We then repeat the two–stage estimation for the subsample of PTCA and CATH patients.

Table 6 contains the second–stage estimation results for $\beta$ from equation (3.2): the causal effect of the report card policy on the measure of illness severity $\delta_{st}$. We do not find any causal effect of report cards on the illness severity of CABG patients. For the linear probability model, the average mortality rate decreases by 0.0008 due to the report cards. The effect is small and not statistically significant. Likewise, for the logistic model we find a non–significant effect. Further, we do not detect any significant effect of report cards on the average illness severity of PTCA or CATH patients.

[Table 6 about here.]

Table 7 contains the estimation results for the AMR and the average partial effects, $\text{APE}_A$ and $\text{APE}_B$. Numbers for the AMR are reported in percent and numbers for the APR are in percentage points. We do not detect any significant decrease in mortality rates. For Pennsylvanian bypass patients during the year 1989 (before report cards), the average mortality rate was $\text{AMR}_{00} = 5.84\%$. Had the report card policy been in place for those patients, the average mortality rate would have decreased to $\text{AMR}_{01} = 5.79\%$. This difference of $\text{APE}_A = -0.05$ percentage points is small and insignificant.

[Table 7 about here.]

Conversely, we calculate an average mortality rate of $\text{AMR}_{11} = 5.00\%$ for Pennsylvanian bypass patients during the years 1990–1992. Had the report card policy not been in place for those patients, the average mortality rate would have increased to $\text{AMR}_{10} = 5.05\%$. This difference of $\text{APE}_B = -0.05$ percentage points is also small and insignificant.

Qualitatively similar results hold for both the PTCA and CATH procedures. The AMR decline marginally but never significantly. If hospitals select low–risk patients for CABG surgery, we would expect the average mortality rates for either PTCA or CATH (or both) to go up.

Combining all results, we cannot reject the hypothesis that mortality rates remained unchanged due to the introduction of hospital report cards in Pennsylvania. An intended effect of report cards is an improved level of care for bypass patients, while an unintended effect is that hospitals select patients strategically. Both effects should lower mortality rates for CABG patients. We provide strong empirical support that both effects are statistically indistinguishable from zero.

*Hospital Heterogeneity.* It is possible that we do not detect any significant selection effects in the analysis because we define the treatment group too broadly. It is possible that not all hospitals in Pennsylvania had an incentive to select patients. If only a subset of hospitals select patients, then our aggregate analysis could fail in detecting selection.

What type of hospitals had the strongest incentive to select patients? Supposedly, private for–profit hospitals are exposed to the highest degree of competition and may be more inclined to select patients. However, virtually all Pennsylvanian hospitals are organized as private non–profit entities; they are perfectly homogeneous along the ownership–dimension. Therefore, exploiting variation in ownership status is not a feasible approach for studying the effects of hospital heterogeneity on selection behavior.

Instead, we focus on the subset of hospitals to which the report cards posed the most immediate threat: those hospitals that prior to the collection of report card data had exceptionally high CABG mortality rates. That subset of hospitals would have had the strongest incentive to select patients. For each state, a hospital is classified as high–mortality if its CABG mortality belongs to the worst quartile of hospitals during the year 1989.[9] We use the year 1989 because for Pennsylvania this is the only pre–treatment year available in the data. For the worst quartile of hospitals, we then proceed with the same two–stage estimation described above to test for patient selection. We then redo this analysis also for the best quartile of hospitals (lowest 25% CABG mortality) for additional robustness.

To make estimation feasible, we need to make a few small adjustments to the estimation and data. First, we include hospital fixed effects, which are a more parsimonious way of controlling for hospital characteristics when we conduct our estimation on a relatively small subset of the data. Next, we include only hospitals that were observed in both the control period (year 1989) and the treatment periods (years 1990–1992). Finally, we drop the states of Colorado and Washington from the sample because they do not have enough hospitals in the data (and we cannot

[9]We choose quartiles to balance the tradeoff between bias and variance. If, for example, we choose the worst mortality *decile* instead then we would end up with too little data, which would adversely affect the variance of the estimators. If, for example, we choose the *median* (or the mean) as a cutoff then the bias for the subset of selective hospitals may be too large.

reasonably assign which hospitals have high mortality rates and which ones have low mortality rates). Overall, these adjustments decrease the patient sample size for the CABG estimation from 121,900 to 106,092. Table 8 contains the estimation results of the hospital heterogeneity estimations. Panel A contains the estimates for the entire adjusted sample. It shows that the changes to the sample did not affect the results qualitatively and only slightly quantitatively. For example, the average mortality rate $AMR_{00}$ decreased from 5.83% by 0.16 percentage points. The corresponding numbers from the original sample, contained in Panel A of Table 7, are 5.84% with a decrease of 0.05 percentage points. We conjecture that the adjustments to the estimation and data did not introduce meaningful bias to the estimates.

[Table 8 about here.]

Panels B and C illustrate that there are differences between hospitals. However, they are not statistically significant. Panel B shows the estimation results for the quartile of hospitals with the highest mortality rates ('worst 25% hospitals'). The average partial effect decreases by 0.85 (from a high rate of 9.59%) or 0.76 (from a rate of 8.51%) percentage points and is not statistically significant. Likewise, Panel C contains the results for the quartile of hospitals with the lowest mortality rates ('best 25% hospitals'). The average partial effects are estimated to be -0.52 and -0.78 percentage points, also statistically indistinguishable from zero. For both the best and worst 25% of hospitals, the average partial effects have increased (in absolute value). However, the difference is small, negative, and statistically insignificant. The middle two quartiles of hospitals, by implication (as the residual group), have lower partial effects (in absolute value): -0.05 and -0.04 percentage points, neither of which is statistically significant (numbers not included in table). In summary, we cannot corroborate econometrically that hospital heterogeneity has an effect on selection behavior.

*Patient Heterogeneity.* Providers have an incentive to select patients, but their opportunity to do so may be limited. A provider's ability to select a patient could depend on the admission type of the patient. Patients who are admitted to a hospital as emergency or urgent cases may be less likely targets for selection than elective patients. Table 9 demonstrates that in Pennsylvania during the year 1989, 32.5% of bypass patients were admitted as emergency cases, 26.6% as urgent cases, and 40.2% as elective cases (this distribution remains reasonably stable over the years). Among all bypass patients, 93.1% were admitted with a principal diagnosis of coronary atherosclerosis (CA, commonly referred to as clogged arteries), while 5.3% were admitted with AMI. Together, these diagnoses, CA and AMI, cover more than 98% of all CABG cases. Among the elective cases, 98.5% were admitted with a diagnosis of CA, while only 1% were admitted with a diagnosis of AMI. A similar picture emerges for the substitute procedure PTCA. There exists considerable patient heterogeneity across admission types and most patients (up to 99.4%) are admitted with either CA or AMI.

How can providers bias report cards in their favor by selecting elective patients? As in Dranove et al. (2003), we assume that there exists a subset of patients with positive measure that are admitted to hospital with heart problems (primary diagnoses of CA or AMI) and who qualify for both CABG surgery and PTCA. This gives providers some flexibility in assigning the appropriate clinical procedure for incoming elective patients. The introduction of the report cards

influences this assignment rule: an elective patient who poses a bad risk is now weakly more likely to be assigned to PTCA (or, put differently, a bad–risk patient is not more likely to be assigned to CABG due to the report cards). Likewise, an elective patient who poses a good risk is weakly more likely to be assigned to CABG (or, not more likely to be assigned to PTCA due to the report cards). We proceed to test whether providers selected patients from the pool of elective patients.

[Table 9 about here.]

Preliminary evidence that providers in Pennsylvania did not systematically choose elective patients to bias report cards has already been given by Table 5. The incidence of elective surgery for bypass patients changed insignificantly due to the report cards. This means that there was neither a net inflow from the pool of elective patients with a diagnosis of CA or AMI into CABG surgery, nor a net outflow. If, for example, providers reacted to the report cards by moving relatively more good–risk elective patients into CABG surgery, then we would expect to see an increase in the incidence of elective surgery for bypass patients. We do not find such an effect.

However, it is possible that selection into and out of the pool of elective patients did occur, but that the effects cancelled each other out. While the incidence of elective surgery overall would not change, the composition of elective patients receiving CABG surgery would shift to more good–risk patients and fewer bad–risk patients. We can test for such a cancelation effect. If providers select from the pool of elective patients, then we expect the mortality rate for patients who underwent CABG surgery and who were admitted as elective cases to decline in response to the report card. Just as before, if the econometric analysis detects a marked decline in mortality, we would interpret this as evidence in support of the hypothesis of selection.

[Table 10 about here.]

Table 10 reports average partial effects of 0.26 and 0.22 percentage points—both insignificant. If a redistribution of patient types did occur among elective patients, we would expect negative average partial effects. However, our estimates, while positive, are indistinguishable from zero. We conclude that providers did not select strategically and systematically from the pool of patients admitted as elective cases.

## 5. DISCUSSION

*Selection.* While our finding that providers do not select patients is in contrast to previous empirical studies, it is consistent with the theoretical results in Chen (2011), who points out that the CABG report cards in Pennsylvania not only measure providers' mortality rates, but also their patient volumes (the number of CABG surgeries performed at the hospital). Chen shows that hospitals need to trade off these two measures. Suppose that there are two hospitals of different quality (high and low) with the same patient volume. The high–quality hospital will have a lower mortality rate in CABG surgery than the low–quality hospital. If the low–quality hospital wants to imitate the mortality rate of the high–quality hospital, it can do so by avoiding sicker patients. This will drive down the patient volume measure, which will be revealed in the report card. In other words, it is not possible for the low–quality hospital to imitate the high–quality

hospital on both measures at the same time. Therefore, the low–quality hospital will not engage in patient selection. The high–quality hospital, knowing this, has no incentive to select patients either.

Secondly, the actual implementation of the CABG report cards by the PHC4 makes it difficult for providers to exploit private information. We have argued that patient selection can only be effective if it is based on a hospital's private information. Yet the value of that private information may be small given the rich clinical data that hospitals are required to provide to the report card authority. Hospitals have to submit an admission severity group (ASG) categorical score, which is constructed based on a complex combination of clinical parameters (including laboratory test results, levels of glucose, or blood urea nitrogen) and comorbidities that are collected by the hospital (see PHC4 1998). The ASG ranges from 0 (no risk of clinical instability) to 4 (maximal risk of clinical instability) and also reflects, to a certain extent, a patient's clinical history (including histories of congestive heart failure, seizures, and surgeries). Given the rich information that comprises the ASG score, it seems conceivable that it effectively serves as a sufficient statistic for private information. Consequently, the scope for the selection of patients by hospitals could be small.

Another reason for the lack of patient selection by hospitals could be altruism. Theoretical work by Newhouse (1970) describes hospitals as complex institutions that include quality of care in their objective functions. This introduces a bias against producing lower–quality products. Further, philanthropic funding arrangements enable non–profit hospitals to deviate from minimum average cost considerations and affords the hospital some latitude for inefficiency. Frank and Salkever (1991) develop two models of altruism: pure and impure. In pure altruism, a hospital only cares about filling the demand for charity care. If other hospitals increase their supply of charity care, then a given hospital will accordingly reduce its supply. The crowding–out effect is therefore perfect. However, such an extreme crowding–out effect is not supported empirically. As such, Frank and Salkever also propose a model of impure altruism in which hospitals perceive it as beneficial to offer charity care themselves, competing for public goodwill by providing and filling the gap for charity care. In contrast to pure (non–competitive) altruism, the crowding–out effect of charity care is smaller and better supported empirically. Rosko (2004) uses Pennsylvanian data from 1995 to 1998 to find empirical evidence in support of impure altruism: a given hospital's supply of uncompensated care increases with the provision of uncompensated care by other hospitals. Rosko also finds that non–profit hospitals tap into financial surpluses to offer uncompensated care to patients. Lastly, Gaskin (1997) gives empirical evidence from New Jersey during the years 1987 to 1992 that supports hospital altruism. New Jersey introduced uncompensated care pools out of which hospitals receive reimbursements for uncompensated care. In response, hospitals increased inpatient uncompensated care by almost 15% on average. Gaskin argues that the introduction of this pool lowered the shadow price for altruistic care.

Lastly, the absence of selection may be explained by a lack of knowledge on part of the hospitals. Hospitals may not know that performance data were being collected and/or they may not know that results would be published. However, either of these possibilities are unlikely. The PHC4 consists of members from the medical community in Pennsylvania, it consults publicly

and disseminated research ideas in collaboration with the hospitals to encourage feedback and enable wider acceptance of its work. Further, as explicitly required by Act 89 of 1986, the PHC4 is essentially set up as a public entity. Its meetings are open to the public, its reports are published in the Pennsylvania Bulletin, and all its actions have to be taken in open public sessions (see Sessa and Bentley 1992 and Sirio et al. 1996).

*Quality Improvements.* Given that the PHC4 uses actual in–hospital mortality as the main outcome measure, rather than improving follow–up care (after patient discharge), hospitals should focus on improving surgical procedures and immediate pre– and post–surgery care. If, instead, hospitals were to invest in improving follow–up procedures after patient discharge, then any reductions in mortality rates would not be reflected in the report cards. Examples of potential quality improvement in bypass surgery include: the hiring of full–time cardiac surgery chiefs to review and redesign the cardiac care systems; retraining of nurses; concentrating pre–surgery, surgery, and post–surgery procedures on a single hospital floor; installing dedicated cardiac anesthesia services (see Chassin 2002).

However, considering the short time frame used in our analysis, it seems possible that any steps undertaken by hospitals to improve quality have not manifested in lower mortality rates. This argument is consistent with our insignificant estimation results. In a recent theoretical paper, Gravelle and Sivey (2010) describe the possibility that even in an environment of competing for–profit hospitals, quality of care may not increase at all in response to public disclosure of the hospitals' qualities. In their model, patients prefer the higher–quality hospital regardless of the magnitude of the quality difference. Given this, when the initial quality difference between two hospitals is sufficiently high, hospitals lack incentives to improve quality. The reason is that any effort the lower–quality hospital exerts to improve its quality of care can only reduce the quality gap, but can never fully close it.

*Period 2–Effect.* Our finding of no provider selection in period 1 does not necessarily mean that there was no provider selection in period 2. Even if hospitals were aware of the timing of the report card scheme, as report cards were still in their infancy, hospitals were facing uncertainty with regard to the impact of the publication of the report cards on consumers. Following publication of report cards, as public awareness grows, so may providers incentives to select patients.

The effect of report card publication can be estimated, in principal, by comparing period 2 to period 0. To address this period 2–effect, we combine the years 1993 and 1994 as period 2 and apply the same estimation strategy as before. We find that the average mortality rate decreases from 5.77% to 4.77%. This means that the mortality rates for a patient from period 0 would decrease by 1.00 percentage point due to the publication of the report cards.[10] These findings are consistent with Dranove et al. (2003). Recall that Dranove et al. do a difference–in–differences analysis in which they compare data from 1987–1992 (i.e., periods 0 and 1 jointly) to data from 1993–1994 (period 2) and they also find a significant effect. In essence, Dranove et al. estimate a pure period 2–effect: Given our main finding of a zero period 1–effect, the fact

[10]More detailed results are available from the authors on request.

that Dranove et al. combine periods 0 and 1 in a comparison with period 2 is innocuous and does yield results that are qualitatively identical to our comparison of period 0 and period 2.

This period 2–effect, however, has no clear interpretation yet as it confounds at least three different factors: hospitals selecting patients, patients selecting hospitals, and quality improvements by hospitals. Separating these factors for period 2 remains a future research direction.

## 6. CONCLUSION

The PHC4 predicted two outcomes of report cards for CABG surgery: the provision of patients "with data that will help them have more informed discussions with their physicians" and the prompting of "providers to take steps to improve the overall quality of bypass surgery". Economists have pointed out a third and unintended consequence: a shift in bypass surgery from sicker patients to healthier ones. These three report card effects—patients selecting hospitals, quality improvement, and hospitals selecting patients—are difficult to separate in empirical work. We exploit a short period of asymmetric information during the implementation of Pennsylvania's report card scheme to shut down the first of the three effects. We compare mortality outcomes of bypass patients from before the implementation of report cards to the period during which the PHC4 collected patient outcome data but prior to the publication of results. During this period, the report cards cannot have altered patients' choice of providers. We then proceed to test for improvement in the quality of patient care and for selection of patients by providers jointly. Through both channels, the introduction of report cards had the potential to decrease average illness severity of bypass patients and therefore mortality outcomes. However, in our empirical analysis, we find no such significant negative effect. We conclude that neither quality improvements nor patient selection occurred. In our empirical strategy, we explicitly account for clustering of the policy variable at the state/year level. We adapt Wooldridge's (2006) efficient non–linear minimum distance estimation and provide detailed asymptotic derivations in the Appendix.

We believe that our approach applies to many other report card schemes—some examples include: the Los Angeles' Department of Health Services restaurant hygiene grade cards from 1998; the Care Quality Commission's collection of information on the quality of National Health Service hospitals in the United Kingdom and the Australian Curriculum, Assessment and Reporting Authority's My School website, launched in 2010 and publishing nationally comparable data on almost 10,000 schools. Our asymptotic derivations can be adapted readily.

## REFERENCES

ASHENFELTER, O. (1978): "Estimating the Effect of Training Programs on Earnings," *The Review of Economics and Statistics*, 60, 47–57.

CHASSIN, M. (2002): "Achieving And Sustaining Improved Quality: Lessons From New York State And Cardiac Surgery," *Health Affairs*, 21, 40–51.

CHEN, Y. (2011): "Why Are Health Care Report Cards so Bad (Good)?" *Journal of Health Economics*, 30, 575–590.

DONALD, S. G. AND K. LANG (2007): "Inference with Difference–in–Differences and Other Panel Data," *The Review of Economics and Statistics*, 89, 221–233.

DRANOVE, D., D. KESSLER, M. MCCLELLAN, AND M. SATTERTHWAITE (2003): "Is more Information Better? The Effects of "Report Cards" on Health Care Providers," *The Journal of Political Economy*, 111, 555–588.

ELIXHAUSER, A., C. STEINER, D. HARRIS, AND R. COFFEY (1998): "Comorbidity Measures for Use with Administrative Data," *Med Care*, 36, 8–27.

EPSTEIN, A. J. (2006): "Do Cardiac Surgery Report Cards Reduce Mortality? Assessing the Evidence," *Medical Care Research and Review*, 63, 403–426.

FRANK, R. G. AND D. S. SALKEVER (1991): "The Supply of Charity Services by Nonprofit Hospitals: Motives and Market Structure," *The RAND Journal of Economics*, 22, pp. 430–445.

GASKIN, D. J. (1997): "Altruism or Moral Hazard: The Impact of Hospital Uncompensated Care Pools," *Journal of Health Economics*, 16, 397–416.

GOLDBERGER, A. S. (1968): *Topics in Regression Analysis*, Macmillan.

GRAVELLE, H. AND P. SIVEY (2010): "Imperfect Information in a Quality–Competitive Hospital Market," *Journal of Health Economics*, 24, 524–535.

HANSEN, C. B. (2007a): "Asymptotic Properties of a Robust Variance Matrix Estimator for Panel Data when $T$ is Large," *Journal of Econometrics*, 141, 597–620.

——— (2007b): "Generalized Least Squares Inference in Panel and Multilevel Models with Serial Correlation and Fixed Effects," *Journal of Econometrics*, 140, 670–694.

MOULTON, B. R. (1990): "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units," *The Review of Economics and Statistics*, 72, 334–338.

NEWHOUSE, J. P. (1970): "Toward a Theory of Nonprofit Institutions: An Economic Model of a Hospital," *The American Economic Review*, 60, 64–74.

PAPKE, L. E. AND J. M. WOOLDRIDGE (2008): "Panel Data Methods for Fractional Response Variables with an Application to Test Pass Rates," *Journal of Econometrics*, 145, 121–133.

PHC4 (1998): "Coronary Artery Bypass Graft Surgery—1994-95 Data." *Pennsylvania Health Care Cost Containment Council. Research Methods and Results*.

ROSKO, M. D. (2004): "The Supply of Uncompensated Care in Pennsylvania Hospitals: Motives and Financial Consequences," *Health Care Management Review*, 29, 229–239.

SESSA, E. J. AND J. M. BENTLEY (1992): "Promoting Competition in the Health Care Market: Pennsylvania's Cost Containment Strategy," *Compensation and Benefits Management*, 8, 7–16.

SIRIO, C. A., E. J. SESSA, AND J. L. MCGEE (1996): "Pennsylvania Health Care Cost Cost Containment Council—The Role of a State Health Care Agency in an Evolving Health Care Market," *American Journal of Medical Quality*, 11, S82–S85.

WOOLDRIDGE, J. M. (2006): "Cluster–Sample Methods in Applied Econometrics: An Extended Analysis," *Mimeo*.

## APPENDIX A. DERIVATION OF STANDARD ERRORS ON AVERAGE MORTALITY RATES AND AVERAGE PARTIAL EFFECTS

A.1. **First–Stage estimation.** First, we estimate the parameters in equation (3.3). For each state separately we pool data across all available years and run a logit of $m_{ist}$ on a constant, a set of hospital characteristics, a set of time dummies, and patient characteristics. (We do not include the derivations for the linear probability model here. Given the results for the logit model, the linear probability model follows straightforwardly.) For state $s$, denote by $N_{hst}$ the total number of patients in hospital $h$ observed during year $t$ and by $H_{st}$ the total number of hospitals observed during year $t$. Then the total number of patients observed during year $t$ equals $N_{st} := N_{hst}H_{st}$. Define by $H_s$ the total number of hospitals observed in state $s$ over all years. Let $T_s$ be the number of years observed in the data for state $s$. Then the total number of observations available for state $s$ equals $N_s := N_{st}T_s$.

Suppose each vector $z_{ihst}$ has dimension $R_s$, where $R_s$ is the number of patient characteristics included for state $s$ (we define vectors as column vectors). We stack vectors in the following order. At the hospital level, define a stacked matrix of patient characteristics $z_{hst} := (z_{1hst},\ldots,z_{N_{hst}hst})'$. At the year level, define a stacked matrix of patient characteristics $z_{st} := (z'_{1st},\ldots,z'_{H_{st}st})'$. At the state level, define a stacked matrix of patient characteristics $z_s := (z'_{s1},\ldots,z'_{sT_s})'$. Note that $\dim(z_s) = N_s \times R_s$.

For each state we now define a bijective mapping from the index set $K_s := \{1,2,\ldots,N_s\}$ to the set of ordered triples $O_{N_s} := \{(i,h,t)|i \in \{1,\ldots,N_{hst}\}, h \in \{1,\ldots,H_{st}\},\ t \in \{1,\ldots,T_s\}\}$, $\kappa_s : K_s \mapsto O_{N_s}$. We can therefore write each element of $z_s$ as $z_{ks}$ with $k \in K_s$. (In what follows, the distinction between $z_{st}$ and $z_{ks}$ whenever $(s,t)$ or $(k,s)$ take on specific numbers will be clear from the context.)

Define a $T_s$-dimensional vector of year dummies, $\mathsf{e}^t_{ks}$, and rewrite equation (3.3)

$$m_{ks} = 1 \cdot \{e^{t\,\prime}_{ks}\theta_s + c'_{ks}\phi_s + z_{ks}{}'\gamma_s + u_{ks} \geq 0\},$$

where $e^t_{ks} := (1, \mathsf{e}^t_{ks})'$ and the definitions of $m_{ks}$, $c_{ks}$, and $u_{ks}$ are analogous to the definition of $z_{ks}$. The associated score function is

$$\mathfrak{s}_{ks}(\theta_s,\phi_s,\gamma_s) := \frac{m_{ks} - F(e^{t\,\prime}_{ks}\theta_s + c'_{ks}\phi_s + z_{ks}{}'\gamma_s)}{F(e^{t\,\prime}_{ks}\theta_s + c'_{ks}\phi_s + z_{ks}{}'\gamma_s)\left(1 - F(e^{t\,\prime}_{ks}\theta_s + c'_{ks}\phi_s + z_{ks}{}'\gamma_s)\right)} \cdot f(e^{t\,\prime}_{ks}\theta_s + c'_{ks}\phi_s + z_{ks}{}'\gamma_s) \begin{bmatrix} e^t_{ks} \\ c_{ks} \\ z_{ks} \end{bmatrix},$$

where $F(y) := \exp(y)/(1+\exp(y))$ and $f(y) := \exp(y)/(1+\exp(y))^2$ are from the logit model. It is well known that the maximum likelihood estimator is defined as the vector $(\hat{\theta}'_s, \hat{\phi}'_s, \hat{\gamma}'_s)'$

that solves, for a given $s$, $\sum_{k=1}^{N_s} \mathfrak{s}_{ks}(\hat{\theta}_s, \hat{\phi}_s, \hat{\gamma}_s) = 0$ and that the following result holds for its asymptotic distribution.

$$\sqrt{N_s}\begin{pmatrix}\hat{\theta}_s - \theta_s \\ \hat{\phi}_s - \phi_s \\ \hat{\gamma}_s - \gamma_s\end{pmatrix} \xrightarrow{d} \mathrm{N}\left(0, \begin{bmatrix}\Upsilon_s^{\theta\theta} & \Upsilon_s^{\theta\phi} & \Upsilon_s^{\theta\gamma} \\ \Upsilon_s^{\phi\theta} & \Upsilon_s^{\phi\phi} & \Upsilon_s^{\phi\gamma} \\ \Upsilon_s^{\gamma\theta} & \Upsilon_s^{\gamma\phi} & \Upsilon_s^{\gamma\gamma}\end{bmatrix}\right) =: \mathrm{N}(0, \Upsilon_s),$$

with

$$\Upsilon_s := \left(E\left[\mathfrak{s}_{ks}(\theta_s, \phi_s, \gamma_s)\mathfrak{s}_{ks}(\theta_s, \phi_s, \gamma_s)'\right]\right)^{-1},$$

due to the information matrix equality. By the analogy principle (see Goldberger (1968)), a valid estimator for the asymptotic variance is

$$\hat{\Upsilon}_s := \left(N_s^{-1}\sum_{k=1}^{N_s}\left[\mathfrak{s}_{ks}(\hat{\theta}_s, \hat{\phi}_s, \hat{\gamma}_s)\mathfrak{s}_{ks}(\hat{\theta}_s, \hat{\phi}_s, \hat{\gamma}_s)'\right]\right)^{-1}.$$

Instead of large $N_s$-asymptotics we will consider, in everything that follows, large $N$-asymptotics. Our interpretation of $N_s$ then is that $N_s = \rho_s N$ with $0 < \rho_s \leq 1$. This results in

$$\sqrt{N}\begin{pmatrix}\hat{\theta}_s - \theta_s \\ \hat{\phi}_s - \phi_s \\ \hat{\gamma}_s - \gamma_s\end{pmatrix} \xrightarrow{d} \mathrm{N}(0, \rho_s^{-1}\Upsilon_s).$$

Let $t = 1, 2, \ldots, T_s$. With the estimator $\hat{\theta}_s = (\hat{\theta}_{1s}, \hat{\theta}_{2s}, \ldots, \hat{\theta}_{T_s s})'$ in hand we can construct an estimator for $\delta_{st}$ for state $s$. We have

$$\hat{\delta}_s := \begin{bmatrix}\hat{\delta}_{s1} \\ \hat{\delta}_{s2} \\ \vdots \\ \hat{\delta}_{sT_s}\end{bmatrix} := \begin{bmatrix}\hat{\theta}_{1s} \\ \hat{\theta}_{1s} + \hat{\theta}_{2s} \\ \vdots \\ \hat{\theta}_{1s} + \hat{\theta}_{T_s s}\end{bmatrix}.$$

To derive the asymptotic distribution of $\delta_{st}$ define by $L_s$ the $T_s$-dimensional identity matrix with the first column replaced by a $T_s$-dimensional column vector of ones. Then, $\delta_s = L_s\theta_s$ and it follows that

$$\sqrt{N_s}(\hat{\delta}_s - \delta_s) \xrightarrow{d} \mathrm{N}(0, \Omega_s),$$

with $\Omega_s := L_s\Upsilon_s^{\theta\theta}L_s'$. Similar to before, we apply apply large $N$-asymptotics, which yields

$$\sqrt{N}(\hat{\delta}_s - \delta_s) \xrightarrow{d} \mathrm{N}\left(0, \rho_s^{-1}\Omega_s\right).$$

Stack all $S$ vectors $\delta_s$ in a new vector $\delta := (\delta_1', \delta_2', \ldots, \delta_S')'$. The dimension of $\delta$ is $D := \left(\sum_{s=1}^{S} T_s\right)$. Then

$$\sqrt{N}(\hat{\delta} - \delta) \xrightarrow{d} \mathrm{N}(0, \Omega),$$

where

$$\Omega := \begin{bmatrix}\rho_1^{-1}\Omega_1 & 0 & \cdots & 0 \\ 0 & \rho_2^{-1}\Omega_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \ldots & \rho_S^{-1}\Omega_S\end{bmatrix}.$$

A.2. **Second–Stage Estimation.** In the second stage we obtain $\hat{\beta}$ based on equation (3.2) by efficient minimum distance estimation. In order to set up the minimum distance estimator properly, denote by $e^s$ a $(D \times (S-1))$-dimensional matrix of state dummies, by $e^t$ a $(D \times (T-1))$-dimensional matrix of time dummies ($T := \max_s\{T_s\}$), and by $e^p$ a $(D \times 1)$-dimensional policy dummy that takes on value one if $s$ and $t$ correspond to the treatment state in the treatment period and zero otherwise. Next, define a $(D \times (S+T))$-dimensional matrix $x$ that collects all these dummy variables and also a constant term, i.e., $x := (1, e^s, e^t, e^p)$. The efficient minimum distance estimator is defined as

$$\hat{\psi} := \operatorname{argmin}_{\tilde{\psi}} \left(\hat{\delta} - x\tilde{\psi}\right)' \Omega^{-1} \left(\hat{\delta} - x\tilde{\psi}\right),$$

where $\psi := (\mu, a_2, \ldots, a_S, b_2, \ldots, b_T, \beta)'$ from equation (3.2). The asymptotic distribution of the efficient minimum distance estimator is

$$\sqrt{N}(\hat{\psi} - \psi) \xrightarrow{d} \mathrm{N}\left(0, (x'\Omega^{-1}x)^{-1}\right).$$

To obtain the asymptotic distribution of $\hat{\beta}$, let $l$ be the $(S+T)$-dimensional vector $(0, 0, \ldots, 0, 1)'$. Then $\beta = l'\psi$ and

$$\sqrt{N}(\hat{\beta} - \beta) \xrightarrow{d} \mathrm{N}\left(0, l'(x'\Omega^{-1}x)^{-1}l\right).$$

An estimator of the asymptotic variance is given by $l'(x'\hat{\Omega}^{-1}x)^{-1}l$, where,

$$\hat{\Omega} := \begin{bmatrix} \rho_1^{-1}\hat{\Omega}_1 & 0 & \cdots & 0 \\ 0 & \rho_2^{-1}\hat{\Omega}_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \rho_S^{-1}\hat{\Omega}_S \end{bmatrix},$$

and

$$\hat{\Omega}_s := L_s \hat{\Upsilon}_s^{\theta\theta} L_s'.$$

A.3. **Average Partial Effects.** To derive standard errors on $\mathrm{APE}_A$ and $\mathrm{APE}_B$ we first propose standard errors on the AMRs. In doing so, we keep notation as generic as possible to facilitate adaptation of the results. In general, an AMR from the set of equations (3.4) has the form

$$\begin{aligned} \mathrm{AMR}_{pq} :=& E\left[F\left(\mu + a_s + b_t + q\beta + c_{hst}'\phi_s + z_{ihst}'\gamma_s\right) | s \in S, t \in T_s^p\right] \\ =& E\left[F\left(l_{stq}'\psi + c_{hst}'\phi_s + z_{ihst}'\gamma_s\right) | s \in S, t \in T_s^p\right], \end{aligned} \tag{A.1}$$

where $p, q \in \{0,1\}$, $T_s^p \subseteq \{1, \ldots, T_s\}$ and $l_{stq}$ is the $(S+T)$-dimensional vector such that $l_{stq}'\psi = \mu + a_s + b_t + q\beta$.

Similar to before, for state $s$ we now define a bijective mapping from the index set $J_s^p := \{1, 2, \ldots, N_s^p\}$ to the set of ordered triples $O_{N_s^p} := \{(i,h,t) | i \in \{1, \ldots, N_{hst}\},\ h \in \{1, \ldots, H_{st}\}, t \in T_s^p\}$, $\iota_s : J_s^p \mapsto O_{N_s^p}$, where $N_s^p$ denotes the total number of patient observations in state $s$ during years $T_s^p$, i.e., $N_s^p := \sum_{\forall t \in T_s^p} N_{st}$. We can therefore index all variables by $j$ instead, for example, $z_{js}$ with $j \in J_s^p$.

The derivations here are similar to Papke and Wooldridge (2008). In order to derive an estimator and standard errors, we rewrite equation (A.1) as

$$\mathrm{AMR}_{pq} := E\left[F\left(l_{stq}'\psi + c_{js}'\phi_s + z_{js}'\gamma_s\right) | s \in S, t \in T_s^p\right],$$

which, by the analogy principle, we estimate by

$$\widehat{\mathrm{AMR}}_{pq} := (N_s^p)^{-1} \sum_{j=1}^{N_s^p} F\left(l'_{stq}\hat{\psi} + c'_{js}\hat{\phi}_s + z'_{js}\hat{\gamma}_s\right).$$

For notational conciseness we define $w_{js} := (c'_{js}, z'_{js})'$ and $\lambda_s := (\phi'_s, \gamma'_s)'$. Next, consider the first order Taylor series expansion

$$(N_s^p)^{-1/2} \sum_{j=1}^{N_s^p} F\left(l'_{stq}\hat{\psi} + w'_j\hat{\lambda}_s\right) = (N_s^p)^{-1/2} \sum_{j=1}^{N_s^p} F\left(l'_{stq}\psi + w'_j\lambda_s\right) + \left((N_s^p)^{-1} \sum_{j=1}^{N_s^p} \nabla_{(\psi,\lambda_s)} F\left(l'_{stq}\tilde{\psi} + w'_j\tilde{\lambda}_s\right)\right) \sqrt{N_s^p} \begin{pmatrix} \hat{\psi} - \psi \\ \hat{\lambda}_s - \lambda_s \end{pmatrix},$$

where $\tilde{\psi}$ lies between $\psi$ and $\hat{\psi}$ and $\tilde{\lambda}_s$ lies between $\lambda_s$ and $\hat{\lambda}_s$. Noting that the asymptotic distributions of $\hat{\psi}$ and $\hat{\lambda}_s$ are normal and letting

$$(N_s^p)^{-1} \sum_{j=1}^{N_s^p} \nabla_{(\psi,\lambda_s)} F\left(l'_{stq}\tilde{\psi} + w'_j\tilde{\lambda}_s\right) \xrightarrow{p} E\left[\nabla_{(\psi,\lambda_s)} F\left(l'_{stq}\tilde{\psi} + w'_j\tilde{\lambda}_s\right)\right] =: B,$$

we obtain

$$\left((N_s^p)^{-1} \sum_{j=1}^{N_s^p} \nabla_{(\psi,\lambda_s)} F\left(l'_{stq}\tilde{\psi} + w'_j\tilde{\lambda}_s\right)\right) \sqrt{N_s^p} \begin{pmatrix} \hat{\psi} - \psi \\ \hat{\lambda}_s - \lambda_s \end{pmatrix} = B\sqrt{N_s^p} \begin{pmatrix} \hat{\psi} - \psi \\ \hat{\lambda}_s - \lambda_s \end{pmatrix} + o_p(1).$$

We therefore have

$$(N_s^p)^{-1/2} \sum_{j=1}^{N_s^p} F\left(l'_{stq}\hat{\psi} + w'_j\hat{\lambda}_s\right) = (N_s^p)^{-1/2} \sum_{j=1}^{N_s^p} F\left(l'_{stq}\psi + w'_j\lambda_s\right) + B\sqrt{N_s^p} \begin{pmatrix} \hat{\psi} - \psi \\ \hat{\lambda}_s - \lambda_s \end{pmatrix} + o_p(1). \tag{A.2}$$

To proceed, we need to derive the asymptotic variance of

$$\sqrt{N_s^p} \begin{pmatrix} \hat{\psi} - \psi \\ \hat{\lambda}_s - \lambda_s \end{pmatrix} = \sqrt{\rho_{sp}}\sqrt{N} \begin{pmatrix} \hat{\psi} - \psi \\ \hat{\lambda}_s - \lambda_s \end{pmatrix},$$

where $\rho_{sp} := N_s^p/N$. Note that, as a corollary from minimum distance estimation, $\sqrt{N}(\hat{\psi} - \psi) = (x\Omega^{-1}x')^{-1}x\Omega^{-1}\sqrt{N}(\hat{\delta} - \delta)$. Recall that $\delta := (\delta'_1, \ldots, \delta'_S)'$ and that $\delta_s := L_s\theta_s$. We can thus construct an expression $L\theta = \delta$ with $L$ being the block diagonal matrix whose diagonal elements are the individual $L_s$. Thence, $\sqrt{N}(\hat{\psi} - \psi) = M\sqrt{N}(\hat{\theta} - \theta)$ and

$$\sqrt{\rho_{sp}}\sqrt{N} \begin{pmatrix} \hat{\psi} - \psi \\ \hat{\lambda}_s - \lambda_s \end{pmatrix} = \sqrt{\rho_{sp}}\sqrt{N} \begin{bmatrix} M & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix} \begin{pmatrix} \hat{\theta} - \theta \\ \hat{\phi}_s - \phi_s \\ \hat{\gamma}_s - \gamma_s \end{pmatrix},$$

where $M := (x\Omega^{-1}x')^{-1}x\Omega^{-1}L$. This is helpful because

$$\sqrt{N} \begin{pmatrix} \hat{\theta} - \theta \\ \hat{\phi}_s - \phi_s \\ \hat{\gamma}_s - \gamma_s \end{pmatrix} \xrightarrow{d} \mathrm{N}(0, \Upsilon),$$

with

$$\Upsilon := \begin{bmatrix} \Upsilon^{\theta\theta}_{<s} & 0 & 0 & 0 & 0 \\ 0 & \Upsilon^{\theta\theta}_{s} & 0 & \Upsilon^{\theta\phi}_{s} & \Upsilon^{\theta\gamma}_{s} \\ 0 & 0 & \Upsilon^{\theta\theta}_{>s} & 0 & 0 \\ 0 & \Upsilon^{\phi\theta}_{s} & 0 & \Upsilon^{\phi\phi}_{s} & \Upsilon^{\phi\gamma}_{s} \\ 0 & \Upsilon^{\gamma\theta}_{s} & 0 & \Upsilon^{\gamma\phi}_{s} & \Upsilon^{\gamma\gamma}_{s} \end{bmatrix}$$

$$\Upsilon^{\theta\theta}_{<s} := \begin{bmatrix} \Upsilon^{\theta\theta}_{1} & 0 & \cdots & 0 \\ 0 & \Upsilon^{\theta\theta}_{2} & \cdots & \\ \vdots & & \ddots & \\ 0 & 0 & \cdots & \Upsilon^{\theta\theta}_{s-1} \end{bmatrix}, \qquad \text{and} \qquad \Upsilon^{\theta\theta}_{>s} := \begin{bmatrix} \Upsilon^{\theta\theta}_{s+1} & 0 & \cdots & 0 \\ 0 & \Upsilon^{\theta\theta}_{s+2} & \cdots & \\ \vdots & & \ddots & \\ 0 & 0 & \cdots & \Upsilon^{\theta\theta}_{S} \end{bmatrix}.$$

Defining $L$ now as $L := \begin{bmatrix} M & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix}$ we get

$$\text{Avar}\left(\sqrt{\rho_{sp}}\sqrt{N}\begin{pmatrix}\hat{\psi}-\psi\\ \hat{\lambda}_s-\lambda_s\end{pmatrix}\right) = \text{Avar}\left(\sqrt{\rho_{sp}}\sqrt{N}L\begin{pmatrix}\hat{\theta}-\theta\\ \hat{\phi}_s-\phi_s\\ \hat{\gamma}_s-\gamma_s\end{pmatrix}\right) = \rho_{sp}^{-1}L\Upsilon L' =: \Xi.$$

Going back to equation (A.2) we obtain

$$\begin{aligned}\sqrt{N_s^p}\left(\widehat{AMR}_{pq} - AMR_{pq}\right) &= \sqrt{N_s^p}\left(\sum_{j=1}^{N_s^p} F\left(l_{stq}'\hat{\psi}+w_j'\hat{\lambda}_s\right) - E\left[F\left(l_{stq}'\psi+w_j'\lambda_s\right)\right]\right)\\ &= (N_s^p)^{-1/2}\left(\sum_{j=1}^{N_s^p} F\left(l_{stq}'\psi+w_j'\lambda_s\right) - E\left[F\left(l_{stq}'\psi+w_j'\lambda_s\right)\right]\right)\\ &\quad + \sqrt{N_s^p}B\begin{pmatrix}\hat{\psi}-\psi\\ \hat{\lambda}_s-\lambda_s\end{pmatrix} + o_p(1).\end{aligned}$$

Both terms on the right hand side have an asymptotic normal distribution and are uncorrelated. We have

$$(N_s^p)^{-1/2}\left(\sum_{j=1}^{N_s^p} F\left(l_{stq}'\psi+w_j'\lambda_s\right) - E\left[F\left(l_{stq}'\psi+w_j'\lambda_s\right)\right]\right) \xrightarrow{d} \text{N}(0,\Gamma),$$

with $\Gamma := \text{Var}\left[F\left(l_{stq}'\psi+w_j'\lambda_s\right)\right]$ and

$$\sqrt{N_s^p}B\begin{pmatrix}\hat{\psi}-\psi\\ \hat{\lambda}_s-\lambda_s\end{pmatrix} \xrightarrow{d} \text{N}(0, B\Xi B').$$

Hence

$$\sqrt{N_s^p}\left(\widehat{AMR}_{pq} - AMR_{pq}\right) \xrightarrow{d} \text{N}(0,\Gamma + B\Xi B').$$

Estimators for the terms in the asymptotic variance follow immediately:

$$\begin{aligned}\hat{\Gamma} &:= (N_s^p)^{-1}\sum_{j=1}^{N_s^p} F\left(l_{stq}'\psi+w_j'\lambda_s\right)^2 - \left((N_s^p)^{-1}\sum_{j=1}^{N_s^p} F\left(l_{stq}'\psi+w_j'\lambda_s\right)\right)^2\\ \hat{B} &:= (N_s^p)^{-1}\sum_{j=1}^{N_s^p} \nabla_{(\psi,\lambda_s)} F\left(l_{stq}'\hat{\psi}+w_j'\hat{\lambda}_s\right)\\ \hat{\Xi} &:= \rho_{sp}^{-1}L\hat{\Upsilon}L'.\end{aligned}$$

Asymptotic distributions for the estimates of the average partial effects, $\text{APE}_A$ and $\text{APE}_B$, now follow straightforwardly because they are merely linear combinations between different average mortality rates $AMR_{pq}$.

RESEARCH SCHOOL OF ECONOMICS, AUSTRALIAN NATIONAL UNIVERSITY, CANBERRA, ACT 0200, AUSTRALIA

*E-mail address*, Yijuan Chen: yijuan.chen@anu.edu.au

*E-mail address*, Juergen Meinecke: juergen.meinecke@anu.edu.au

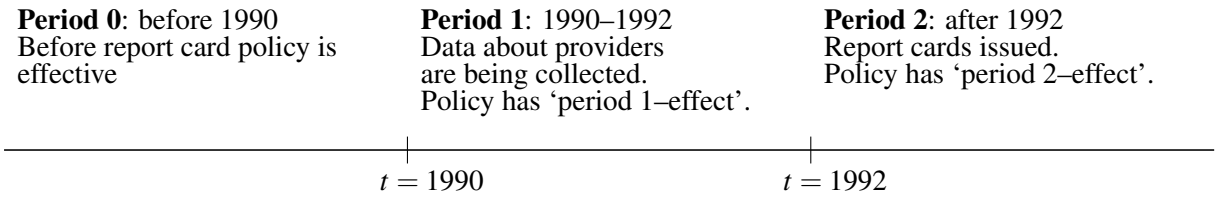


Figure 1: Timing of Report Card Policy in Pennsylvania

Table 1. —Nationwide Inpatient Sample (NIS) Data Coverage

| Calendar year | States covered | Number of hospitals | Number of in–patient stays |
|---|---|---|---|
| 1988 | CA, CO, FL, IL, IA, MA NJ, WA | 759 | 5,265,756 |
| 1989 | AZ, CA, CO, FL, IL, IA, MA NJ, PA, WA, WI | 882 | 6,110,064 |
| 1990 | AZ, CA, CO, FL, IL, IA, MA NJ, PA, WA, WI | 871 | 6,268,515 |
| 1991 | AZ, CA, CO, FL, IL, IA, MA NJ, PA, WA, WI | 859 | 6,156,188 |
| 1992 | AZ, CA, CO, FL, IL, IA, MA NJ, PA, WA, WI | 856 | 6,195,744 |

Note.—Source: NIS documentation, release 3.

Table 2. —Descriptive Statistics: Control States

| **Variable** | **1988** | **1989** | **1990** | **1991** | **1992** |
|---|---|---|---|---|---|
| Male | 51.7 | 51.8 | 52.2 | 52.7 | 53.0 |
| White | 86.0 | 83.8 | 82.0 | 83.2 | 84.8 |
| Age | 74.4 | 74.6 | 74.7 | 74.8 | 74.9 |
| Emergency | 49.2 | 51.1 | 52.0 | 51.7 | 52.0 |
| Urgent | 33.8 | 32.0 | 32.9 | 32.8 | 32.4 |
| Elective | 15.7 | 15.8 | 14.6 | 15.0 | 15.3 |
| Principal procedure | | | | | |
| CABG | 11.2 | 10.9 | 11.8 | 12.6 | 12.9 |
| PTCA | 7.5 | 8.5 | 10.0 | 11.4 | 12.5 |
| CATH | 16.7 | 17.1 | 17.4 | 17.7 | 18.0 |
| Died | | | | | |
| CABG | 5.8 | 5.4 | 5.1 | 4.9 | 4.5 |
| PTCA | 2.1 | 1.9 | 1.9 | 1.9 | 1.6 |
| CATH | 2.1 | 1.9 | 1.8 | 1.8 | 1.8 |
| N | 195,461 | 152,964 | 157,802 | 164,826 | 168,116 |

Note.—All numbers in percent (except sample size N). Sample: Patients, age 65–99 who have received principal diagnoses of acute myocardial infarction, coronary atherosclerosis, non–specific chest pain, pulmonary heart disease, other and ill-defined heart disease, conduction disorders, cardiac dysrhythmias, cardiac arrest and ventricular fibrillation, or congestive heart failure (non–hypertensive). Control states included are only those that were always in the NIS sample between 1988 and 1992 (i.e., CA, CO, FL, IL, IA, MA NJ, WA). Variable 'White' for control states based on CA, IA, MA, NJ only. Variables 'Emergency', 'Urgent', and 'Elect' for control states based on CA, FL, IL, MA, NJ, WA only. See main text for explanation of procedure acronyms.

Table 3. —Descriptive Statistics: Pennsylvania

| **Variable** | **1989** | **1990** | **1991** | **1992** |
|---|---|---|---|---|
| Male | 47.8 | 50.0 | 50.3 | 51.1 |
| Age | 74.2 | 74.1 | 74.3 | 74.4 |
| Emergency | 58.6 | 58.5 | 58.8 | 59.9 |
| Urgent | 23.0 | 20.2 | 21.4 | 22.4 |
| Elective | 15.3 | 18.8 | 19.3 | 17.6 |
| Principal procedure | | | | |
| CABG | 10.8 | 12.6 | 13.1 | 13.7 |
| PTCA | 7.1 | 9.6 | 10.3 | 11.2 |
| CATH | 15.4 | 16.6 | 17.1 | 18.1 |
| Died | | | | |
| CABG | 5.8 | 5.4 | 4.9 | 4.7 |
| PTCA | 1.7 | 1.0 | 1.5 | 1.7 |
| CATH | 1.8 | 1.5 | 1.4 | 1.5 |
| N | 23,256 | 28,839 | 29,953 | 30,983 |

Note.—All numbers in percent (except sample size N). Sample: Patients, age 65–99 who have received principal diagnoses of acute myocardial infarction, coronary atherosclerosis, non–specific chest pain, pulmonary heart disease, other and ill-defined heart disease, conduction disorders, cardiac dysrhythmias, cardiac arrest and ventricular fibrillation, or congestive heart failure (non–hypertensive). Variable 'White' not available for PA. See main text for explanation of procedure acronyms.

Table 4. —Description of Covariates included in Estimation

| Vector | Category | Variable | Description |
|---|---|---|---|
| $z_{ihst}$ | Demographic | Age dummies | Dummy equal 1 if $65 \leq$ Age $\leq 69$ |
| | | | Dummy equal 1 if $70 \leq$ Age $\leq 74$ |
| | | | Dummy equal 1 if $75 \leq$ Age $\leq 79$ |
| | | Admission type | Dummy equal 1 if emergency admission |
| | | | Dummy equal 1 if urgent admission |
| | | Male | Dummy equal 1 if male |
| | | White | Dummy equal 1 if white |
| | Clinical/comorbidities | AMI | Dummy equal 1 if diagnosed with acute myocardial infarction |
| | | Complications | Dummy equal 1 if complications during hospitalization (complications defined as in Elixhauser et al. 1998) |
| | | Peripheral vascular disorders | Dummy equal 1 if diagnosed with peripheral vascular disorders |
| | | Hypertension | Dummy equal 1 if diagnosed with hypertension |
| | | Chronic pulmonary disease | Dummy equal 1 if diagnosed with chronic pulmonary disease |
| | | Diabetes, uncomplicated | Dummy equal 1 if diagnosed with diabetes, uncomplicated |
| | | Fluid and electrolyte disorders | Dummy equal 1 if diagnosed with fluid and electrolyte disorders |
| $c_{hst}$ | Hospital | Urban status | Dummy equal 1 if urban |
| | | Teaching status | Dummy equal 1 if urban–teaching hospital |
| | | Size | Dummy equal 1 if medium size hospital |
| | | | Dummy equal 1 if large hospital |
| | | Ownership status | Dummy equal 1 if public hospital |
| | | | Dummy equal 1 if private–profit hospital |

Table 5. —Estimation Results: Patient Characteristics

| **Characteristic** | $\hat{\beta}$ | **Confidence interval** | | |
|---|---|---|---|---|
| Demographics | | | | |
| Age 65–69 | 0.0054 | [ -0.0458 | ; | 0.0565 ] |
| Age 70–74 | -0.0023 | [ -0.0520 | ; | 0.0474 ] |
| Age 75–79 | 0.0011 | [ -0.0412 | ; | 0.0434 ] |
| Male | 0.0017 | [ -0.0481 | ; | 0.0515 ] |
| Admission type | | | | |
| Emergency | 0.0053 | [ -0.0432 | ; | 0.0537 ] |
| Urgent | -0.0685 | [ -0.1153 | ; | -0.0217 ] |
| Elective | 0.0220 | [ -0.0295 | ; | 0.0735 ] |
| Clinical/comorbidities | | | | |
| AMI | 0.0160 | [ -0.0101 | ; | 0.0421 ] |
| Complications | -0.0707 | [ -0.1209 | ; | -0.0204 ] |
| Peripheral vascular disorders | -0.0069 | [ -0.0275 | ; | 0.0138 ] |
| Hypertension | 0.0063 | [ -0.0358 | ; | 0.0483 ] |
| Chronic pulmonary disease | -0.0032 | [ -0.0324 | ; | 0.0260 ] |
| Diabetes, uncomplicated | 0.0058 | [ -0.0275 | ; | 0.0390 ] |
| Fluid and electrolyte disorders | -0.0325 | [ -0.0531 | ; | -0.0119 ] |

Note.—Sample: CABG patients, age 65–99. Confidence interval at 95% level. N = 121,900.

Table 6. —Estimation Results: Second Stage

| **Procedure** | **Functional form** $F(\cdot)$ | $\hat{\beta}$ | **Confidence interval** | **N** |
|---|---|---|---|---|
| CABG | LPM | -0.0008 | $[-0.0111; 0.0095]$ | 121,900 |
| | Logit | -0.0104 | $[-0.2180; 0.1973]$ | |
| PTCA | LPM | -0.0030 | $[-0.0098; 0.0037]$ | 101,122 |
| | Logit | -0.1623 | $[-0.6188; 0.2941]$ | |
| CATH | LPM | -0.0019 | $[-0.0068; 0.0029]$ | 172,534 |
| | Logit | -0.1245 | $[-0.4376; 0.1885]$ | |

Note.—Sample: CABG/ PTCA/ CATH patients, age 65–99. LPM: linear probability model. Confidence interval at 95% level.

Table 7. —Estimation Results: Average Mortality Rates and Average Partial Effects

| PANEL A: — CABG | | |
|---|---|---|
| **Parameter** | **Point estimate** | **95% Confidence interval** |
| $AMR_{00}$ | 5.84 | [ 5.49 ; 6.20 ] |
| $AMR_{01}$ | 5.79 | [ 5.49 ; 6.08 ] |
| $APE_A$ | -0.05 | [ -0.52 ; 0.41 ] |
| $AMR_{10}$ | 5.05 | [ 4.71 ; 5.39 ] |
| $AMR_{11}$ | 5.00 | [ 4.85 ; 5.16 ] |
| $APE_B$ | -0.05 | [ -0.43 ; 0.33 ] |
| N = 121,900 | | |

| PANEL B: — PTCA | | |
|---|---|---|
| **Parameter** | **Point estimate** | **95% Confidence interval** |
| $AMR_{00}$ | 1.65 | [ 1.41 ; 1.88 ] |
| $AMR_{01}$ | 1.41 | [ 1.25 ; 1.58 ] |
| $APE_A$ | -0.24 | [ -0.52 ; 0.06 ] |
| $AMR_{10}$ | 1.69 | [ 1.44 ; 1.94 ] |
| $AMR_{11}$ | 1.45 | [ 1.35 ; 1.55 ] |
| $APE_B$ | -0.24 | [ -0.51 ; 0.03 ] |
| N = 101,122 | | |

| PANEL C: — CATH | | |
|---|---|---|
| **Parameter** | **Point estimate** | **95% Confidence interval** |
| $AMR_{00}$ | 1.82 | [ 1.63 ; 2.01 ] |
| $AMR_{01}$ | 1.62 | [ 1.47 ; 1.77 ] |
| $APE_A$ | -0.20 | [ -0.44 ; 0.04 ] |
| $AMR_{10}$ | 1.68 | [ 1.51 ; 1.85 ] |
| $AMR_{11}$ | 1.50 | [ 1.41 ; 1.58 ] |
| $APE_B$ | -0.18 | [ -0.37 ; 0.00 ] |
| N = 172,534 | | |

Note.—Sample: CABG/ PTCA/ CATH patients, age 65–99. AMR reported in percent; APE in percentage points.

Table 8. —Estimation Results: Average Mortality Rates and Average Partial Effects (CABG only)

| PANEL A: — ALL HOSPITALS | | | | | |
|---|---|---|---|---|---|
| **Parameter** | **Point estimate** | **95% Confidence interval** | | | |
| $AMR_{00}$ | 5.83 | [ | 5.47 | ; | 6.20 ] |
| $AMR_{01}$ | 5.67 | [ | 5.36 | ; | 5.99 ] |
| $APE_A$ | -0.16 | [ | -0.65 | ; | 0.33 ] |
| $AMR_{10}$ | 5.06 | [ | 4.71 | ; | 5.41 ] |
| $AMR_{11}$ | 4.92 | [ | 4.75 | ; | 5.09 ] |
| $APE_B$ | -0.14 | [ | -0.53 | ; | 0.25 ] |
| N = 106,092 | | | | | |

| PANEL B: — WORST 25% HOSPITALS | | | | | |
|---|---|---|---|---|---|
| **Parameter** | **Point estimate** | **95% Confidence interval** | | | |
| $AMR_{00}$ | 9.59 | [ | 8.52 | ; | 10.65 ] |
| $AMR_{01}$ | 8.74 | [ | 7.71 | ; | 9.76 ] |
| $APE_A$ | -0.85 | [ | -2.33 | ; | 0.63 ] |
| $AMR_{10}$ | 8.51 | [ | 7.43 | ; | 9.59 ] |
| $AMR_{11}$ | 7.75 | [ | 7.10 | ; | 8.40 ] |
| $APE_B$ | -0.76 | [ | -2.03 | ; | 0.50 ] |
| N = 20,162 | | | | | |

| PANEL C: — BEST 25% HOSPITALS | | | | | |
|---|---|---|---|---|---|
| **Parameter** | **Point estimate** | **95% Confidence interval** | | | |
| $AMR_{00}$ | 3.10 | [ | 2.51 | ; | 3.69 ] |
| $AMR_{01}$ | 2.59 | [ | 2.19 | ; | 2.98 ] |
| $APE_A$ | -0.52 | [ | -1.23 | ; | 0.20 ] |
| $AMR_{10}$ | 4.81 | [ | 3.83 | ; | 5.79 ] |
| $AMR_{11}$ | 4.03 | [ | 3.70 | ; | 4.36 ] |
| $APE_B$ | -0.78 | [ | -1.81 | ; | 0.25 ] |
| N = 27,306 | | | | | |

Note.—Sample: CABG patients, age 65–99. AMR reported in percent; APE in percentage points.

Table 9. —Descriptive Statistics: Principal Diagnosis and Admission Types of CABG/PTCA Patients in Pennsylvania in 1989

| | **CABG** | **PTCA** |
|---|---|---|
| Admission Types | | |
| Emergency | 32.5 | 41.6 |
| Urgent | 26.6 | 30.1 |
| Elective | 40.2 | 28.6 |
| Principal diagnosis: All patients | | |
| CA | 93.1 | 84.3 |
| AMI | 5.3 | 14.0 |
| Principal diagnosis: Only elective patients | | |
| CA | 98.5 | 96.4 |
| AMI | 1.0 | 3.0 |
| N | 2,504 | 1,656 |

Note.—All numbers in percent (except sample size N). Sample: Patients, age 65–99 who received CABG or PTCA in Pennsylvania in 1989. Acronyms: CA = coronary atherosclerosis, AMI = acute myocardial infarction.

Table 10. —Estimation Results: Average Mortality Rates and Average Partial Effects (CABG only)

| **Parameter** | **Point estimate** | **95% Confidence interval** |
|---|---|---|
| $AMR_{00}$ | 4.37 | [ 3.83 ; 4.92 ] |
| $AMR_{01}$ | 4.63 | [ 4.16 ; 5.11 ] |
| $APE_A$ | 0.26 | [ -0.46 ; 0.98 ] |
| $AMR_{10}$ | 3.72 | [ 3.20 ; 4.24 ] |
| $AMR_{11}$ | 3.94 | [ 3.69 ; 4.18 ] |
| $APE_B$ | 0.22 | [ -0.35 ; 0.80 ] |
| $N = 39,548$ | | |

Note.—CABG Patients, age 65–99. AMR reported in percent; APE in percentage points.