Reference Pricing: The Case of Screening Colonoscopies

Marion Aouad*, Timothy T. Brown†, & Christopher M. Whaley‡

March 13, 2019

Abstract

We study the introduction of reference pricing to the California Public Employees' Retirement System. Reference pricing changes the relative price of using a hospital versus an ambulatory surgery center (ASC) for patients receiving a colonoscopy, leading to as good as random variation in patients' use of ASCs. We find a 10 percentage point increase in the share of patients using an ASC, leading to a \$2300 to \$1700 reduction in prices paid for patients who switch to ASCs. Our results suggest that the use of ASCs has a causal effect on prices paid and has no negative effect on patient health outcomes.

*Corresponding Author: maouad@stanford.edu; Stanford University School of Medicine, S-SPIRE

†University of California Berkeley, School of Public Health

‡RAND Corporation, University of California Berkeley, School of Public Health.

Thank you to David Card, Hilary Hoynes, Ben Handel, and Will Dow for invaluable comments. We also thank Patrick Kline, Jesse Rothstein, Casper Nordal Jørgensen, Ada Gonzalez-Torres, Jen Kwok, Zarek Brot-Goldberg, as well as participants in the UC Berkeley Labor Lunch, Labor Seminar, IRLE seminar, CLE seminar, ASHEcon, and the Berkeley Center for Health Technology. This project was supported by grant number R01 HS022098 from the Agency for Healthcare Research and Quality. Support is also provided by NIA grant T32-AG000246. Data on Anthem Blue Cross PPO enrollees were provided by Anthem, Inc. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality, CalPERS, or Anthem, Inc.

Health care spending currently accounts for approximately 17.5 percent of GDP and has grown at a rate faster than GDP over the last 50 years [CMS (2014)].[1] Reasons for this growth include new medical technologies, the aging population, and increasing medical prices (Blumenthal et al. (2013)). While these are important factors, consumer decisions regarding how much medical care to receive and where this care is received are also important. Many insurance plans use cost-containment strategies that affect consumer demand as a means of constraining "wasteful" medical spending.[2] These strategies typically expose consumers to a greater share of their medical expenses to reduce moral hazard.

Several studies have documented the wide degree of price dispersion for health care services (e.g. Cooper et al. (2018)), while other studies have noted that higher prices are a key reason for higher health care spending in the United States compared to other countries (Anderson et al. (2003), Papanicolas et al. (2018)). The large variation in prices presents sizable opportunities for savings. For example, in the data used in this study, moving patients from the 90th percentile to the 10th percentile would reduce the average expenditures by $2,594.

One common explanation for differences in observed prices is patient selection bias - higher prices simply reflect providers treating more medically complex patients. However, the validity of this claim is not well understood due to the inherent selection of patients to providers. For instance, higher-risk patients may be more likely to select higher-priced providers. Similarly, referring physicians may be more likely to refer higher-risk patients to more expensive health care providers.

Thus, understanding the causal effect of visiting a lower-priced provider on patient outcomes is important for determining the appropriateness of incentivizing the use of lower-priced providers. If higher-priced providers are more expensive due to differences in patient characteristics, then encouraging the use of lower-priced providers may be harmful to patients. However, if higher prices are not explained by differences in patient characteristics, then increasing the use of lower-priced providers may be an effective

[1] "Average annual health [inflation-adjusted] spending growth was 5.5 percent between 1960 and 2013 compared to 3.1 percent growth in GDP (Catlin, Aaron C. and Cowan, Cathy A. (2016))."

[2] Wasteful care can be defined as medical care for which the health benefits are smaller than the costs.

way to reduce waste. Studies that estimate the causal effect of visiting higher-priced providers on patient outcomes are rare, in large part due to the few natural experiments that exogenously shift patients to higher- or lower-priced providers.

In this paper, we are able to study the effect of moving patients to lower-priced providers by investigating the introduction of *reference pricing* by one of the largest purchasers of health care services in the United States, the California Public Employees' Retirement System (CalPERS). We specifically focus on the application of reference pricing to screening colonoscopy procedures. Within the context of this reform, our paper addresses the following questions: 1) How does changing the relative prices of health care for consumers, via increased consumer cost-sharing, affect the location where people receive their care, and 2) How does the change in the site of care translate to changes in mean prices, the distribution of prices, and the quality of care received?[3] By investigating this reform, we are able to determine how much is saved by the movement of patients to lower-priced providers.

While several papers have evaluated the intention-to-treat effects of the CalPERS reference pricing program (e.g. Robinson et al. (2015)–a, Robinson et al. (2015)–b, Robinson et al. (2015)–c), this paper uses the natural shock in patient cost-sharing caused by the program as an instrument to estimate the causal effects of receiving care from lower-priced providers on prices paid and patient outcomes for those patients who comply with the policy. We also consider distributional impacts.

Our analyses show that the CalPERS program provides strong incentives to receive medical care from ambulatory surgery centers (ASCs), which tend to be lower-priced than hospital outpatient departments (Hospitals). Our first-stage estimates show that the program leads to a 10 percentage-point (14.6 percent) increase in the use of ASCs. The estimated local average treatment effect (LATE) indicates a reduction in average procedure prices of approximately \$2300. Controlling for potential changes in the prices set by hospitals and ASCs in response to reference pricing, we find an average reduction in prices paid of approximately \$1700.[4] Our analyses also indicate that a one dollar increase

[3]The term *price* represents both the procedure price set by the hospital/ASC and the procedure price paid by the patient and employer/insurer. Where needed, we provide additional distinction.

[4]Our strategy also controls for general price changes due to inflation.

in patient cost-sharing leads to a reduction in the average total price paid for a procedure of approximately \$0.26. At the mean, this suggests a price elasticity of per-procedure prices paid with respect to patient cost-sharing of approximately -0.05.

Regarding distributional effects, we find that using an ASC leads to a 67 percentage point increase in the probability of prices paid being below \$1500. We also find that the increase in the probability of being below a particular threshold decreases for compliers as we move up from the reference price of \$1500. This finding indicates that the largest relative increases in the mass of the price distribution occur at regions below the reference price threshold.

We also estimate the LATE for medical complications associated with screenings. We find no change in the probability of patients experiencing a medical complication (serious or non-serious). Finally, utilization estimates suggest that the share of people receiving screenings does not appear to be affected by the introduction of reference pricing.

This work fits into an extensive literature that analyzes consumers' responses to changes in the price of medical care. These studies find that consumers are price sensitive and decrease medical spending in response to increased prices. Examples of this work include the analysis of consumers' response to the introduction of high-deductible health plans as discussed in Brot-Goldberg et al. (2017) as well as changes to the deductible and rates of cost-sharing in the RAND Health Insurance Experiment (Joseph P. Newhouse and RAND Corporation (1993), Aron-Dine et al. (2013)). Related work by Robinson et al. (2015)–a, Robinson et al. (2015)–b, Robinson et al. (2015)–c, and Robinson et al. (2016) study the reduced form effects of reference pricing over several medical procedures.[5] Additionally, Whaley et al. (2017) discuss the moral hazard effects of the reference pricing program while Aouad et al. (2016) discuss the unconditional distributional effects of the reference pricing program, by evaluating the quantile treatment effects of the program. Similar studies show that narrow and tiered insurance networks, which also use variations in patient cost-sharing to incentivize the use of lower-priced providers, lead to changes in patient behavior and reduce spending (e.g. Sinaiko and Rosenthal (2014), Gruber and McKnight (2016), Prager (2018)).

[5]These procedures include hip and knee replacement, arthroscopy, diagnostic and screening colonoscopies, and cataract surgery. The previous studies did not use instrumental variables.

Our study also relates to the literature on "cream-skimming" at ASCs. If cream-skimming occurs, it is difficult to compare average expenditures and quality outcomes between ASCs and hospitals as these differences could be attributable to patient selection (Munnich and Parente (2014)). Consistent with this, David and Neuman (2011) and Munnich and Parente (2018) find that ASCs tend to treat healthier patients, while Plotzke and Courtemanche (2011) find that ASCs perform more-profitable surgeries. Additionally, Munnich and Parente (2018) control for patient selection into ASCs (versus hospitals) using Medicare's predicted ASC payments as an instrument, and find that patient health outcomes are better at ASCs. However, a better understanding of the financial savings from movement to ASCs is still needed.

Thus, this work contributes to the literature in several ways. First, we evaluate a specific exogenous policy that uses financial incentives to encourage the use of ASCs: reference pricing. By using this policy as an instrument, we are able to add to the literature by examining changes in financial outcomes, such as average prices paid, as well as changes in health outcomes that are causally due to the movement of patients to ASCs. Second, the setting analyzed is one of the few instances where cost-sharing has been introduced on a preventive medical procedure after the Affordable Care Act was implemented.[6] In this setting, concern over reduced utilization of medically recommended, preventive care may be mitigated by the existence of a no-cost option (the ASC). Finally, by using the claims data for a large public agency such as CalPERS paired with a comparison group consisting of California members of Anthem Blue Cross, we believe our results have strong external validity to populations within and outside of California.

The rest of this paper proceeds as follows; Section 1 discusses the reference price program details and background. In Section 2 we describe the data. Section 3 discusses the methodology used to estimate the program's impact on cost-savings. Section 4 presents the empirical strategies used. Section 5 presents the results and is followed by the conclusion.

[6] The Affordable Care Act expanded patient coverage for medical screenings that have been given a recommendation of "A" by the US Preventive Services Task Force.

1 Reference Pricing Background

Reference pricing was introduced in January 2012 for enrollees in CalPERS Preferred Provider Organizations (PPOs), which consisted of approximately 22 percent of CalPERS active health plan members in 2011. The program set a maximum reimbursable amount of \$1500 for colonoscopy screenings received at hospitals.[7] At ASCs, reimbursements remained the same. Figure 1 helps illustrate the program in the simplest case where there is no health plan deductible or co-insurance. At a hospital, patients pay nothing until the total expense is above \$1500, but are responsible for every dollar spent above this threshold. Additionally, the expense incurred above the \$1500 threshold does not go toward the annual deductible or annual out-of-pocket maximum. In contrast, if patients choose an ASC, they remain along the horizontal axis in Figure 1.

Reference pricing was chosen as a strategy to reduce health care spending over alternative strategies, such as high-deductible health plans (HDHPs). The program was seen by CalPERS as more palatable to enrollees than HDHPs, as it allows enrollees to avoid higher prices simply by changing their site of care. In the particular case considered, hospitals are generally more expensive than ASCs. For example, Figure 2 shows the per-procedure variation in prices paid for screening colonoscopy at ASCs and hospitals in the three years prior to reference pricing implementation. It also shows that hospitals are generally higher-priced.[8] However, CalPERS determined that there was no discernible difference in the quality of care provided by hospitals and ASCs.

Price variation across health care facility types may be driven by the fact that Medicare compensates hospitals at a higher rate than the rate it pays to ASCs (see Munnich and Parente (2014)), which may affect the way private insurers reimburse the two facility types, as discussed in Clemens and Gottlieb (2017). Reasons for the higher facility fees at hospitals include the higher costs associated with running a hospital, such as added regulatory requirements.

Hospitals and ASCs also vary along other dimensions. For example, ASCs are typ-

[7]Prior to January 2012, no such maximum had existed. Exemptions from the reference pricing program were made for those who lived more than 30 miles from an ASC and for those who had a medical condition that required the use of a hospital for their screening.

[8]The per-procedure price refers only to the facility price and any complications occurring on the day of the procedure.

ically smaller, free-standing facilities that specialize in fewer medical procedures. They are not required to accept uninsured patients and only perform outpatient procedures. This is in contrast to hospitals, which perform a larger and more diverse set of medical procedures, are required to serve the uninsured in many circumstances, and offer inpatient, outpatient, and emergency room services. Additionally, Munnich and Parente (2014) find evidence that ASCs are faster at performing a given procedure than hospitals, suggesting that part of the difference in prices between hospitals and ASCs may be due to the efficiency gains that ASCs derive from procedural specialization.

Reference pricing was only applied to facility fees, which are one component of a procedure's overall price. We refer to the facility fee as the procedure price throughout this paper, as facility fees are generally the largest component of the expenses associated with a screening colonoscopy. Facility fees cover the operating expenses associated with running a hospital or ASC. Other expenses that a patient may face when receiving a screening are physician fees. [9] The patient is also generally responsible for ancillary services (e.g. laboratory tests).

CalPERS PPO health insurance plans are administered by Anthem Blue Cross Health Insurance ("Anthem"). As an insurance benefits administrator, Anthem negotiates prices with individual medical practices, ASCs, and hospitals for the members of CalPERS, among other responsibilities. Anthem also administers the health care benefits for other California PPO health plans besides CalPERS. The members of these other California PPO plans form our comparison group. Thus, CalPERS members and comparison group members face the same price at any given facility (ASC or hospital) conditional on services delivered, which is a function of health status.[10]

Lastly, for purposes of context, screening colonoscopy is a medical procedure in which the large intestine is examined using a lighted scope. It is one colorectal cancer screening method recommended by the US Preventive Services Task Forces.[11] Adults age 50 years

[9]Physician fees for screenings colonoscopies performed by in-network physicians are not billed to patients. When physician fees are non-zero, they make up approximately 30 percent of the procedure's total price.

[10]The provisions of the health plan such as coinsurance and deductible, can vary across CalPERS PPO members and the members of the comparison group. However, since screenings are not subject to coinsurance and deductibles, both the comparison group and CalPERS enrollees face the same price at a given health care facility.

[11]Source: American Cancer Society (2016).

and older are recommended to get a screening every ten years.[12]

2 Data

We use medical claims data for CalPERS PPO members and non-CalPERS Anthem PPO members, residing in California, who visited a California ASC or California hospital during the sample period. The non-CalPERS Anthem PPO members serve as a comparison group. The provider networks are established at the Anthem-level, so the non-CalPERS PPO members have access to the same set of providers.

The data set consists of repeated cross-sectional colonoscopy claims data from 2011 to 2013. We use 2011 as the pre-period because in that year both CalPERS members and members of the comparison group had no cost-sharing for screenings per the requirements of the newly-implemented Affordable Care Act (ACA).[13] [14] Prior to 2011, comparison group enrollees were subject to cost-sharing; CalPERS enrollees were not. The years, 2012-2013, represent the post-period.

The data set includes information on each procedure's total price (the sum of the price paid by the insurer plus the price paid by the patient), the date of the procedure, and the name and zip code of the health care facility at which the procedure was performed. We use the facility name to categorize health care facilities as either an ASC or a hospital, based on the presence or lack of an emergency room. The presence of an emergency room identifies the facility as a hospital, otherwise a facility is an ASC. [15] Additionally, we use the zip code of the health care facility to classify the Hospital Referral Region (HRR) of the facility. HRRs represent health care markets. Prices have been shown to vary by

[12]Other methods to check for colon cancer are fecal occult blood testing and sigmoidoscopy (US Preventative Task Force (2016)).

[13]Source: The Henry J. Kaiser Family Foundation (2015). The ACA mandated that private insurers could not impose cost-sharing on preventive services, such as colonoscopy, that had been given an "A" rating by the U.S. Preventive task force.

[14]Previous evaluations of the CalPERS program have included 2009-2010 pre-implementation data. We exclude these years in order to isolate the effects of the reference pricing program from the ACA's policy, where the latter was introduced in 2011.

[15]Our classification procedure may not be exact. We also try and gauge the number of services provided at the health care facility – generally, if there are fewer than five, we classify the facility as an ASC. Additionally, the name of the facility can be telling. For example, a name such as, "Endoscopy Center of Los Angeles," would be categorized as an ASC.

region and may be correlated within an HRR; (see Cooper et al. (2018)).[16]

Patient demographic information includes gender, age (18 - 64 years), and zip code (used to classify the patient-specific HRR). Additional patient information includes both ICD-9-CM (International Classification of Diseases, Ninth Revision, Clinical Modification) and CPT (Current Procedural Terminology) codes, which are medical diagnosis and procedure codes. These are observed over the previous year for all patients. We use these to create the following variables.

– We create a Charlson Comorbidity Index, which is a standard measure of the expected impact of comorbid health conditions on mortality. In the Charlson Comorbidity Index, each diagnosed condition (up to 22 conditions) from the last 12 months is assigned a score of 1, 2, 3, or 6, depending on the risk of dying associated with each one. Scores are summed to provide a total score. (Charlson et al. (1987)). This may be correlated with the procedure's price (i.e. the less healthy the patient, the more expensive the screening).

– We determine if a patient has a medical intervention during their screening (e.g. polyp removal, biopsy) using the diagnosis and procedure codes.

– We categorize post-screening complications into either 1) serious or 2) non-serious. Serious complications include intestinal perforation or bleeding in the intestine. Non-serious complications include paralytic ileus, nausea or abdominal pain.

– Following the algorithm discussed in (Robinson et al. (2015)), ICD-9-CM and CPT codes are used to distinguish between screening (i.e. preventive in nature) and diagnostic colonoscopies.[17]. This classification algorithm classifies colonoscopies in our sample as diagnostic by looking at the previous 6 months of medical and diagnosis codes to detect particular diagnoses.[18] Those colonoscopies in our sample not

[16] "HRRs represent regional health care markets for tertiary medical care that generally requires the services of a major referral center. The regions were defined by determining where patients were referred for major cardiovascular surgical procedures and for neurosurgery;" Source: The Dartmouth Atlas of Health Care (2016)

[17] In the data provided, we observe if a patient has a colonoscopy but no information on whether it was for screening/preventive or diagnostic purposes

[18] The particular diagnoses used to classify colonoscopies as diagnostic are: Clostridium difficile colitis, Ulcerative colitis, Crohn's disease, Ischemic colitis, Colitis NOS, Anorectal bleed, Melena, Gastrointesti-

classified as diagnostic (because they do not meet the medical conditions/diagnoses considered), are classified as screenings colonoscopies ("screenings").[19]

Table 1 gives the descriptive statistics of our sample. From the descriptive statistics, the comparison and CalPERS samples appear similar on many dimensions.[20] The comparison group has a slightly higher share of men obtaining screenings, however the age distribution across the two groups are similar and steady over time. Consistent with medical recommendations for the age at first screening (age 50), the average age observed in both groups is 55 and 56 years, respectively. We also observe that the Charlson Comorbidity Index is equal to zero for a similar share of comparison and CalPERS members and that a relatively similar share of patients receive some sort of intervention in the pre and post reference price periods.

3 Methodology

To measure how the introduction of reference pricing affects the use of ASCs and how this translates into changes in the price paid for a procedure and other related outcomes, we assume a heterogeneous treatment effects framework and identify the effect for the subset of compliers – that is, we identify the local average treatment effect (LATE). In this potential outcomes framework, treatment status, D_i, is influenced by exogenous variation in the binary instrument, Z_i. Our instrument, Z_i, is the exposure to reference pricing by being a member of CalPERS in the post-period (i.e. $Z_i = Treat \times Post$ in a regression that includes $Post$ and $Treat$). Thus, Z_i is equal to 1 if the individual is a

nal bleeding, Abdominal pain, Abdominal swelling, Abdominal tenderness, Abdominal bloating, Megacolon, Change in bowel habits, Diverticulitis or Diverticular hemorrhage, Volvulus, a history of Colorectal cancer, Iron deficiency anemia, Abnormal stool contents, and Fecal occult blood testing coded within the 6 months prior to colonoscopy. A colonoscopy observed in the data is also classified as diagnostic if a patient had a prior colonoscopy or barium enema within the last three years.

[19] We also exclude upper gastrointestinal colonoscopies and colonoscopies done in an in-patient setting. The facility type of the latter will presumably be less a function of the patients' choice.

[20] Differences in the mean patient prices paid at ASCs vs. hospitals seem to be driven by right-tail prices. The 75^{th} percentile of patient prices paid for 1) ASCs and 2) hospitals in the pre-period for CalPERS members and the comparison group are 1) \$0 and 2) \$0 for CalPERs and 1) \$511 and 2) \$617.40 for the comparison group. In the post-period, the 75^{th} percentiles for patient prices paid at ASCs and hospitals are 1) \$0 and 2) \$911 for CalPERS 1) \$223.20 and 2) \$230.90 for the comparison group.

member of CalPERS after reference pricing is introduced in 2012, and is 0 otherwise.[21]

The instrument induced variation in an individual's treatment status is defined as $D_i = D_{0i} + (D_{1i} - D_{0i}) \times Z_i$. Here, D_i takes value 1 if the i^{th} person uses an ASC and 0 if they use a hospital. This should hold for the i^{th} person in a given time period, t.

To capture the as good as random variation in our instrument, we use a difference-in-difference (DiD) model as our first stage equation, which will be displayed in the next section (Equation 3). This is a standard strategy in the applied economics literature (e.g. Duflo (2001); Hudson et al. (2015)). By using a DiD strategy, where we show common time trends between the comparison group and CalPERS and assume the existence of a time-invariant, group-specific effect, we capture the causal effect of exposure to reference pricing on ASC use. We believe that using this DiD strategy to capture the exogenous variation in the instrument and its effect on ASC use is valid because: 1) reference pricing was introduced with relatively little notice to CalPERS members (consisting of information conveyed during the annual open enrollment period) and 2) given that the recommendation for the receipt of a screening is related to age, there may be less ability for an individual to manipulate the timing of a screening.[22]Additionally, we have a valid comparison group for whom the parallel trends assumptions are likely to hold. This is because the comparison group faces the same price at any given health care facility, conditional on health status and health insurance plan provisions.

To test the parallel trends assumption between our comparison group and members of CalPERS, we run an event study and estimate the following:

$$\mathbb{1}(ASC_i) = \beta_0 + \sum_{r=-12, r\neq -1}^{7} \beta^r \mathbb{1}(L_i)_r + \sum_{r=-12}^{7} \gamma^r \mathbb{1}(L_i)_r \times \mathbb{1}(Treat_i) + X_i'\beta + \epsilon_i$$

$\mathbb{1}(L_i)_r$ equals 1 if person i is observed r quarters away from the introduction of reference pricing; it is 0 otherwise. We omit the lag dummy from the quarter before the introduction of the reference pricing program (i.e. We omit $\mathbb{1}(L_i)_{-1}$). $\mathbb{1}(ASC)$ is 1 if a person uses an ASC and is 0 if they use a hospital, while $\mathbb{1}(Treat)$ is 1 if an

[21] Treat equals 1 if an individual is a member of CalPERS and is 0 if they are part of the comparison group. Post equals 1 if a person is observed after reference pricing is introduced in January 2012. It is 0 if a person is observed before January 2012.

[22] Anecdotal evidence suggests little plan switching between the CalPERS health insurance plans, around the introduction of reference pricing.

individual is a member of CalPERS and 0 if they are part of the comparison group. X are patient demographic covariates, which are further discussed in the next section. Our estimates are presented in Figure 3 and show the per-quarter difference in the mean share using ASCs between CalPERS and the comparison group (i.e. the coefficient estimates of γ^r), controlling for patient characteristics. Figure 4 is analogous to Figure 3 but instead uses $Price_i$ as the outcome variable. Both figures confirm the validity of the parallel trends assumption by showing a relatively stable trend in ASC use between the comparison group and CalPERS before reference price implementation. Additionally, the figures provide visual evidence for the existence of the first stage and the intention-to-treat effects. Appendix Figures 6-9 show similar event study plots for the complication outcomes.

We estimate the LATE using a two-stage least squares (2SLS) estimator for the Wald DiD, which is the ratio of two DiD estimates (a similar approach is used by Duflo (2001)). This is presented below:[23]

$$\frac{\Big(E[Y_i|Treat_i=1,Post_i=1]-E[Y_i|Treat_i=1,Post_i=0] \\ -E[Y_i|Treat_i=0,Post_i=1]-E[Y_i|Treat_i=0,Post_i=0]\Big)}{\Big(E[D_i|Treat_i=1,Post_i=1]-E[D_i|Treat_i=1,Post_i=0]- \\ E[D_i|Treat_i=0,Post_i=1]-E[D_i|Treat_i=0,Post_i=0]\Big)}\Bigg/$$

To ensure identification of LATE and the interpretation of the instrumental variable (IV) estimate as the effect of ASC use on procedure prices for those who now use an ASC because of reference pricing, we make the standard assumptions as discussed by Imbens

[23] We also consider the conditions necessary to identify the LATE in a fuzzy design setting such as ours. Here, there is no sharp treatment - some members of CalPERS do not use an ASC while some members of the comparison group use ASCs after reference pricing is introduced (i.e. no group is fully treated/untreated). However, the rate of ASC use goes up by more for CalPERS after reference pricing is introduced than it does for the comparison group. Recent work by De Chaisemartin and D'HaultfŒuille (2017) show that the LATE can still be identified if the following two conditions hold: 1) The treatment rate in the comparison group is stable over time and 2) Time has the same effect on both means of the potential outcomes, $E[Y_{0i}]$ and $E[Y_{1i}]$. We can show that the first condition holds and that the share of ASC users in the comparison group is steady over time at approximately 71 percent. Our alternatively constructed price variables, that hold time factors fixed, should satisfy the second requirement.

and Angrist (1994). Firstly, this requires that the independence assumption be satisfied. In other words, we require that $\{(Y_i(D_{1i}, 1), Y_i(D_{0i}, 0), D_{1i}, D_{0i}\} \perp\!\!\!\perp Z_i$; $Y_i(d, z)$ is the potential outcome for person i if they had treatment status $D_i = d$ and the value of the instrument $Z_i = z$. Secondly, the instrument exclusion restriction, $Y_i(d, 0) = Y_i(d, 1) \equiv Y_{di}$ for $d = 0$, 1, must hold. Thirdly, the instrument must strongly influence the treatment, $E[D_{1i} - D_{0i}] \neq 0$ (i.e. the first stage condition must hold). The final necessary assumption is the monotonicity assumption: $D_{1i} \geq D_{0i}$ $\forall i$.

These assumptions appear valid in this context given the nature in which this program was implemented and our use of the DiD strategy to exploit the variation induced by our instrument. Additionally, we have evidence for the strength of our first stage and the assumption of no-defiers seems reasonable in this setting. However, there may be concern over the assumption that reference pricing has no direct impact on prices set by hospitals/ASCs. We discuss the validity of the instrument exclusion restriction in the next section.

3.1 Methodological Concerns

Our approach faces three key challenges. The first concern is whether we satisfy the instrument exclusion restriction. This assumption may not hold if we believe there are supply-side responses to reference pricing, namely that health care facilities respond to the introduction of reference pricing by changing the price they set/are willing to accept for colonoscopies. We test for this in the data and construct an alternate procedure price measure that is solely a function of pre-period covariates. With this alternative measure, potential supply-side responses should no longer be a concern and should strengthen our belief in the validity of the instrument exclusion restriction.

Similarly, we may be concerned that the reference pricing program leads patients to select different types of ASCs and hospitals, rather than a uniform shift from hospitals to ASCs. Such a scenario may violate the exclusion restriction if reference pricing affects the choices of the never takers by shifting them from higher-priced to lower-priced hospitals. This concern may be lessened by the absence of a price transparency tool. CalPERS members where given a list of ASCs, but did not have access to prices for specific ASCs or hospitals.

We may also be concerned about sample selection bias due to non-random reductions in the utilization of screenings. To test for this response, we construct utilization estimates using enrollment data provided by CalPERS. We compare this to the observed number of screenings in our data to see if there are changes in the overall utilization patterns for members of CalPERS. These results are presented in section 5.4

Finally, one may also be concerned about forward-looking behavior and if we are capturing the response of consumers to the "spot" price (i.e. the price right now) of medical care or some "forward price" (i.e. the price after taking into account all spending over the year). The presence of forward-looking behavior would imply that patients' anticipated spending in the year could affect their current behavior/choices because of the presence of a deductible (e.g. see Aron-Dine et al. (2015)). We do not believe this should be a concern in this setting because screening colonoscopies are not subject to deductibles or coinsurance, unlike many other medical procedures.[24] Additionally, any expenditure incurred above the \$1500 reference price threshold does not go toward the annual deductible or annual out-of-pocket maximum, thus it will not discount future care in any way.

4 Empirical Strategy

We present the structural, reduced form and first-stage equations below. Here, i indexes the individual and t represents time {pre-period, post-period}. We pool data from 2011 to 2013 and examine the mean effects of the program using Equation 2 and our Wald-DiD, where the outcome variable is $Price$. To understand the distributional impacts of the program, we use the outcome, $\mathbb{1}(Price_i \leq Z)$. By varying the value of Z, we can understand how the introduction of reference pricing changes the probability of observing a price below our specified cut-point or threshold.

[24] Screening colonoscopies are not subject to deductibles or coinsurance because they have an "A" rating from The US Preventive Services Task Force.

Structural Equation:

$$Y_i = \pi_0 + \pi_1 \times Post_i + \pi_2 \times Treat_i + \pi_3 \times ASC_i + X_i'\pi + e_i \tag{1}$$

Reduced Form Equation:

$$Y_i = \beta_0 + \beta_1 \times Post_i + \beta_2 \times Treat_i + \beta_3 \times Post_i \times Treat_i + X_i'\beta + \varepsilon_i \tag{2}$$

- $Y_i \in \{Price_i,\ \mathbb{1}(Price_i \leq Z)\}$
- $\mathbb{1}(Price_i \leq Z)\} = \begin{cases} 1, & \text{if the procedure price at chosen facility is} \leq \text{Z} \\ 0, & \text{if the procedure price at chosen facility is} > \text{Z} \end{cases}$
- $Z \in \{\$1500, \$2000, \$2500, \$5000, \$10000\}$

Our vector of covariates, X, includes the demographic characteristics of patients and is presented below along with the definitions for $Treat$, $Post$ and ASC:

- $X =< \mathbb{1}(\text{Male}), \mathbb{1}(\text{Age Category}), \text{Comorbidity Index},$
  $\text{Patient Health Care Market Dummy} >$

  $$Treat_i = \begin{cases} 1, & \text{if person i is a member of CalPERS} \\ 0, & \text{if person i is a member of the comparison group} \end{cases}$$

  $$Post_t = \begin{cases} 1, & \text{if the observation is from the post-period (2012, 2013)} \\ 0, & \text{if the observation is from the pre-period (2011)} \end{cases}$$

  $$ASC_i = \begin{cases} 1, & \text{if facility choice of person i is an ASC} \\ 0, & \text{if facility choice of person i is a hospital} \end{cases}$$

First Stage Equation:

$$ASC_i = \delta_0 + \delta_1 \times Post_t + \delta_2 \times Treat_i + \delta_3 \times Post_i \times Treat_i + X_i'\delta + v_i \tag{3}$$

5 Results

Table 2 presents our estimates of the first stage using Equation 3, without and with the inclusion of covariates. Our results show that the instrument is statistically significant in the first stage and is sufficiently strong, as evidenced by an effective F-statistic of 24.69. The F-statistic, which is estimated using the method of Montiel-Olea and Pflueger, which is robust to heteroscedasticity and autocorrelation (Olea and Pflueger, 2013), is also greater than the critical value of 23.11. This implies a worst-case bias of only 10 percent. The results also indicate an increase in the share of CalPERS enrollees using ASCs of approximately 10 percentage points when covariates are included.[25] Prior to reference pricing, the share of CalPERS members using ASCs was approximately 67 percent. Thus, the increase in the share of members using an ASC is approximately 15 percent.

Table 3 presents the reduced form, OLS estimates, and estimates of the LATE using 2SLS regression. In columns 3-6 ("OLS"' and "IV"), the estimated effect of using an ASC is given in each cell with the relevant outcome variables listed in the left-most column. Columns 3-4 give coefficient estimates using OLS while columns 5-6 give 2SLS regressions estimates. In columns 1-2 ("Reduced Form") each cell gives the estimate for β_3 from Equation 2, which measures how exposure to reference pricing affects each outcome variable.

The reduced form results show that exposure to reference pricing has a statistically significant effect on average per-procedure price for screenings among CalPERS members. Exposure to the program leads to a reduction in mean per procedure price of approximately \$224 (with covariates included in the regression). This is a relatively large reduction considering that the CalPERS pre-reference unadjusted mean price for a screening was approximately \$1814. For the IV estimates, we find estimates of the LATE to be statistically significant and larger than our reduced form estimates, with a reduction in the mean per-procedure price of approximately \$2300. We also include the OLS estimates for comparison to our IV estimates and see that they differ. To decompose the OLS estimate further, we can write π^{OLS} from Equation 1 as:

[25]Estimates of the increase in the share of CalPERS patients using an ASC due to reference price exposure are slightly smaller than those found in Robinson et al. (2015) because this analysis includes CalPERS members who are exempt from the reference price policy.

$$\pi^{OLS} = \underbrace{E[Y_{1i} - Y_{0i}|ASC_i = 1, X]}_{\text{ATT}} + \underbrace{\left(E[Y_{0i}|ASC_i = 1, X] - E[Y_{0i}|ASC_i = 0, X]\right)}_{\text{Selection Bias}}$$

The first term gives the average effect of ASC use on mean prices for those who choose to use ASCs, while the second term shows the selection bias. Because ASC use is not randomly assigned, and instead is a joint decision by both the patient and the patient's physician, selection bias will likely be non-zero causing our OLS estimates for the treatment-on-the-treated (ATT) to be biased. In a heterogeneous treatment effects framework, those who receive the treatment consist of always takers and compliers.[26] We do not necessarily expect the effect of treatment for these two groups to be equal, that is to say that the LATE does not necessarily equal the ATT.[27] For example, compliers may live in geographic areas where the difference between ASC and hospital prices is larger. This would imply that the LATE is larger in absolute value than the ATT, but could still imply that there is negative selection bias in our OLS estimates. We further explore complier versus always taker characteristics in Section 5.5.

To understand where in the price distribution patients move when facing reference pricing, we present the distributional impacts of the program in Table 3. These estimates show that the largest relative increase occurs for prices below the \$1500 threshold, with declines in the increased probability as Z increases. This is likely a function of the lower-price structure of ASCs. For example, the mean, median and 75^{th} percentile of procedure prices for CalPERS members who used ASCs in the pre-period were \$1335.68, \$730 and \$1273, respectively.

5.1 Testing for Market Effects

In this section, we explore the validity of our instrument exclusion restriction. To do this, we determine if health care facilities respond to reference pricing by changing the amount that they are willing to accept from Anthem insurance. This amount is referred to as the "allowed amount" (or "negotiated amount") and is the amount that the insurer

[26]Always takers are the subset of people who always receive the treatment, even when not encouraged to do so as determined by the instrument (e.g. $D_{0i} = D_{1i} = 1$).

[27]In the pre-period, ASC users consist of always takers while in the post-period, ASC users consist of always takers and compliers.

has agreed to pay a health care facility for the specific medical services performed by the facility.[28] This test is related to work by Whaley and Brown (2018). To test for a supply-side response, we assume that the responsiveness to reference pricing is a function of facilities' pre-reference pricing exposure to CalPERS members. Thus, facilities with a higher share of CalPERS members may respond differently than those with a lower share. To analyze this hypothesis, we construct a measure of pre-reference pricing period exposure to CalPERS member, $\widetilde{Treat}$, for those health care facilities that we observe in both the pre and post-reference pricing periods. Thus, $\widetilde{Treat}$ is in the [0,1] interval.[29] One caveat for our measure, $\widetilde{Treat}$, is that it will not perfectly capture a facility's share of patients who are from CalPERS if there are many other non-Anthem patients who use the facility (we do not observe these patients). If this share of patients with non-Anthem insurance is high, then we likely overestimate the share of a facility's patients who belong to CalPERS.

Specifically, for the i^{th} screening colonoscopy procedure, at the n^{th} health care facility, in time period t (pre, post), we pool data from 2011 to 2013 and estimate:

$$P_{in} = \gamma_0 + \gamma_1 \times Post_i + \gamma_2 \times Post_i \times \widetilde{Treat}_n + X_i'\gamma + \alpha_n + u_{in} \tag{4}$$

Here, $P_{in} \in \left\{ln\left(NegotiatedAmount_{in}\right), NegotiatedAmount_{in}\right\}$.
$\widetilde{Treat}_n = \frac{CalPERS_{nN}}{CalPERS_{nN}+Control_{nN}}$, is the ratio of the number (N) of CalPERS members observed at facility n, ($CalPERS_{nN}$), to the sum of all patients in our data observed at facility n, ($CalPERS_{nN} + Control_{nN}$), in the year prior to reference pricing (2011). Lastly, α_n represents a health care facility dummy.

We estimate Equation 4 separately for ASCs and hospitals, using both the natural log of the negotiated amount and the negotiated amount (in \$'s) as outcomes.[30] We limit

[28] The allowed amount is the amount that is accepted by health care facilities that are in-network, by definition of their in-network status. If a facility is out-of-network, the allowed amount exists and is "the amount that Anthem Blue Cross or the local Blue Cross and/or Blue Shield Plan determines is appropriate considering the particular circumstances and the services rendered." (CalPERS Plan Booklet (2012)). Thus, the allowed amount set by the insurer may not be accepted as full payment for services performed; patients would then be responsible for the excess.

[29] The minimum and maximum observed for $\widetilde{Treat}$ are 0 and 1, respectively. However, the median and 75^{th} percentiles for health care facilities for observations in the sample are 0.156 and 0.25, which suggests most facilities have a relatively low exposure to CalPERS in the pre-period, based on our measure.

[30] Our estimates will be biased by the inclusion of out-of-network facilities if the negotiated amounts

our samples to patients who go to health facilities observed in both the pre and post reference pricing periods.

Table 4 shows the health care facility's response results and demonstrates that results are sensitive to the model specification. Although no effects are found for the hospital sample, in the case of ASCs, we find that in the log specification, ASCs whose 2011 exposure to CalPERS patients was higher have lower negotiated prices in the post-period than those with a lower share of CalPERS patients. Specifically, the sign on our estimate for γ_2 when the outcome is $ln(NegotiatedAmount)$ among the ASC sample is negative with a coefficient estimate of -0.26. It is also statistically different from zero at the 5 percent significance level.[31] These negative estimates may occur because of the bargaining power of Anthem insurance. Ho and Lee (2013) and Gaynor and Town (2011) discuss the importance of (hospital) bargaining power in price setting with insurers. These results may be explained if health care facilities that have lower levels of bargaining power are willing to accept a lower price for increased volume.

5.2 Controlling for Market Effects

To control for any potential market effects, we construct two alternative price variables that are solely a function of pre-period data. Doing this should remove any potential supply-side effects as well as other price changes due to factors like inflation. Thus, our results should only reflect the savings from patients' movement to ASCs. We begin by limiting our sample to those patients who go to a health care facility that is observed in both the pre and post-periods.[32] From this sample, we define two alternative procedure price variables.

We construct the first alternative dependent variable using an exact matching strategy

listed for them are artificially low since these facilities have not agreed to the price set by the insurer. For example, the set allowed amount for out-of-network ASCs is \$350 and \$380. For these ASCs, we use the sum of the insurer amount paid plus the patient amount paid (\$'s) to get the amount that these facilities are charging. The same thing is done for hospitals since the maximum allowed amount is now \$1500 for CalPERs members.

[31] Negative coefficient estimates suggest that increased pre-period exposure to CalPERS results in lower negotiated amounts than would have been realized if the CalPERS exposure was lower. This may be indicative of a slower growth in negotiated amount/prices among the facilities more likely to be visited by members of CalPERS.

[32] This restriction removes approximately 8 percent of our total observations

to match patients observed in the post-period, to patients observed in the pre-period that have the same observable characteristics. Specifically, we group observations into "cases" defined by 1) The specific health care facility used 2) Whether or not the patient had any medical intervention and 3) The patient's Charlson Comorbidity Index.[33] We create 1,113 unique cases using this strategy, leaving approximately 1.63 percent unmatched observations from the pre and post periods. Within these cases, we take the mean procedure price for the pre-period observations and apply this mean to every observation within the case.[34]

The second alternative dependent variable is defined using a regression adjustment strategy, which generates price estimates that are purely a function of pre-period observables. Specifically, we predict what a person observed in the post-period would have paid had they been observed in the pre-period instead. To determine this, we limit our sample to pre-period only observations and estimate the coefficients from the regression, $Price_i^{Pre} = X_i^{Pre} \times \beta^{pre} + \alpha_n^{pre} + u_i^{pre}$. X consists of the covariates discussed in the Empirical Section and α_n^{pre} represents a facility fixed effect. From this regression, we generate predicted prices for post-period observations as $\widehat{Price_i^{Post}} = X_i^{Post} \times \hat{\beta}^{pre} + \hat{\alpha}_n^{pre}$.[35]

Our result are shown in Tables 5 and 6. Both tables give the reduced form, OLS, and 2SLS regression estimates using the two alternative price definitions. The estimates are similar in direction to those presented in Table 3, although they are generally smaller in magnitude, depending on the model specification. In Table 5, the estimate of the LATE for average price reductions when using the exact matching definition is statistically significant and is -$1670 (versus the analogous estimate of approximately -$2300 presented in Table 3). Additionally, the increased probability of being below $1500 among compliers is equal to approximately 60 percent in contrast to an approximately 67 percent increase in Table 3. Using the regression adjusted definition, we find similar estimates across our specifications in Table 6. Together, these results suggest that supply-side responses

[33] We think these are the relevant observables on which to categorize since insurer negotiated prices vary at the facility-intervention status level, as discussed earlier.

[34] We omit cases that do not contain of at least one pre and one post-reference pricing observation (i.e. we omit those cases for whom only pre or post period observations are contained in the case).

[35] We use the actual observed price for those patients observed in the pre-period. The R^2 from this regression is approximately 0.84. We also limit the sample to those with a minimum procedure price above $300; the minimum from 2011.

may not be a large contributor to the price reductions associated with reference pricing. Rather, it appears that the reductions in prices paid are largely driven by patient's switching from high-priced to lower-priced sites of care.

5.3 Cost-Sharing Impact: An Alternative Approach

To further understand the effects of reference pricing on compliers, we estimate how changes in cost-sharing, induced by the policy, affect the total price of the facility chosen by CalPERS patients. To examine this, we estimate the cost-sharing payments that a patient would face if they visited either an ASC or a hospital. Since patients are observed at only one of the two facilities, we impute a patient's cost-sharing for the facility type at which they are not observed. For example, for those patients who use ASCs, we impute the cost-sharing amount they would face if they instead, visited a hospital. The method is further discussed in the Appendix. We then compute the difference in cost-sharing that a patient would face across a hospital and an ASC, (i.e. cost-sharing payments at the hospital minus cost-sharing payments at the ASC). We refer to this cost-sharing difference as $Pay\,Diff_i$ and estimate how increased cost-sharing differences, stemming from the introduction of the policy, affect the price of the facility chosen by patients. The equations estimating this relationship are presented below. Equation 5 is the first-stage equation and the parameter of interest, δ_3, tells us how exposure to reference pricing affects the cost-sharing difference across hospitals and ASCs. Equation 6 represents the second-stage equation and the parameter of interest, θ_3, indicates how changes in the cost-sharing difference, induced by exposure to reference pricing, affect the price of the facility selected by patients.

$$Pay\,Diff_i = \delta_0 + \delta_1 \times Post_i + \delta_2 \times Treat_i + \delta_3 \times Post_i \times Treat_i + \delta' X + \epsilon_i \quad (5)$$

$$Price_i = \theta_0 + \theta_1 \times Post_i + \theta_2 \times Treat_i + \theta_3 \times \widehat{Pay\,Diff}_i + \theta' X + u_i \quad (6)$$

The results are presented in Table 7. The first-stage results indicate that exposure to reference pricing does affect the difference in cost-sharing across hospitals and ASCs, as expected. The coefficient estimate for δ_3 is approximately \$926.36 and is statistically significant at the 1% significance level. Also, the F-Statistic estimated using the method of Montiel-Olea and Pflueger is approximately 41 and is statistically significant at the 5% significance level (Olea and Pflueger (2013)). This result implies that exposure to reference pricing increases the difference in cost-sharing across ASCs and hospitals by about \$926, on average, which is in line with the program's impact on hospital cost-sharing. The second-stage results, which are statistically significant at the 5% level, indicate that if the cost-sharing difference increases by \$1, due to exposure to reference pricing, the price of the facility chosen decreases by approximately \$0.26. Using the average cost-sharing amount and average procedure price incurred among CalPERS patients in the pre-period, this result implies a point elasticity estimate of -0.05. This estimate is smaller than the RAND HIE estimate of -0.2, but is consistent with that found in Ellis et al. (2017) who estimate smaller elasticities ranging from -0.02 to -0.11 for preventative services.

5.4 Health Outcomes

To better understand the health impacts of this program, we estimate the LATE for several patient outcomes. We focus on the following measured patient outcomes: 1) Any Complications (1-30 days post-screening) 2) Serious Complications (1-30 days post-screening) 3) Non-Serious Complications (1-30 days post-screening) and 4) Whether or not a patient has any medical intervention performed during the screening (i.e $\mathbb{1}$(Intervention)).[36] The health outcome variables are defined as follows:

$$HealthComplication_i = \begin{cases} 1, & \text{if person } i \text{ has a health complication} \\ 0, & \text{Otherwise} \end{cases}$$

Table 8 presents the LATE for the four health outcomes discussed above, with each cell giving the coefficient estimate on ASC from a 2SLS regression using $Post \times Treat$ as an instrument for ASC choice. Almost all quality outcome coefficients are negative (the exception is interventions), but none of the regression estimates for the LATE are statistically different from zero. This result implies that the CalPERS program does not

[36]Serious complications include intestinal perforation or bleeding in the intestine while non-serious complications include paralytic ileus, nausea or abdominal pain and are determined using the medical CPT and ICD-9M codes of the patient. Any complications encompasses serious and non-serious complications and also includes cardiac and anesthesia-related complications

have negative health impacts on patients. It also implies that the quality of care received for compliers, as measured by these outcomes, is similar at ASCs and hospitals.[37] These results imply that, at least along the quality dimensions we study, ASCs have no worse outcomes than hospitals. Our reduced form results analyzing health outcomes, presented in Table 12 of the Appendix, also support this finding as do the reduced-form event studies (Figures 6 - 9 of the Appendix), which show that there is no change in health outcomes between CalPERS and comparison group patients after reference pricing is introduced.

We are also interested in whether CalPERS members change their utilization of screenings in response to the reference pricing program. Utilization reductions are often a concern when implementing policies that increase patients' medical cost-sharing and are especially concerning if they occur for preventative procedures, since these are medically encouraged. To analyze this question, we obtain PPO enrollment data from CalPERS for those who "should" be obtaining a screening based on their age (i.e. the population of 50-64 years olds)[38] We then create our utilization estimate, which is the ratio of CalPERS screenings observed in the data (for those ages 50-64 years) to the CalPERS enrollment number (for those ages 50-64 years). Unfortunately, we do not have enrollment data for the comparison group health plan.

Figure 5 plots these estimates of CalPERS screening utilization over time. Screening utilization rates are relatively steady between the year before reference price implementation and the years after it was introduced. Specifically, the CalPERS utilization rates calculated are 3.5 percent, 3.6 percent, and 3.5 percent in 2011, 2012, and 2013, respectively. Thus, this provides suggestive evidence that utilization rates may not have been impacted by reference pricing.

5.5 Complier Characteristics

We are also interested in the characteristics of compliers in order to understand how they differ from those who used ASCs prior to reference pricing (i.e. the always takers) and those members of CalPERS who continue to use hospitals after reference pricing

[37]In the medical literature, a good measure of the quality of a colonoscopy is the withdrawal time. This measures the length of time taken to remove the scope from the colon. We do not observe this in the data. Source: Rex et al. (2015).

[38]Enrollment data tell us the number of people enrolled in a health plan.

implementation (i.e. the never takers). To the extent that compliers are similar to the average population considered, this would imply that the results of this analysis are more widely extrapolatable and provide more insight into the average price reductions that can result from patient movement to ASCs. To estimate mean complier characteristics, we implement the estimation strategy of Abadie (2002).

Table 9 presents our results. We see that the share male appears to be lower among compliers as is the share that are between 50 and 59 years old. However, while the average age is similar across the groups, the share that are between 60 and 64 years old is higher among the compliers. Additionally, the share with a comorbidity index equal to 0 is lower, while the share with a medical intervention is higher in the complier versus always taker and never taker samples. As shown in Table 8, we test for changes in the probability of having an intervention driven by increased ASC use due to reference pricing, and do not find estimates that statistically differ from zero. This suggests that while the average intervention rate may be similar at hospitals and at ASCs among compliers, the probability of needing an intervention appears to be slightly higher for compliers, relative to the other two groups. Compliers are also more likely to use a facility within their own HRR than other groups. Our results indicate that approximately 90 percent of compliers visit a facility within their own HRR while that estimate is approximately 82 percent for always takers and 64 percent for the never takers.

Additionally, we find that compliers are more likely to come from Northern California versus always takers and never takers. This may be a result of differences in the mean price between Northern California hospitals and ASCs, which is generally large compared to other regions. In the pre-period, the difference between the mean hospital and mean ASC price was approximately $1840 in Northern California. In contrast, this difference was approximately $924 in Southern California and $1225 in Central California. This difference in complier characteristics may explain the difference between the OLS and IV results. The potential savings from shifting demand to ASCs are larger in the higher-priced Northern California market than in the lower-priced Southern California market.

5.6 Welfare Analysis

To understand the efficiency gains or losses associated with the reference pricing program, we determine if the program is efficient in the sense of Kaldor (1939) and Hicks (1939).[39] To do this, we compare the savings ($'s) generated from the movement of the compliers to ASCs to the losses ($'s) experienced by those members of CalPERS who use hospitals after reference price implementation (i.e. the never takers). Assuming a constant share of always takers, there are approximately 68 percent of CalPERS members who are classified as always takers.[40] With 7,243 members of CalPERS observed in the post-period and 5,685 members of CalPERS using an ASC, this implies that there are approximately 4,925 always takers and 760 compliers (i.e. $5,685 - 7,243 \times 0.68$).

Taking the number of compliers, we estimate the average total savings generated by their move to ASCs from hospitals to be approximately $1,748,000 (= $-\$2300 \times 760$), where we use the LATE on procedure prices from Table 3. We observe the additional out-of-pocket amount ($'s) paid by CalPERS members who use hospitals in the post-period in the data. The additional out-of-pocket amount paid by never takers is equal to: $H_i = Price_i$ - $1500. We exclude from our analysis those hospital-goers with a procedure price less than $1500 (N = 158), leaving us with 638 CalPERS members who use hospitals in the post reference pricing period.[41] For these 638, $\sum_{i=1}^{638} H_i = \$855,810.70$. This implies an average per person additional expense of approximately, $1341.40. Thus, the overall savings from this program for 2012 and 2013, accounting for the never taker losses, are approximately $892,189.30.

Applying the continuous mapping theorem, we compute the 95 percent confidence interval for our total savings estimate.[42] Using our estimates in Table 3, the 95 percent confidence interval for our savings estimate from compliers switching to ASCs is [$326,921,

[39] Here, we assume that total cost is equal to marginal cost. However, to the extent that hospital prices reflect subsidizations of free care (e.g. to indigent populations), we are not able to capture this.

[40] ASC_{2009} = 68 percent, ASC_{2010} = 68.97 percent, ASC_{2011} = 67.41 percent for members of CalPERS.

[41] Those with procedure prices less than $1500 should not have an additional out-of-pocket expense. We also exclude the CalPERS members who are exempt from the reference price program since they did not face additional out-of-pocket expenses after the program implementation.

[42] Let $G(\hat{\beta}) = J \times \hat{\beta}$, where J is equal to the number of compliers and $\hat{\beta}$ is the estimate for the LATE on total costs. By the continuous mapping theorem, $s.e.(G(\hat{\beta})) \to J \times s.e.(\hat{\beta})$. This implies that the 95 percent confidence interval is: $G(\hat{\beta}) \pm 1.96 \times \left(J \times s.e.(\hat{\beta})\right)$.

$3,169,078]. Given that we directly observe the amount paid by the never takers in the data, we do not compute a confidence interval for this amount. Our estimate for the additional expenses incurred by the never takers (i.e. $\sum_{i=1}^{638} H_i$) is included in the 95 percent confidence interval. Thus, while the estimated savings of $892,189 are positive, we cannot say that they are statistically different from zero.

6 Conclusion

In response to high prices and a wide variation in these prices, many employers and insurers have implemented innovative insurance benefit designs. One such design, reference pricing, uses targeted financial incentives to direct patients to lower-price providers. Understanding the mechanisms by which savings are achieved as well as the related health consequences of such policies are important to determine their appropriateness. This study examines a reference pricing program for screening colonoscopies by CalPERS. By focusing on the program design, which incentivized CalPERS members to choose lower-priced ASCs over hospitals, we are able to understand how shifts toward ASCs affect average prices paid and the distribution of prices paid.

Using both DiD and IV-DiD strategies, we show that the CalPERS reference pricing program leads to a large increase in the share of patients using ASCs (approximately 10 percentage points). We find that the associated reductions in average per-procedure prices range between approximately $1700 and $2300, depending on model specification. Additionally, the largest relative increase in the distribution occurs for the area below the reference price threshold of $1500, likely driven by the lower ASC pricing structure. Also important, we show that there are no health complications arising as a result of this program and provide evidence that screening colonoscopy utilization may not be greatly affected.

Future areas for research include determining to what other types of medical procedures reference pricing could be similarly applied. While, this study focuses on a single setting, the observed price variation which drives the effectiveness of reference pricing, are present in many health care markets (Cooper et al. (2018)). For services like screening colonoscopies where there are different health care facility types that provide a similar quality of care, the introduction of a reference pricing program reduces spending without impacting measurable quality of care. While CalPERS introduced reference pricing for three other procedures (arthroscopy, cataracts, and hip and knee replacement surgery) around the same time, a better understanding of how extrapolatable our results are to a broader setting is still needed. To the extent that CalPERS members are similar to other health consumers within and outside of California, our estimates may be externally valid. Understanding the supply-side responses, including how many more ASCs would be needed to meet the potentially increased demands for service as well as the associated changes in negotiated prices due to a changing market structure, are also important considerations.

References

Abadie, A. (2002). Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models. *Journal of the American Statistical Association 97*(457), 284–292.

American Cancer Society (2016). `www.cancer.org` [Accessed: September 2016].

Anderson, G. F., U. E. Reinhardt, P. S. Hussey, and V. Petrosyan (2003, May). Its The Prices, Stupid: Why The United States Is So Different From Other Countries. *Health Affairs 22*(3), 89–105.

Aouad, M., T. T. Brown, and C. Whaley (2016). Understanding the Distributional Effects of Health Insurance Reforms: Evidence from a Consumer Cost-Sharing Program. *UC Berkeley Working Paper*.

Aron-Dine, A., L. Einav, and A. Finkelstein (2013). The RAND Health Insurance Experiment, Three Decades Later. *The Journal of Economic Perspectives 27*(1), 197–222.

Aron-Dine, A., L. Einav, A. Finkelstein, and M. Cullen (2015). Moral Hazard in Health Insurance: Do Dynamic Incentives Matter? *Review of Economics and Statistics 97*(4), 725–741.

Blumenthal, D., K. Stremikis, and D. Cutler (2013). Health Care Spending—a Giant Slain or Sleeping? *New England Journal of Medicine 369*(26), 2551–2557.

Brot-Goldberg, Z. C., A. Chandra, B. R. Handel, and J. T. Kolstad (2017). What Does a Deductible Do? The Impact of Cost-sharing on Health Care Prices, Quantities, and Spending Dynamics. *The Quarterly Journal of Economics 132*(3), 1261–1318.

CalPERS Plan Booklet (2012). `https://www11.anthem.com/ca/national/f0/s0/t0/pw_e175308.pdf?refer=popcontent` [Accessed: October 2016].

Catlin, Aaron C. and Cowan, Cathy A. (2016). History of Health Spending in the United States, 1960-2013. `www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/Downloads/HistoricalNHEPaper.pdf` [Accessed: September 2016].

Charlson, M. E., P. Pompei, K. L. Ales, and C. R. MacKenzie (1987). A New Method of Classifying Prognostic Comorbidity in Longitudinal Studies: Development and Validation. *Journal of Chronic Diseases 40*(5), 373–383.

Clemens, J. and J. D. Gottlieb (2017). In the Shadow of a Giant: Medicare's Influence on Private Physician Payments. *Journal of Political Economy 125*(1), 1–39.

CMS (2014). National Health Expenditures 2014 Highlights. `www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/nationalhealthexpenddata/downloads/highlights.pdf` [Accessed: September 2016].

Cooper, Z., S. V. Craig, M. Gaynor, and J. Van Reenen (2018). The Price ain't Right? Hospital Prices and Health Spending on the Privately Insured. *Quarterly Journal of Economics*.

David, G. and M. D. Neuman (2011). Physician Division of Labor and Patient Selection for Outpatient Procedures. *Journal of Health Economics 30*(2), 381–391.

De Chaisemartin, C. and X. D'HaultfŒuille (2017). Fuzzy differences-in-differences. *The Review of Economic Studies 85*(2), 999–1028.

Duflo, E. (2001). Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment. *American Economic Review 91*(4), 795–813.

Ellis, R. P., B. Martins, and W. Zhu (2017). Health Care Demand Elasticities by Type of Service. *Journal of Health Economics 55*, 232–243.

Gaynor, M. and R. J. Town (2011). Competition in health care markets. *National Bureau of Economic Research*.

Gruber, J. and R. McKnight (2016). Controlling Health Care Costs through Limited Network Insurance Plans: Evidence from Massachusetts State Employees. *American Economic Journal: Economic Policy 8*(2), 219–50.

Hicks, J. R. (1939). The foundations of welfare economics. *The Economic Journal 49*(196), 696–712.

Ho, K. and R. S. Lee (2013). Insurer Competition in Health Care Markets. *National Bureau for Economic Research.*

Hudson, S., P. Hull, and C. Liebersohn (2015). Interpreting instrumented difference-in-differences. Technical report, Working Paper (available upon request).

Imbens, G. W. and J. D. Angrist (1994). Identification and estimation of local average treatment effects. *Econometrica 62*(2), 467–475.

Joseph P. Newhouse and RAND Corporation (1993). *Free for All?: Lessons from the RAND Health Insurance Experiment.* Harvard University Press.

Kaldor, N. (1939). Welfare propositions of economics and interpersonal comparisons of utility. *The Economic Journal*, 549–552.

Munnich, E. L. and S. T. Parente (2014). Procedures Take Less Time at Ambulatory Surgery Centers, Keeping Costs Down and Ability to Meet Demand Up. *Health Affairs 33*(5), 764–769.

Munnich, E. L. and S. T. Parente (2018). Returns to Specialization: Evidence from the Outpatient Surgery Market. *Journal of Health Economics 57*, 147–167.

Olea, J. L. M. and C. Pflueger (2013). A Robust Test for Weak Instruments. *Journal of Business & Economic Statistics 31*(3), 358–369.

Papanicolas, I., L. R. Woskie, and A. K. Jha (2018, March). Health Care Spending in the United States and Other High-Income Countries. *JAMA 319*(10), 1024–1039.

Plotzke, M. R. and C. Courtemanche (2011, July). Does Procedure Profitability Impact whether an Outpatient Surgery is Performed at an Ambulatory Surgery Center or Hospital? *Health Economics 20*(7), 817–830.

Prager, E. (2018). Consumer Responsiveness to Simple Health Care Prices: Evidence From Tiered Hospital Networks. *Working Paper.*

Rex, D. K., P. S. Schoenfeld, J. Cohen, I. M. Pike, D. G. Adler, M. B. Fennerty, J. G. Lieb, W. G. Park, M. K. Rizk, M. S. Sawhney, et al. (2015). Quality Indicators for Colonoscopy. *The American Journal of Gastroenterology 110*(1), 72.

Robinson, J. C., T. Brown, and C. Whaley (2015). Reference-Based Benefit Design Changes Consumers' Choices and Employers' Payments for Ambulatory Surgery. *Health Affairs 34*(3), 415–422.

Robinson, J. C., T. T. Brown, C. Whaley, and K. J. Bozic (2015). Consumer Choice Between Hospital-Based and Freestanding Facilities for Arthroscopy. *J Bone Joint Surg Am 97*(18), 1473–1481.

Robinson, J. C., T. T. Brown, C. Whaley, and E. Finlayson (2015). Association of Reference Payment for Colonoscopy with Consumer Choices, Insurer Spending, and Procedural Complications. *JAMA Internal Medicine 175*(11), 1783–1789.

Robinson, J. C., C. Whaley, and T. T. Brown (2016). Association of Reference Pricing for Diagnostic Laboratory Testing with Changes in Patient Choices, Prices, and Total Spending for Diagnostic Tests. *JAMA Internal Medicine 176*(9), 1353–1359.

Sinaiko, A. D. and M. B. Rosenthal (2014, August). The Impact of Tiered Physician Networks on Patient Choices. *Health Services Research 49*(4), 1348–1363. 00007.

The Dartmouth Atlas of Health Care (2016). `http://www.dartmouthatlas.org/` [Accessed: September 2016].

The Henry J. Kaiser Family Foundation (2015). Preventative Services Covered by Private Health Plans Under the Affordable Care Act. `http://kff.org/health-reform/fact-sheet/preventive-services-covered-by-private-health-plans/` [Accessed: September 2016].

US Preventative Task Force (2016). `www.uspreventiveservicestaskforce.org/Page/Document/UpdateSummaryFinal/colorectal-cancer-screening` [Accessed: September 2016].

Whaley, C. and T. T. Brown (2018). Firm Responses to Targeted Consumer Incentives: Evidence from Reference Pricing for Surgical Services. *Journal of Health Economics 61*, 111–133.

Whaley, C. M., C. Guo, and T. T. Brown (2017). The Moral Hazard Effects of Consumer Responses to Targeted Cost-Sharing. *Journal of Health Economics 56*, 201–221.

Table 1: Descriptive Statistics

	Pre-Period (2011)		Post-Period (2012-2013)	
	Control	CalPERS	Control	CalPERS
Share Male	0.51	0.47	0.50	0.46
Mean Age	55.11	56.28	55.20	56.39
Share 40-49 Yrs	0.06	0.04	0.05	0.04
Share 50-59 Yrs	0.67	0.63	0.67	0.60
Share 60-64 Yrs	0.25	0.32	0.26	0.35
Share w/Comorbidity Index= 0	0.95	0.95	0.96	0.94
Share w/Intervention	0.49	0.48	0.52	0.51
Share using ASC	0.72	0.67	0.72	0.78
Share patient from Northern Cal	0.41	0.56	0.37	0.55
Share patient from Southern Cal	0.50	0.39	0.55	0.40
Share patient from Central Cal	0.09	0.05	0.08	0.05
Mean Procedure Price ($)	1794.43	1813.94	2061.10	1820.44
Median Procedure Price ($)	996.60	1129.00	1128.00	979.50
Mean Procedure Price ($) - ASC	1479.91	1335.68	1811.39	1515.24
Mean Procedure Price ($) - Hospital	2603.14	2803.03	2714.66	2934.07
Mean Patient Cost-Sharing ($)	643.79	317.42	734.91	516.35
Median Patient Cost-Sharing ($)	0.00	0.00	0.00	0.00
Mean Patient Cost-Sharing ($) - ASC	721.82	412.28	897.25	472.46
Mean Patient Cost-Sharing ($) - Hospital	443.14	121.25	310.00	676.50
N	22,020	3,651	53,492	7,243

Note: Data consists of screening colonoscopy claims data for CalPERS and the comparison group, pooled from 2011-2013. Means are presented. Note: Northern Cal = HRRs in Alameda, Chico, Contra Costa, Modesto, Napa, Redding, Sacramento, Salinas, San Francisco, San Jose, San Mateo, Santa Rosa, Santa Cruz, and Stockton. Southern Cal = HRRs in Los Angeles, Orange County, Palm Springs, San Bernardino, San Diego, San Luis Obispo, Santa Barbara, and Ventura. Central Cal = HRRs in Bakersfield and Fresno.

Table 2: First Stage Estimates

ASC	No Cov. (1)	W/Cov. (2)
$Post \times Treat$	0.107	0.098
	(0.020)	(0.020)
$F - Statistic$	27.975	24.692
N	86,406	86,406

1: Data consists of screening colonoscopy claims data from 2011-2013. Column 1 excludes covariates, while column 2 includes covariates: $\mathbb{1}$(Male),$\mathbb{1}$(Age Category), Comorbidity Index, Patient Health Care Market Dummy.

2: Cluster s.e. at health care facility level – 541 clusters

Table 3: Regression Estimates

Outcome Variable	Reduced Form No Cov. (1)	Reduced Form W/Cov. (2)	OLS No Cov. (3)	OLS W/Cov. (4)	IV No Cov. (5)	IV W/Cov. (6)
Price	-260	-224	-1025	-984	-2425	-2300
	(101)	(89.8)	(253)	(267)	(965)	(954)
$\mathbb{1}(Price \leq \$1.5K)$	.0691	.0657	.608	.602	.644	.673
	(.0196)	(.0172)	(.0597)	(.0499)	(.138)	(.135)
$\mathbb{1}(Price \leq \$2K)$	.0497	.046	.451	.436	.464	.472
	(.0294)	(.0282)	(.0629)	(.0547)	(.255)	(.278)
$\mathbb{1}(Price \leq \$2.5K)$	.0358	.0291	.303	.306	.334	.298
	(.0154)	(.0135)	(.0563)	(.0533)	(.138)	(.135)
$\mathbb{1}(Price \leq \$5K)$	.0127	.0104	.0144	.0173	.118	.106
	(.00597)	(.00544)	(.0199)	(.0207)	(.0615)	(.0612)
$\mathbb{1}(Price \leq \$10K)$	.00605	.00515	-.0231	-.025	.0564	.0527
	(.004)	(.0036)	(.00889)	(.0104)	(.0398)	(.0393)
N	86,406	86,406	86,406	86,406	86,406	86,406

1: Data consists of screening colonoscopy claims data pooled from 2011-2013. Columns 1-2 give the effect of $Post \times Treat$ on outcomes (per-procedure price), while Columns 3-6 give the effect of ASC on outcomes.
2: Columns 1, 3, and 5 exclude covariates, while columns 2, 4, and 6 include covariates: $\mathbb{1}$(Male),$\mathbb{1}$(Age Category), Comorbidity Index, Patient Health Care Market Dummy.
3: Cluster s.e. at health care facility level– 542 clusters

Table 4: Coefficient Estimates of γ_2

Outcome	ASC Sample (1)	Hospital Sample (2)
$\ln$ Negotiated Amount	-0.262	-0.048
	(0.120)	(0.110)
Negotiated Amount ($)	-339.662	-185.996
	(215.777)	(266.911)
N	58,637	23,257

1: Each cell represents the coefficient estimate from γ_2 from Equation 4.
2: Data is limited to health care facilities observed in all three years of 2011, 2012, and 2013.

3: Cluster s.e. at health care facility level– 209 ASC clusters, 167 Hospital clusters

Table 5: Regression Estimates using Exact Matching Price Definition

	Reduced Form		OLS		IV	
Outcome Variable	No Cov.	W/Cov.	No Cov.	W/Cov.	No Cov.	W/Cov.
	(1)	(2)	(3)	(4)	(5)	(6)
Price	-209	-169	-1162	-1107	-1832	-1670
	(77.9)	(69.8)	(229)	(239)	(641)	(648)
$\mathbb{1}(Price \leq \$1.5K)$	.0689	.0603	.605	.592	.604	.595
	(.02)	(.0177)	(.068)	(.0583)	(.126)	(.123)
$\mathbb{1}(Price \leq \$2K)$	.0617	.0534	.419	.404	.541	.527
	(.0187)	(.0166)	(.0718)	(.0608)	(.137)	(.141)
$\mathbb{1}(Price \leq \$2.5K)$	.0441	.0357	.287	.285	.387	.352
	(.0164)	(.0141)	(.0659)	(.0615)	(.136)	(.131)
$\mathbb{1}(Price \leq \$5K)$	.0038	.0025	-.00535	-.0101	.0333	.0247
	(.00838)	(.00805)	(.0167)	(.018)	(.0734)	(.0791)
$\mathbb{1}(Price \leq \$10K)$	.00266	.00201	-.0117	-.0146	.0233	.0198
	(.00325)	(.00286)	(.00831)	(.00993)	(.029)	(.0285)
N	80,550	80,550	80,550	80,550	80,550	80,550

1: Data consists of screening colonoscopy claims data pooled from 2011-2013. Columns 1-2 give the effect of $Post \times Treat$ on outcomes (per-procedure price), while Columns 3-6 give the effect of ASC on outcomes.

2: Columns 1, 3, and 5 exclude covariates, while columns 2, 4, and 6 include covariates: $\mathbb{1}$(Male),$\mathbb{1}$(Age Category), Comorbidity Index, Patient Health Care Market Dummy.

3: Cluster s.e. at health care facility level– 373 clusters

Table 6: Regression Estimates using Regression Adjusted Price Definition

Outcome	Reduced Form No Cov. (1)	Reduced Form W/ Cov. (2)	OLS No Cov. (3)	OLS W/ Cov. (4)	IV No Cov. (5)	IV W/ Cov. (6)
Price	-204	-166	-1146	-1082	-1803	-1630
	(79.4)	(70.7)	(226)	(238)	(662)	(658)
$\mathbb{1}(Price \leq$ \$1.5K)	.0695	.0619	.611	.6	.612	.607
	(.0205)	(.0183)	(.0646)	(.0552)	(.13)	(.128)
$\mathbb{1}(Price \leq$ \$2K)	.0484	.0405	.449	.433	.427	.39
	(.0248)	(.023)	(.0644)	(.0576)	(.195)	(.21)
$\mathbb{1}(Price \leq$ \$2.5K)	.0475	.0396	.279	.28	.419	.389
	(.0166)	(.0148)	(.0624)	(.0587)	(.14)	(.14)
$\mathbb{1}(Price \leq$ \$5K)	-.00412	-.00616	-.000522	-.00568	-.0363	-.0605
	(.00773)	(.00777)	(.0146)	(.0161)	(.067)	(.0746)
$\mathbb{1}(Price \leq$ \$10K)	.00227	.00156	-.0121	-.0151	.02	.0153
	(.00329)	(.00288)	(.00825)	(.00995)	(.0294)	(.0285)
N	81,837	81,837	81,837	81,837	81,837	81,837

1: Data consists of screening colonoscopy claims data pooled from 2011-2013, omitting those with a predicted price lower than \$300 (the minimum observed expense from 2011). Columns 1-2 give the effect of $Post \times Treat$ on outcomes (per-procedure price), while Columns 3-6 give the effect of ASC on outcomes.
2: Columns 1, 3, and 5 exclude covariates, while columns 2, 4, and 6 include covariates: $\mathbb{1}$(Male),$\mathbb{1}$(Age Category), Comorbidity Index, Patient Health Care Market Dummy.
3: Cluster s.e. at health care facility level– 376 clusters

Table 7: Cost-Sharing Response on Price

First-Stage Results - Outcome is $Pay\,Diff$ (\$'s)	
$Post$	-107.833
	(20.197)
$Treat$	-66.141
	(31.751)
$Post \times Treat$	926.364
	(166.559)
N	84,277
Second-Stage Results - Outcome is $Price$ (\$'s)	
$Post$	219.600
	(85.949)
$Treat$	24.033
	(57.264)
$\widehat{Pay\,Diff}$	-0.257
	(0.108)
N	84,277

1: The first-stage results give the coefficient estimates from estimating Equation 5. The second-stage results give the coefficient estimates from estimating Equation 6.
2: Data consists of screening colonoscopy claims data pooled from 2011-2013. All estimates include the covariates: $\mathbb{1}$(Male), $\mathbb{1}$(Age Category), Comorbidity Index, Patient Health Care Market Dummy.
4: Cluster s.e. at health care facility level– 535 clusters

Table 8: Health Complication/Intervention Outcomes

	IV		Pre-Period Occurrence
Health Outcome	No Cov.	W/Cov.	
	(1)	(2)	(3)
Any Complications (1-30d)	-0.025	-0.031	0.73 percent
	(0.021)	(0.023)	
Serious Complications (1-30d)	-0.005	-0.006	0.17 percent
	(0.010)	(0.011)	
Non-Serious Complications (1-30d)	-0.006	-0.008	0.34 percent
	(0.012)	(0.013)	
$\mathbb{1}$(Intervention)	0.025	0.036	48.53 percent
	(0.113)	(0.118)	
N	86,406	86,406	

1: Columns 1-2 give the effect of ASC on outcomes using a 2SLS regression with $Post \times Treat$ as the instrument.
2: The pre-period mean for each binary outcome variable is presented in column 3.
3: Data consists of screening colonoscopy claims data pooled from 2011-2013. Column 1 excludes covariates, while column 2 includes covariates: $\mathbb{1}$(Male), $\mathbb{1}$(Age Category), Comorbidity Index, Patient Health Care Market Dummy.
4: Cluster s.e. at health care facility level– 542 clusters

Table 9: Characteristics of Compliers, Always Takers, and Never Takers

	Compliers	Always Takers	Never Takers
Share Male	0.36	0.50	0.50
	(0.08)	(0.01)	(0.02)
Average Age	55.69	55.32	56.41
	(1.04)	(0.07)	(0.21)
Share 40-49 Yrs	0.07	0.05	0.03
	(0.04)	(0.00)	(0.01)
Share 50-59 Yrs	0.49	0.67	0.61
	(0.10)	(0.00)	(0.02)
Share 60-64 Yrs	0.42	0.26	0.35
	(0.10)	(0.00)	(0.02)
Share w/Comorbidity	0.92	0.96	0.92
Index = 0	(0.03)	(0.00)	(0.01)
Share w/Intervention	0.59	0.51	0.48
	(0.10)	(0.01)	(0.02)
Share Visiting Facility	0.90	0.82	0.64
in their own HRR	(0.09)	(0.03)	(0.06)
Share Northern Cal	0.64	0.39	0.49
	(0.14)	(0.05)	(0.07)
Share Southern Cal	0.17	0.53	0.46
	(0.15)	(0.05)	(0.07)
Share Central Cal	0.19	0.08	0.05
	(0.10)	(0.04)	(0.02)

Covariate means for compliers, in column 1, are computed using the technique described in Abadie (2002), using pooled 2011-2013 data. Clustered standard errors are presented in parentheses. Note: Covariate means for Always Takers and Never Takers are are estimated in the data using the years 2011-2013. Never Taker means exclude the reference price exempt population. Clustered standard errors are presented in parentheses. Data consists of screening colonoscopy claims data from members of CalPERS and the comparison group who use ASCs. Means and standard deviations, in parentheses, are presented. Note: Northern Cal = HRRs in Alameda, Chico, Contra Costa, Modesto, Napa, Redding, Sacramento, Salinas, San Francisco, San Jose, San Mateo, Santa Rosa, Santa Cruz, and Stockton. Southern Cal = HRRs in Los Angeles, Orange County, Palm Springs, San Bernardino, San Diego, San Luis Obispo, Santa Barbara, and Ventura. Central Cal = HRRs in Bakersfield and Fresno.

Figure 1: How Reference Pricing Works

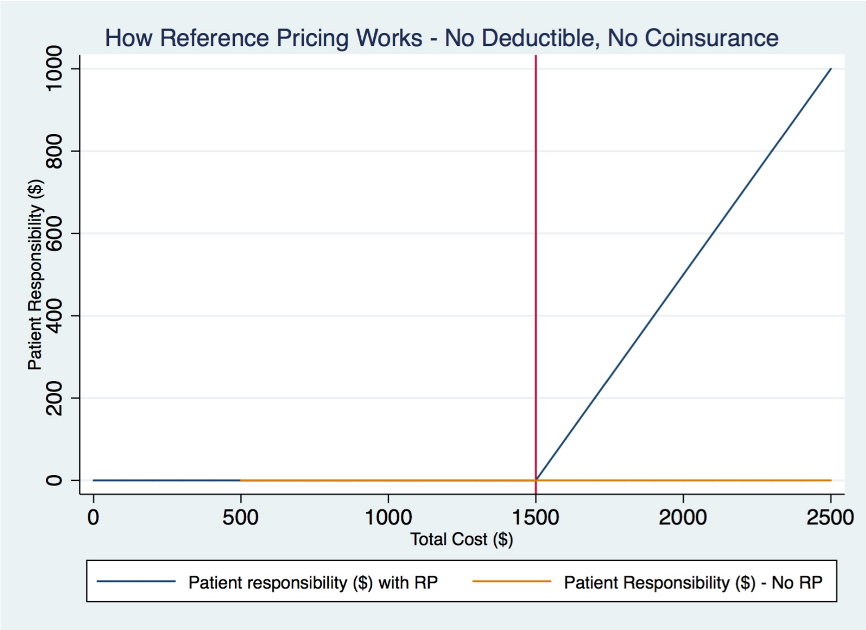
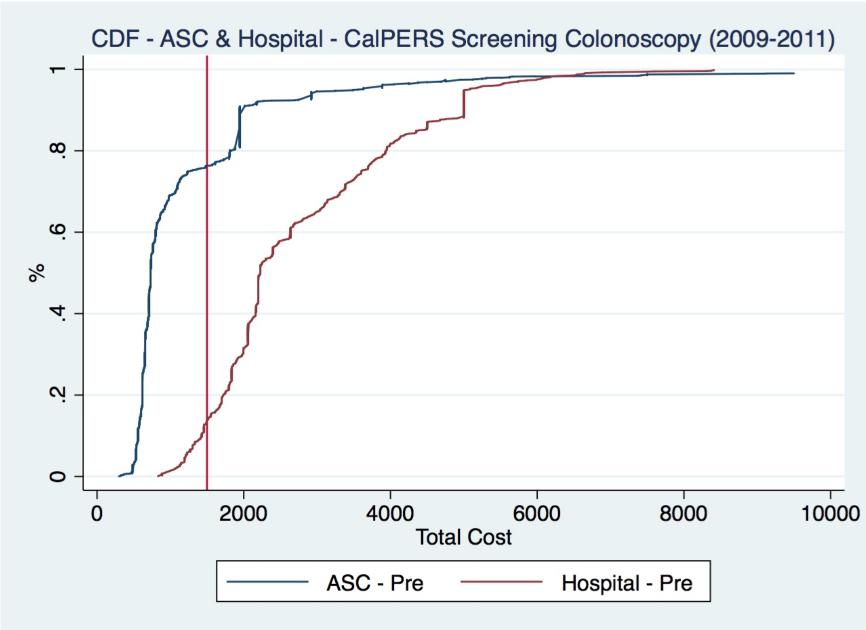
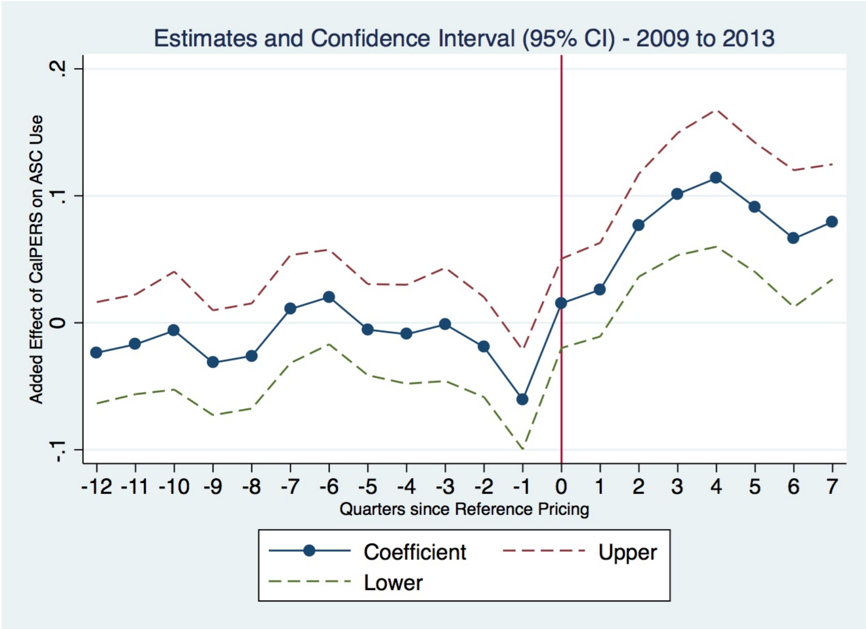


Figure 2: ASC vs. Hospital Per-Procedure Price Distribution (CDF)



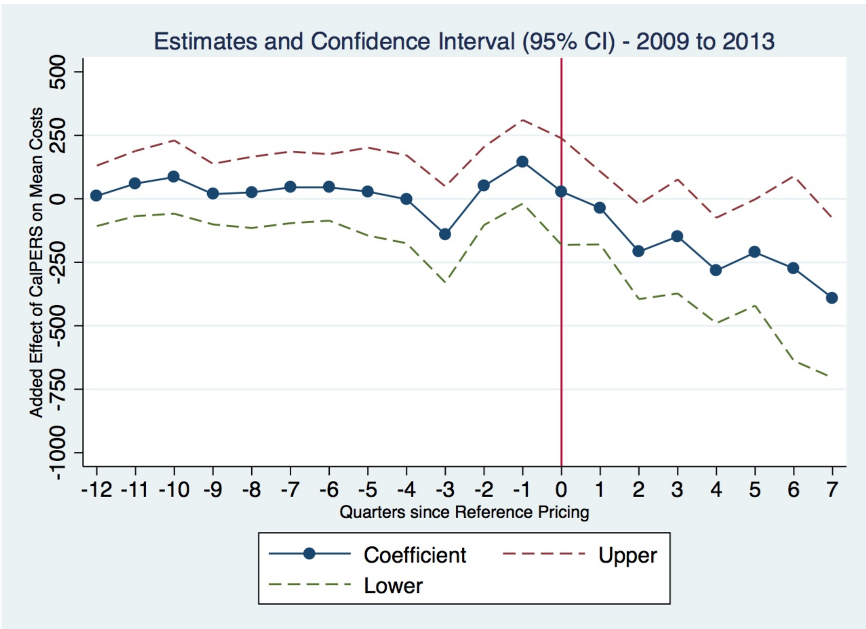
Note: Data consists of screening colonoscopy claims data for CalPERS members for the year 2011.

Figure 3: Event Study - Added Effect of CalPERS on ASC Use



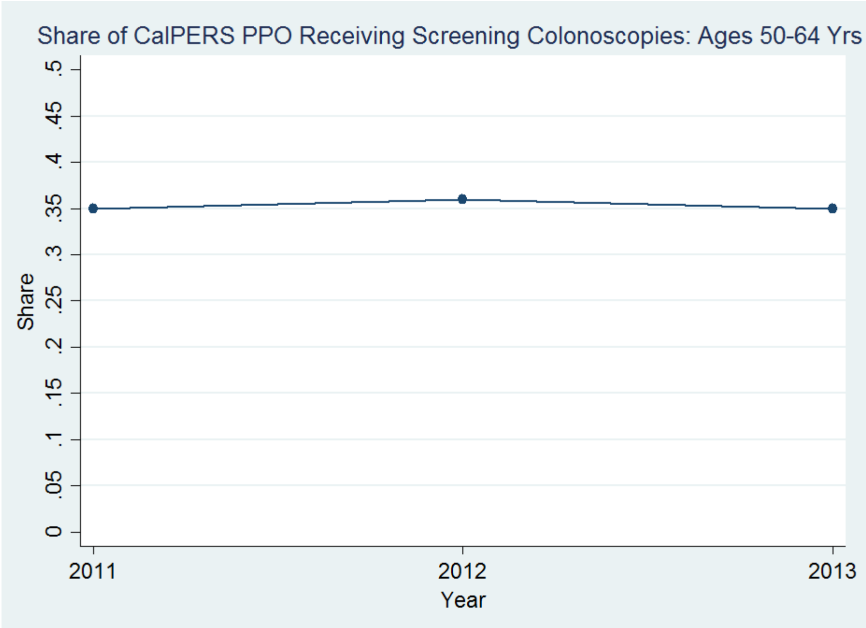
Note: Figure plots the coefficient estimates of γ^r from the event study presented in the methodology section, where the outcome variable is ASC.

Figure 4: Event Study - Added Effect of CalPERS on Mean Procedure Price



Note: Figure plots the coefficient estimates of γ^r from the event study presented in the methodology section, where the outcome variable is $price$.

Figure 5: Estimated CalPERS Screening Colonoscopy Rates by Year



Note: Data used to construct the utilization estimates consists of 1) Active PPO enrollment numbers provided by CalPERS (members ages 50-64 years) and 2) Screening colonoscopy claims data for CalPERS members in the corresponding years.

Appendix

Table 10: Data Description - ASCs

Outcome Variable	ASC-Pre Control	ASC-Pre CalPERS	ASC-Post Control	ASC-Post CalPERS
Share Male	0.51	0.47	0.50	0.45
Share Charlson Index=0	0.96	0.95	0.96	0.95
Avg. Age	55.08	56.24	55.25	56.31
Share 40-49 Yrs	0.06	0.04	0.05	0.04
Share 50-59 Yrs	0.68	0.62	0.68	0.61
Share 60-64 Yrs	0.25	0.33	0.26	0.35
Share w/Intervention	0.49	0.47	0.53	0.52
Share Northern Cal	0.40	0.54	0.36	0.53
Share Southern Cal	0.52	0.41	0.56	0.41
Share Central Cal	0.09	0.05	0.08	0.06
Mean Procedure Price ($)	1479.91	1335.68	1811.39	1515.24
Mean Patient Cost-Sharing ($)	721.82	412.28	897.25	472.46
Median Procedure Price ($)	735.00	730.00	783.00	783.00
Median Patient Cost-Sharing ($)	0.00	0.00	0.00	0.00
N	15,854	2,461	38,704	5,685

Note: Data consists of screening colonoscopy claims data from members of CalPERS and the comparison group who use ASCs. Means and standard deviations, in parentheses, are presented. Note: Northern Cal = HRRs in Alameda, Chico, Contra Costa, Modesto, Napa, Redding, Sacramento, Salinas, San Francisco, San Jose, San Mateo, Santa Rosa, Santa Cruz, and Stockton. Southern Cal = HRRs in Los Angeles, Orange County, Palm Springs, San Bernardino, San Diego, San Luis Obispo, Santa Barbara, and Ventura. Central Cal = HRRs in Bakersfield and Fresno.

Table 11: Data Description - Hospitals

Outcome Variable	Hospital-Pre Control	Hospital-Pre CalPERS	Hospital-Post Control	Hospital-Post CalPERS
Share Male	0.52	0.46	0.51	0.49
Share Charlson Index=0	.0.94	0.95	0.95	0.92
Avg. Age	55.20	56.36	55.08	56.66
Share 40-49 Yrs	0.06	0.04	0.06	0.04
Share 50-59 Yrs	0.66	0.64	0.66	0.58
Share 60-64 Yrs	0.27	0.32	0.26	0.37
Share w/Intervention	0.48	0.49	0.51	0.48
Share Northern Cal	0.43	0.60	0.39	0.61
Share Southern Cal	0.46	0.36	0.51	0.35
Share Central Cal	0.11	0.05	0.09	0.04
Mean Procedure Price ($)	2603.14	2803.03	2714.66	2934.07
Mean Patient Cost-Sharing ($)	443.14	121.25	310.00	676.50
Median Procedure Price ($)	2199.00	2225.00	2344.00	2442.23
Median Patient Cost-Sharing ($)	0.00	0.00	0.00	363.13
N	6,166	1,190	14,788	1,558

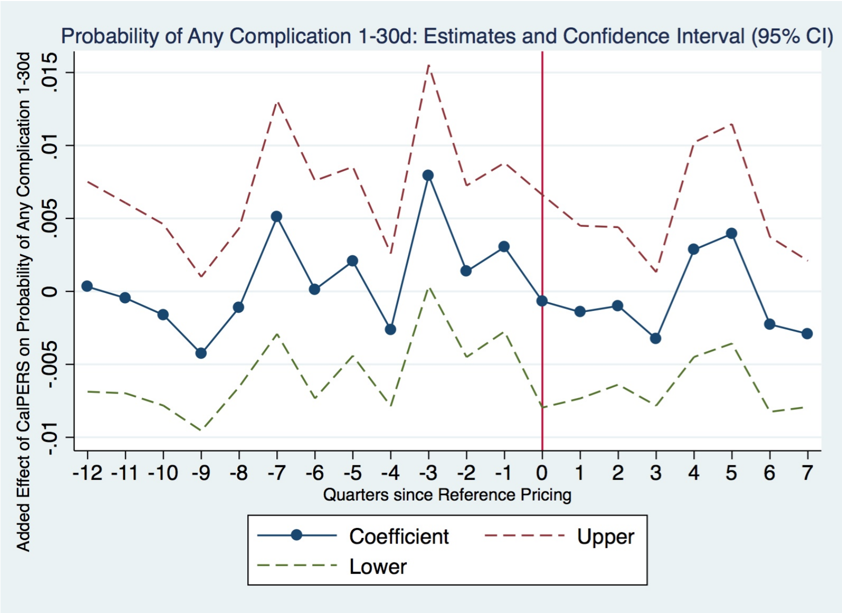
Note: Data consists of screening colonoscopy claims data from members of CalPERS and the comparison group who use hospitals. Means and standard deviations, in parentheses, are presented. Note: Northern Cal = HRRs in Alameda, Chico, Contra Costa, Modesto, Napa, Redding, Sacramento, Salinas, San Francisco, San Jose, San Mateo, Santa Rosa, Santa Cruz, and Stockton. Southern Cal = HRRs in Los Angeles, Orange County, Palm Springs, San Bernardino, San Diego, San Luis Obispo, Santa Barbara, and Ventura. Central Cal = HRRs in Bakersfield and Fresno.

Table 12: Health Complication/Intervention Outcomes

Health Outcome	Reduced Form Estimates No Cov. (1)	Reduced Form Estimates W/Cov. (2)	Pre-Period Occurrence (3)
Any Complications (1-30d)	-0.003	-0.003	0.73 percent
	(0.002)	(0.002)	
Serious Complications (1-30d)	-0.001	-0.001	0.17 percent
	(0.001)	(0.001)	
Non-Serious Complications (1-30d)	-0.001	-0.001	0.34 percent
	(0.001)	(0.001)	
$\mathbb{1}$(Intervention)	0.003	0.003	48.53 percent
	(0.012)	(0.011)	
N	86,406	86,406	

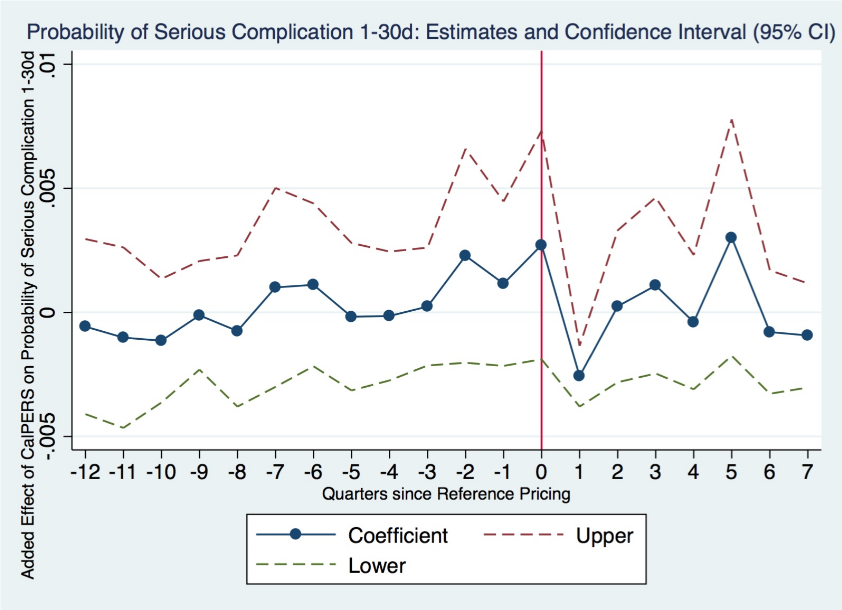
1: Columns 1-2 give estimates of β_3 from Equation 2 where the outcome is the health outcome specified in the row.
2: The pre-period mean for each binary outcome variable is presented in column 3.
3: Data consists of screening colonoscopy claims data pooled from 2011-2013. Column 1 excludes covariates, while column 2 includes covariates: $\mathbb{1}$(Male), $\mathbb{1}$(Age Category), Comorbidity Index, Patient Health Care Market Dummy.
4: Cluster s.e. at health care facility level– 542 clusters

Figure 6: Event Study - Added Effect of CalPERS on Probability of Any Complication 1-30d



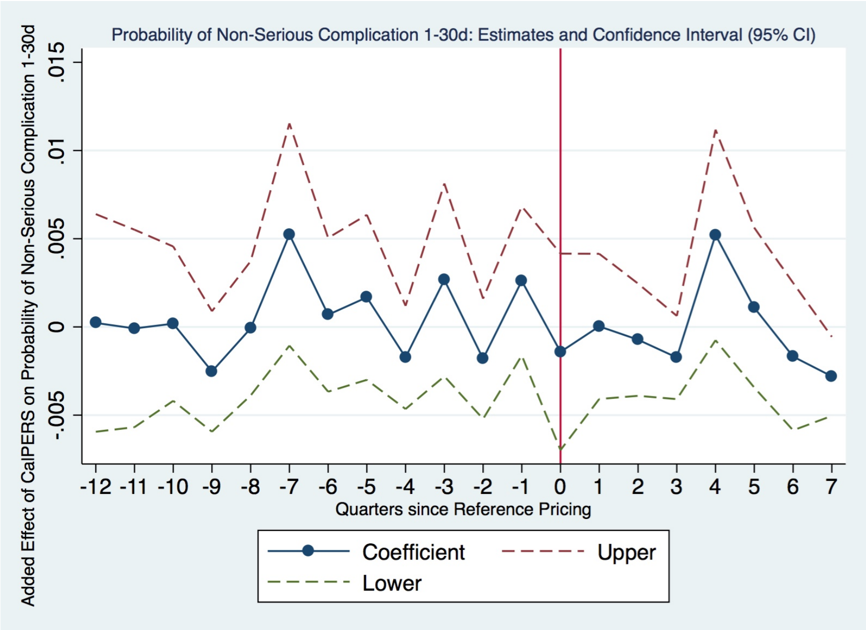
Note: Figure plots the coefficient estimates of γ^r from the event study presented in the methodology section, where the outcome variable is $Any\,Complication(1-30d)$.

Figure 7: Event Study - Added Effect of CalPERS on Probability of Serious Complication 1-30d



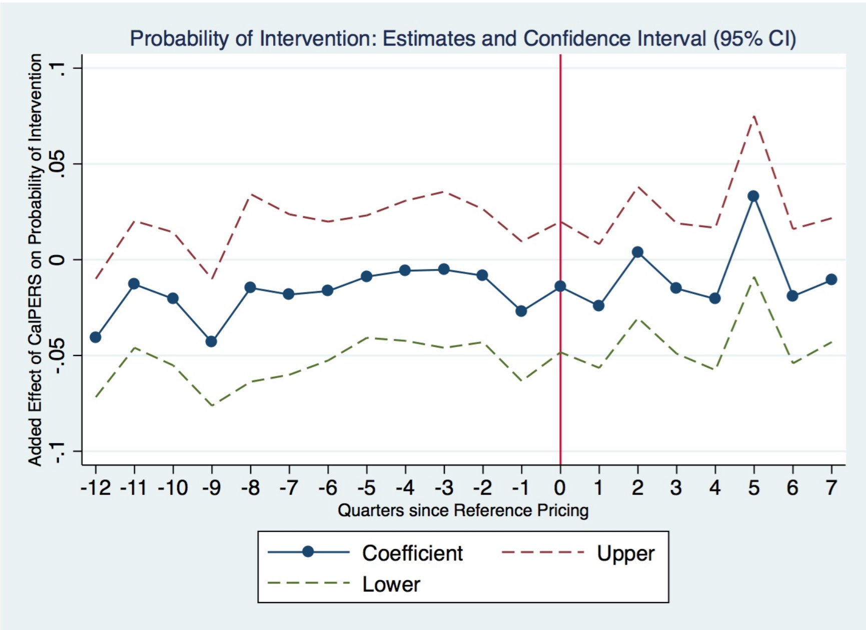
Note: Figure plots the coefficient estimates of γ^r from the event study presented in the methodology section, where the outcome variable is $Serious\,Complication(1-30d)$.

Figure 8: Event Study - Added Effect of CalPERS on Probability of Non-Serious Complication 1-30d



Note: Figure plots the coefficient estimates of γ^r from the event study presented in the methodology section, where the outcome variable is $Non - Serious\, Complication(1 - 30d)$.

Figure 9: Event Study - Added Effect of CalPERS on Probability of Intervention



Note: Figure plots the coefficient estimates of γ^r from the event study presented in the methodology section, where the outcome variable is $Intervention$.

Cost-Sharing Analysis

Let $PatPay_i$ represent the cost-sharing amount of the patient after visiting an ASC or Hospital. Specifically, let $PatPayASC_i$ be the patient's cost-sharing amount when visiting an ASC and $PatPayHosp_i$ be the patient's cost-sharing amount when visiting a hospital. Also, let $Price_i$ represent the total price for the medical service patient i receives. The different cases, presented below, are used to determine what each patient would have paid had they gone to an ASC or hospital. Since patients are only observed going to either an ASC or hospital, and not both, the patient's cost-sharing amount for the non-observed scenario must be imputed. We impute a patient's cost-sharing amount for the non-observed setting by computing the average patient cost-sharing amount in the HRR of the facility/provider that the patient visits - this average represents the imputed value. For example, if a pre-period CalPERS patient is observed at an ASC in HRR "X," the patient's cost-sharing amount at the ASC is directly observed in the data (i.e. $PatPayASC_i$). However, we compute the cost-sharing amount the patient would have faced at a hospital by averaging over the cost-sharing amounts of patients j who visited hospitals in the same period (pre or post) in the HRR (i.e. $\frac{1}{N_{HRR,Hosp}} \times \sum_{HRR\,Provider_{j,\,j \neq i}} PatPayHosp_j$). After reference pricing is introduced (Case 2), the CalPERS patient cost-sharing amount estimates are adjusted for this new institutional feature. Specifically, the imputed hospital payment for a CalPERS patient who visits an ASC averages over the cost-sharing amount of individuals who visit hospitals in the same HRR in 2011 (before reference pricing is introduced). To capture the effects of the policy, for hospital goers who experience an expense greater than or equal to \$1500, we use the cost-sharing amount less the \$1500 reference price threshold. For hospital goers who experience an expense less than \$1500, we use the actual patient cost-sharing amount.

Case 1 (a. Pre-Period CalPERS, b. Pre-Period Control and c. Post-Period Control):

$$PatPayASC_i = \begin{cases} PatPayASC_i, \text{ if } i \text{ visits ASC} \\ \dfrac{1}{N_{HRR,ASC}} \times \displaystyle\sum_{HRR\,Provider_{j,\,j\neq i}} PatPayASC_j, \text{ if } i \text{ visits hospital} \end{cases}$$

$$PatPayHosp_i = \begin{cases} PatPayHosp_i, \text{ if } i \text{ visits Hospital} \\ \dfrac{1}{N_{HRR,Hosp}} \times \displaystyle\sum_{HRR\,Provider_{j,\,j\neq i}} PatPayHosp_j, \text{ if } i \text{ visits ASC} \end{cases}$$

Case 2: Post-Period CalPERS

$$PatPayASC_i = \begin{cases} PatPayASC_i, \text{ if } i \text{ visits ASC} \\ \dfrac{1}{N_{HRR,ASC}} \times \displaystyle\sum_{HRR\,Provider_{j,\,j\neq i}} PatPayASC_j, \text{ if } i \text{ visits hospital} \end{cases}$$

$$PatPayHosp_i = \begin{cases} PatPayHosp_i, \text{ if } i \text{ visits Hospital} \\ \dfrac{1}{N_{HRR,Hosp}} \times \displaystyle\sum_{HRR\,Provider_{j,\,j\neq i}} Hosp_j = \begin{cases} Price_j - \$1500, \text{if}\, Price_j \geq \$1500 \\ PatPayHosp_j, \text{if}\, Price_j < \$1500 \end{cases} \end{cases}$$