Understanding the Distributional Impacts of Health Insurance Reform: Evidence from a Consumer Cost-Sharing Program

Marion Aouad[1*], Timothy T. Brown[2†], and Christopher M. Whaley[3‡]

January 3, 2021

Abstract

We examine the heterogeneous effects of reference pricing, a health insurance reform introduced by the California Public Employees' Retirement System (CalPERS), on the distribution of spending by patients and insurers. Using medical claims data for CalPERS and a comparison group not subject to reference pricing, we use the changes-in-changes approach to estimate the quantile treatment effects of the program across different medical procedures. We find that the quantile treatment effects vary across the patient spending distributions, with a range of positive and negative estimates of the QTE, depending on the medical procedure considered. However, across all procedures, the insurer's spending distributions tend to shift left, with the largest reductions occurring in the right-tail of the spending distributions. These effects are not captured by mean estimates but have important policy implications.

[1*] Corresponding Author: marion.aouad@uci.edu; University of California Irvine, Department of Economics 3151 Social Sciences Plaza, Irvine CA, 92697

[2†] University of California Berkeley, School of Public Health.

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

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, NCI grant R21-CA219229, and the NIHCM Foundation. 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.

1. Introduction

Between 2000 and 2018, spending on health care in the United States increased from 13.3% of GDP to 17.7%, and is expected to outpace GDP growth by 1.1 percentage points per year over the next decade (CMS, 2019). Adding to this concern, some estimates suggest that in the United States, approximately one-third of health care spending is wasteful, with market pricing failures accounting for approximately 14% ($131 billion) of wasteful spending (Berwick and Hackbarth, 2012). Furthermore, health care prices are often characterized by a wide degree of price variation, even within narrow categories of common health care services (Newman et al. 2016). The variation in prices has led to significant interest among policymakers in finding effective ways to reduce wasteful spending in order to improve the efficiency of health care markets.

This paper examines how health insurance reforms can affect the variability of medical spending, for insurers and patients. We do so by estimating the distributional effects of an innovative health insurance design on insurer and patient spending using the changes-in-changes (CIC) method (Athey and Imbens, 2006; Melly and Santangelo, 2015). To our knowledge, this is the first analysis to apply the CIC methods to examine the effects of a health insurance reform.

Specifically, we analyze a reference pricing reform implemented by one of the largest health care purchasers in the United States, the California Public Employees' Retirement System (CalPERS). Reference pricing encourages patients to use lower-priced providers by changing the relative prices patients face when deciding on the setting in which to receive medical care. Reference pricing has been widely applied to pharmaceutical products within Europe, including Sweden, Spain, Norway, Italy, Hungary, Belgium, and Germany, and outside of Europe in Canada and New Zealand (Galizzi et al., 2011).

In the setting we analyze, the introduction of reference pricing changes the relative prices faced by CalPERS members for several medical procedures. Specifically, CalPERS sets a maximum reimbursable amount (i.e., the reference price) for medical services received in higher-price settings, requiring patients who choose higher-price settings to cover the difference between the price and the maximum reimbursement amount. This design explicitly addresses the observed price dispersion across medical facilities and incentivizes the use of lower-price providers by exposing patients to the full marginal costs of choosing higher-price providers.

Using detailed medical claims data for both CalPERS and a comparison group covered by the same health insurer, Anthem Blue Cross, we focus on the effect of the reference pricing policy across four categories of medical procedures: total joint replacement, colonoscopy, arthroscopy, and cataract surgery. These procedures are chosen by the insurer because they are very common standardized procedures that are elective, such that shopping can easily occur.

Our results indicate that there are heterogeneous effects of reference pricing across the distributions of spending within and across the four categories of medical procedures. We find that the largest reductions in insurer spending occur at the higher quantiles. That is, the right-tail shifts in by more than other parts of the distribution. However, consistent with the incentive structure of reference pricing, patient out-of-pocket spending generally remains the same or *increases* across the quantiles.

This work adds to the broad literature analyzing the effects of health insurance reform on patients' health behaviors and choices (e.g. Gruber and McKnight (2016), Eisenberg et al. (2017), Zhang et al. (2018)). Additionally, this work expands upon prior research on the effects of the CalPERS reference pricing program. These early reference pricing studies find that, on average, patients respond to the program's incentives by switching from higher-price settings to lower-price settings (Robinson and Brown, 2013; Robinson et al., 2015a;

Robinson et al., 2015b). Similarly, these basic results have been extended to examine the program's impact on moral hazard (Whaley et al., 2017), provider responses to the reference pricing program (Whaley and Brown, 2018), and the local average treatment effect (LATE) of compliance with the program (Aouad et al., 2019).

However, previous examinations of health insurance reforms have typically only examined the mean effects, often by estimating difference-in-differences models. This study expands upon these earlier studies by moving beyond mean estimates and focusing on the distributional effects of reference pricing. This setting is ideal for this type of examination because the reference pricing reform was explicitly designed to address the pre-policy variability in health care prices.[4] Thus, the pre-existing price dispersion (within and across lower/higher-priced health care facilities) along with the program's incentives provide for a setting in which heterogeneous impacts along the spending distributions are likely to occur.

Additionally, examining the distributional effects of the reform has important implications for health care policy. In particular, finding a larger right-tail reduction in the distribution of spending (relative to the left-tail or median) is important if right-tail spending is deemed wasteful. This would be the case if higher spending is not positively correlated with higher quality health care.[5] This finding can provide insight for policy-makers interested in designing health insurance reforms to reduce right-tail spending without sacrificing the quality of care received by patients.[6]

Furthermore, since this policy financially exposes patients to the wide degree of variation

---

[4] For example, Appendix Figure 1 provides evidence of the price dispersion across health care providers and health care procedures. Figure 1 presents the probability density functions for total spending across arthroscopy and colonoscopy procedures in the years prior to reference pricing.

[5] For a review of this subject, please see Hussey et al. (2013).

[6] In fact, numerous studies support the contention that higher spending is not positively correlated with higher quality health care in the CalPERS reference pricing context (Brodke et al., 2019; Robinson et al., 2015a; Robinson et al., 2015b).These studies also find that there is no measurable change in complication rates due to the implementation of reference pricing.

in provider prices, the results of this study can provide an alternative way of examining the effects of the reference pricing program on patients. Specifically, this analysis shows how patient out-of-pocket payments are affected at the different quantiles of the patient spending distributions. This, in turn, can be informative for understanding the *variability* in patient financial burden (as measured by an increase in patient out-of-pocket payments) resulting from the policy's enactment. For example, if higher quantiles (i.e. "the right-tail") in the patient spending distribution have a positive QTE estimate, while lower quantiles have a negative QTE or a QTE of zero, then some patients must pay more as a result of the policy change. Furthermore, such a result shows that regardless of one's assumptions about patient sorting across the spending distribution (i.e. rank invariance vs. rank similarity or another sorting scheme), there exist patients who experience a greater financial burden as a result of the policy. Moreover, our analysis details where along the patient spending distribution these increases occur and how much more costly these patient payments are.

In sum, from the insurer's standpoint, the reform reduces costs and affects the variability of overall spending by limiting insurers' exposure to higher-priced providers. However, the financial responsibility for patients tends to increase and stems from the upper end of the spending distribution, where price dispersion tends to be high. Thus, depending on policy-makers' objectives, these two outcomes should be weighed against each other. In particular, a policy maker placing higher weight on patients who are most adversely affected by the reform should focus on these distributional effects as they detail where in the distribution of spending costly increases occur.[7]

The paper proceeds as follows: Section 2 provides an overview of the institutional background, Section 3 describes the data used in this paper, and Sections 4 and 5 discuss the

[7] Since this program does generate net-savings (see Aouad et al. (2019)), the program can be Pareto improving under some redistributional scheme.

methodology and results, respectively. We conclude in Section 6.

2. Institutional Background

CalPERS offers health insurance to State of California employees and their dependents. Many California municipalities and public agencies also receive health insurance from CalPERS. Covering approximately 1.4 million lives, CalPERS is the second largest purchaser of commercial health insurance benefits in the US. CalPERS provides health insurance benefits primarily through three health insurers: Kaiser Permanente (which offers an integrated health maintenance organization (HMO) plan), Blue Shield (which offers HMO plans), and Anthem Blue Cross (which offers preferred provider organization (PPO) plans).

Reference pricing was only implemented in CalPERS PPO plans, which cover approximately 225,000 lives. The decision to implement the reference pricing program in the PPO plan was largely motivated by the observed variation in prices, coupled with rising PPO insurance premiums. Switching out of a PPO plan to avoid reference pricing was not a reasonable strategy for most CalPERS patients, since while CalPERS PPO plans provide a choice of settings, HMO plans restrict this choice. Those who switched out of a PPO plan and into an HMO plan would also be subject to an even higher out-of-pocket price if they chose a higher-price setting for the relevant procedures since such settings would virtually always be subject to out-of-network pricing in an HMO.

Reference pricing for total joint replacement was introduced in 2011, with the reference price set at $30,000. For total joint replacement, patients who visited hospitals designated by CalPERS as Value-Based Purchasing Design (VBPD) were not subject to reference pricing.[8]

[8] VBPD hospitals charged a price at or below the reference price ($30,000) *and* were deemed to have acceptable quality. For a further discussion of the program in the context of total joint replacement, see Robinson and Brown (2013).

However, patients that visited hospitals not designated as VBPD (i.e. a non-VBPD hospital) were subject to the reference pricing policy. Furthermore, non-VBPD hospitals virtually always charged fees greater than the reference price.

Thus, the out-of-pocket price paid by patients can be described as follows: let r represent the reference price, d represent the deductible, and c represent the coinsurance rate, then patients who select a hospital with a price p are subject to reference pricing, and their out-of-pocket payments will be as follows:

$$(d + c(r - d) + (p - r)) \ \ if\ p > r > d \tag{1}$$

$$(d + c(p - d)) \ \ if\ r \geq p > d \tag{2}$$

$$p\ if\ p \leq d \tag{3}$$

If visiting a non-VBPD hospital where $p > r$, patient-out-of-pocket payments are described by equation (1) whereas patient out-of-pocket payments at non-VBPD hospitals where $p \leq r$ and are described by equation (2). Equation (3) is included for completeness.

For colonoscopy, arthroscopy, and cataract surgery, reference pricing started in 2012. In this setting, patients choosing a hospital outpatient department (HOPD), which tends to be a higher-price setting, were subject to reference pricing of $1,500 for colonoscopy, $6,000 for arthroscopy, and $2,000 for cataract surgery. However, no reference price was set for patients choosing an ambulatory surgical center (ASC), which tends to be a lower-price setting.

Thus, the application of reference pricing is asymmetric in facility choice. In the general case, if patients choose an HOPD with price p, their out-of-pocket payments are represented by equations 1-3 above. In contrast, if a patient chose an ASC with price p', then their out-of-pocket payments are as follows:

$$(d + \alpha(p' - d)) \ \ if\ p' > d \tag{4}$$

$$p'\ if\ p' \leq d \tag{5}$$

Patients are further discouraged from using non-VBPD hospitals for total joint replacement and from using HOPDs for colonoscopy, arthroscopy and cataract surgery because patient payments above the reference price are neither applied towards the patient's annual health insurance deductible nor applied toward the patient's out-of-pocket maximum.

An important point is that patients who live more than 30 miles from a VPBD or ASC are exempt from the reference pricing policy. Patients receiving one of the outpatient procedures are also exempt from reference pricing if their physician deems it clinically necessary that they receive care in an HOPD due to age and/or comorbid conditions.

3. Data

To estimate the distributional effects of the program, we make use of detailed medical claims data that cover the 2008-2014 (total joint replacement) and 2009-2013 (arthroscopy, colonoscopy, cataract surgery) periods for two populations. The first population consists of CalPERS employees and dependents who are enrolled in an Anthem PPO and received one of the procedures in the four relevant medical categories. The second group consists of a comparison group of Anthem PPO enrollees, who face the same provider networks and provider prices, but who are not CalPERS enrollees. Thus, these individuals face the same insurer-negotiated prices but are not subject to reference pricing.[9]

For both the treatment and comparison populations, we restrict the sample populations to California residents ages 18-64 who have received at least one of the four procedure categories subject to reference pricing within California. These services are identified using the Current Procedural Terminology (CPT) codes in the claims data. For simplicity, our arthroscopy category contains only the two most common types of arthroscopy (knee and

[9] The provider networks and prices faced by CalPERS and the comparison group are identical because Anthem negotiates the provider networks and providers jointly for both groups, independent of employer influence.

shoulder) and our total joint replacement category contains only the two most common types of joint replacement (knee and hip). Similarly, our colonoscopy category contains the three common types of colonoscopy (screening, diagnostic, and interventional). Finally, our cataract surgery category contains cataract removal procedures.

The claims data contain detailed information on patient characteristics, provider identifiers, and characteristics of the surgical procedure. In our regressions, we include controls for patient age, gender, and differences in patient health. To adjust for differences in patient health, we use each patient's Charlson Comorbidity risk score (Charlson et al. (1987)). The Charlson comorbidity score is based on the diagnosis codes in each patient's medical claims and is a standard measure of comorbid conditions.

Our key dependent variables are: total spending of the insurer and total patient spending. We define the insurer's spending as the total allowed amount less patient spending. We define patient out-of-pocket spending as the sum of all copayments, coinsurance, deductible payments, and any additional non-covered payments (particularly those above the reference price). We will refer to insurer spending as the insurer payment, and patient spending as the patient out-of-pocket payment.[10]

We analyze these two payments separately because of the design of reference pricing, which specifically alters patient cost-sharing. In particular, reference pricing may increase, decrease, or leave unaffected patient out-of-pocket payments, depending on the size of the consumer response to the program.

4. Methods

To examine the distributional effects of the reform, we apply the CIC method of Athey

[10] Expenditures are deflated to 2013 levels.

and Imbens (2006). To provide intuition behind the CIC model, suppose that patients are randomly assigned to a PPO health plan. Some plans include a reference pricing policy (forming a treatment group) and some do not (forming a comparison group). In this case, the quantile treatment effect (QTE) can be directly identified from the insurer payment and patient out-of-pocket payment distributions. At any given quantile, the QTE is the difference in the i^{th} quantile of the distribution of interest in the treatment group versus the comparison group. The QTE difference is illustrated in Figure 1, Panel A, in which the QTEs are measured by the horizontal differences in the cumulative distribution functions (CDFs) of the outcome across the treatment and the comparison group.

However, with observational data, identification of the QTE requires specific assumptions about the counterfactual CDF for the treatment group. Using the CIC approach, we assume that the change (growth or contraction) in payments for the comparison group from t=0 to t=1 is a reasonable estimate of the change (growth or contraction) in payments that would have occurred in the treatment group had the treatment group not been subject to the implementation of reference pricing. Furthermore, we make the less stringent assumption of rank similarity, rather than rank preservation (Melly and Santangelo, 2015)

As a stylized example of the CIC approach, we use the notation of Havnes and Mogstad (2015). Specifically, let $F_t(y)(G_t(y))$ be the distribution of payments for the treatment (comparison) group, where y is the amount paid, and t=0 refers to the pre-implementation period, while t=1 refers to the post-implementation period.

Figure 1, Panel B presents illustrations of how the QTE are estimated using the CIC approach. For example, in Panel B, F_0(y), the dashed blue line, is the CDF of treatment group in the pre-period (t=0), and F_1(y), the solid blue line, is the CDF of the treatment group in the post period (t=0). G_0(y), the dashed red line, is the CDF of the comparison group in the pre-period (t=0), and G_1(y), the solid red line, is the CDF of the comparison group in the post period (t=0). The three steps to estimate the CIC estimator are as follows: first, for each value of y that is on the support of $F_1(y)$ (e.g., point A), determine the quantile of y in G_1 (e.g., point C); second, determine the level of y at that quantile in G_0 (e.g., point D); third, determine the quantile of that level of y in F_0 (e.g., point B). The counterfactual assumed by the CIC-estimator is then (A – B) – (C – D) or $F_1(y)_{CIC} = F_0(G_0^{-1}(G_1(y))$.

We implement the CIC estimator with covariates using the estimator developed by Melly and Santangelo (2015). For all results, the covariates include age, sex, and the Charlson comorbidity index. We estimate standard errors via bootstrapping using 500 repetitions for all but one result, for which we used 100 repetitions due to processing time considerations.

Due to differences in where the reference price is set in the price distribution, we separately examine the distributional impacts of the program for each procedure. The specification for colonoscopies also includes indicators for screening colonoscopies (versus diagnostic colonoscopies). The specification for arthroscopies includes an indicator for knee arthroscopy (versus shoulder). The specification for total joint replacement analyzes hip replacements and knee replacements together to be consistent with earlier research.

5. Results

5.1 Descriptive Statistics

Table 1 presents descriptive statistics for the CalPERS and non-CalPERS populations for both the pre-reference pricing and post-reference pricing periods. The pre-reference period includes 2008-2010 for total joint replacement and 2009-2011 for arthroscopy, colonoscopy, and cataract surgery. The post-reference pricing period includes 2011-2014 for total joint replacement and 2012-2013 for arthroscopy, colonoscopy, and cataract surgery. Across all procedures, there is little baseline difference between the CalPERS and the non-CalPERS populations, and the demographics of these two populations do not meaningfully change following the implementation of the reference pricing program.

Across the pre- and post-implementation periods of the reference pricing program for total joint replacement, simple unconditional difference-in-differences (DiD) estimates computed from the table show that CalPERS patients using a VBPD increased by 19.3 percentage points. Across the pre- and post-implementation periods of the reference pricing program for colonoscopy, arthroscopy, and cataract surgery, simple unconditional DiD estimates show that CalPERS patients using an ASC increased by 8.6 percentage points for knee arthroscopy, 5 percentage points for shoulder arthroscopy, 9.2 percentage points for screening colonoscopy, 6.6 percentage point for diagnostic colonoscopy, and 3.5 percentage points for cataract surgery.[11]

Table 1 additionally shows that the reference pricing policy led to sizable changes in the mean insurer payment per procedure. Simple unconditional DiD estimates of the mean impact of reference pricing are informative. For total joint replacement, the DiD for the mean insurer payment is -$5,924 (-19.2%), while the unconditional DiD for the mean patient out-of-pocket payment is $2,036 (96.3%). For the outpatient procedures, the unconditional DiD for mean insurer payment and the unconditional DiD for mean patient out-of-pocket payments, respectively, are as follows: knee arthroscopy -$468 (-17.6%), -$296 (-7.4%); shoulder arthroscopy -$700 (-21.1%), -$995 (-15.3%); screening colonoscopy $112 (31.5%), -$354 (-26.9%); diagnostic colonoscopy -$76 (-11.9%), -$240 (-17.2%); and cataract surgery $346 (36.5%), $503 (19.7%). However, these estimates obscure the distributional effects of the policy.

5.2 Changes-in-Changes Results

[11] Note that unlike previously published research, individuals exempt from the CalPERS reference pricing policy are included in these estimates, since we are unable to determine who would similarly be "exempt" from the comparison group.

5.2.1 Patient Out-of-Pocket Payments

Our results show that the distributional effects of reference pricing on patient out-of-pocket payments vary across the four procedure categories. Full procedure and quantile results are shown in Tables 1 to 8 in the Appendix.

As shown in Figure 2, the pointwise CIC for joint arthroscopy results show very small reductions in patient out-of-pocket payments starting at the 0.6^{th} quantile. At the 0.6^{th} quantile, we observe a $94 reduction in patient out-of-pocket payments. This reduction increases constantly to a $161 reduction at 0.8^{th} quantile. However, we observe a non-statistically significant $1,654 increase in patient cost sharing at the 0.9^{th} quantile.

The results are different for colonoscopy (Figure 3). For colonoscopies, the QTE estimates for out-of-pocket payments are zero for all quantiles below the 0.6^{th} quantile. Starting at the 0.6^{th} quantile, we observe a $103 increase in patient cost sharing, which steadily increases to a $586 increase at the 0.9^{th} quantile. The large mass of patients with no cost sharing is likely due to ACA cancer screening coverage protections that eliminate cost sharing for most screening colonoscopies. The distributional impacts are even more extreme for cataract surgery (Figure 4) and joint replacement (Figure 5), where we do not find statistically significant changes below the 0.9^{th} quantile. However, at the 0.9^{th} quantile, we observe a large increase, $3,381, in patient out-of-pocket payments for cataract surgery and an even larger increase of $6,146 for joint arthroscopy.

While existing studies have generally found that the CalPERS program leads to increases in mean patient out-of-pocket payments, these results show that simply examining mean impacts obscures the distributional effects. In particular, this analysis shows where in the spending distributions these increases are derived - specifically, the increase in payments is largely driven by an increase in the right-tail of the patient spending distributions.

5.2.2 **Insurer Payments**

Given the program design, insurer payments are expected to decrease, although estimates of the QTE might vary across quantile and across medical procedures. Figure 6 shows the distributional changes in insurer payments for arthroscopy. At the 0.5th quantile and below, we observe a non-statistically significant increase in insurer payments. However, starting at the 0.65th quantile, we observe a reduction of $349, which increases to $1,703 by the 0.9th quantile. For colonoscopies, we observe reductions across the entire distribution of insurer payments (Figure 7). However, the magnitude of the reductions increases from $53 at the 0.1th quantile to $847 at the 0.9th quantile.

Similar to the patient out-of-pocket payment results, for cataract surgery and joint replacement, we do not observe impacts until the upper-end of the distribution of insurer payments. For cataract surgery (Figure 8), we do not find any statistically-significant changes in insurer payments below the 0.7th quantile. At the 0.70th quantile, we find a $908 reduction in insurer payments, which nearly doubles to $1,809 at the 0.9th quantile. As shown in Figure 9, the joint arthroscopy distributional results are even more pronounced. We again find no change below the 0.7th quantile, where we observe a $5,832 reduction in insurer payments. This reduction increases to $8,876 at the 0.75th quantile, $13,492 at the 0.8th quantile, $18,990 at the 0.85th quantile, and $21,151 at the 0.9th quantile.

The observed effects of reference pricing on the distribution of insurer payments are due to the design of the program. Specifically, patients who receive care from more expensive providers are now responsible for higher out-of-pocket payments than patients who go to lower-priced providers. Thus, there is a stronger incentive to avoid higher-priced providers, and so savings are largest at the upper-end of the price distribution. While intuitive, given the structure of the CalPERS program, standard difference-in-differences approaches that focus on mean effects miss these important distributional impacts. In addition, the distributional differences are largest for the procedures in which the reference price is set lower in the price distribution. For example, for cataract surgery and joint replacement, where the reference price is set relatively low in the price distribution, the mean spending reductions are driven by large reductions at the 0.9^{th} quantile and above.

6. Conclusion

Reference pricing is a reimbursement policy in which a health insurer sets a maximum reimbursable amount (the reference price) that they will pay for a given medical procedure. While variants of reference pricing have been used in many OECD countries to reimburse for pharmaceuticals, its application in the US for medical procedures is still relatively new.

Previous studies have found that the CalPERS program reduces mean spending per procedure. While the mean effects of the program are interesting, they do not allow us to distinguish whether reductions in spending are due to movements across all parts of the distribution (i.e., a location shift), or if there are differential shifts across the spending distributions. To address this question, we estimate the distributional impacts of reference pricing using the Changes-in-Changes method of Athey and Imbens (2006). This approach allows us to show where exactly in the patient and insurer spending distributions program savings are derived.

Our findings indicate that the reference pricing program has important distributional effects and also demonstrate the impact of a reform that specifically targets the distribution of provider prices. Rather than a uniform reduction in spending, we find that the mean changes are driven by changes at the upper end of the spending distributions. Across all procedures, increases in patient out-of-pocket payments and reductions in insurer payments are largest at the 0.9th quantile of the respective distributions. These results are informative for understanding the *variability* in patients' financial burden, relative to that of the insurer.

The observed findings may be expected given the design of the reform; yet, the skewed distributional effects still have important policy implications. In particular, while the reform generates aggregate savings (e.g. see Aouad et al. (2019)), this finding must be reconciled with patients' financial exposure to the wide-degree of procedure price variability, particularly at the upper-tail of price/spending distributions where price dispersion tends to be high. This is of particular concern for policy-makers who place higher weight on patients most adversely affected by the reform.

While the setting we study is limited to a single insurance design and employer (CalPERS), it is a large employer. To our knowledge, these types of distributional impacts have not been examined in other similar programs. Thus, future studies should examine other programs and populations. As we show, examining the distributional impacts of insurance reforms is important for informing policy makers of how patients are affected by such reforms and how subsequent program savings are generated.

References

Aouad, M., Brown, T. T., and C. M. Whaley (2019). Reference pricing: The Case of Screening Colonoscopies. *Journal of Health Economics*, *65*, 246-259.

Athey, S., and G. W. Imbens. (2006). Identification and inference in nonlinear difference-in-differences models. *Econometrica*, *74*(2), 431-497.

Berwick, D. M. and A. D. Hackbarth (2012, April). Eliminating Waste in US Health Care. *JAMA 307* (14).

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.

CMS (2019). National Health Expenditures Fact Sheet. Technical report, Centers for Medicare & Medicaid Services.

Eisenberg, M. D., Haviland, A. M., Mehrotra, A., Huckfeldt, P. J., & Sood, N. (2017). The Long Term Effects of "Consumer-Directed" Health Plans on Preventive Care Use. *Journal of health economics*, *55*, 61-75.

Galizzi, M. M., Ghislandi, S., & Miraldo, M. (2011). Effects of reference pricing in pharmaceutical markets. *Pharmacoeconomics*, *29*(1), 17-33.

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

Havnes, T., and M. Mogstad. (2015). Is universal child care leveling the playing field? *Journal of Public Economics*, *127*, 100-114.

Hussey, P. S., S. Wertheimer, and A. Mehrotra (2013). The Association between Health Care Quality and Cost: A Systematic Review. *Annals of*

Internal Medicine 158 (1), 27–34.

Manning, Willard G., Joseph P. Newhouse, Naihua Duan, Emmett B. Keeler, and Arleen Leibowitz. "Health insurance and the demand for medical care: evidence from a randomized experiment." *The American Economic Review* (1987): 251-277.

Melly, B., & Santangelo, G. (2015). The changes-in-changes model with covariates. *Universität Bern, Bern*.

Newman, D., Parente, S. T., Barrette, E., & Kennedy, K. (2016). Prices for common medical services vary substantially among the commercially insured. *Health Affairs*, *35*(5), 923-927.

Robinson, J.C. and Brown, T.T., 2013. Increases in consumer cost sharing redirect patient volumes and reduce hospital prices for orthopedic surgery. *Health Affairs*, *32*(8), pp.1392-1397.

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

Robinson, James C., Timothy T. Brown, and Christopher Whaley. 2017. “Reference Pricing Changes The ‘Choice Architecture’ Of Health Care For Consumers.” *Health Affairs* 36 (3): 524–30.

Whaley, C. M. 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.

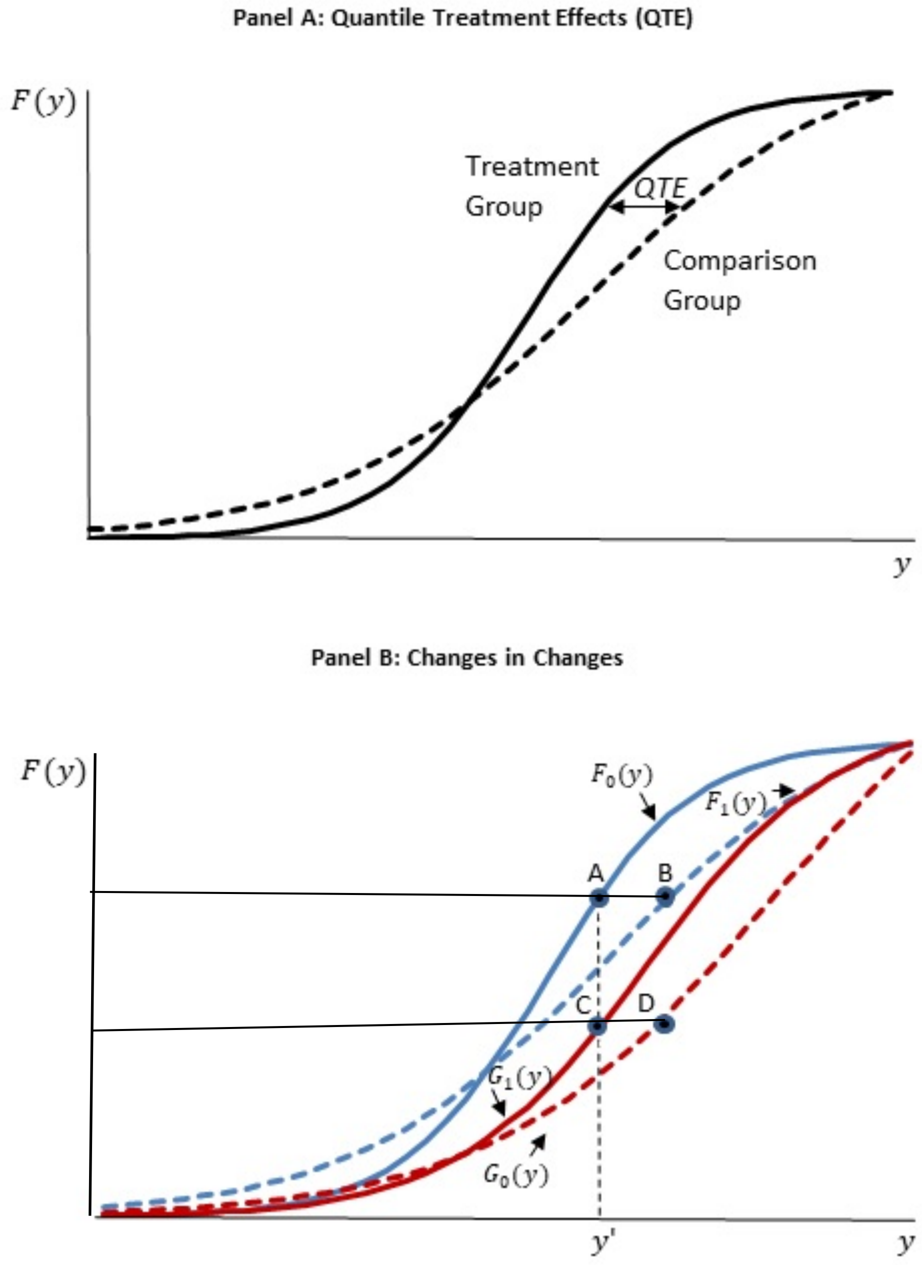
Whaley, C. M., C. Guo, and T. T. Brown (2017). The Moral Hazard Effects of Consumer Responses to Targeted Cost-sharing. *Journal of Health Economics 56*, 201–221.

Zhang, X., Haviland, A., Mehrotra, A., Huckfeldt, P., Wagner, Z., & Sood, N. (2018). Does Enrollment in High-Deductible Health Plans Encourage Price Shopping?. *Health Services Research*, *53*, 2718-2734.

Table 1. Descriptive Statistics

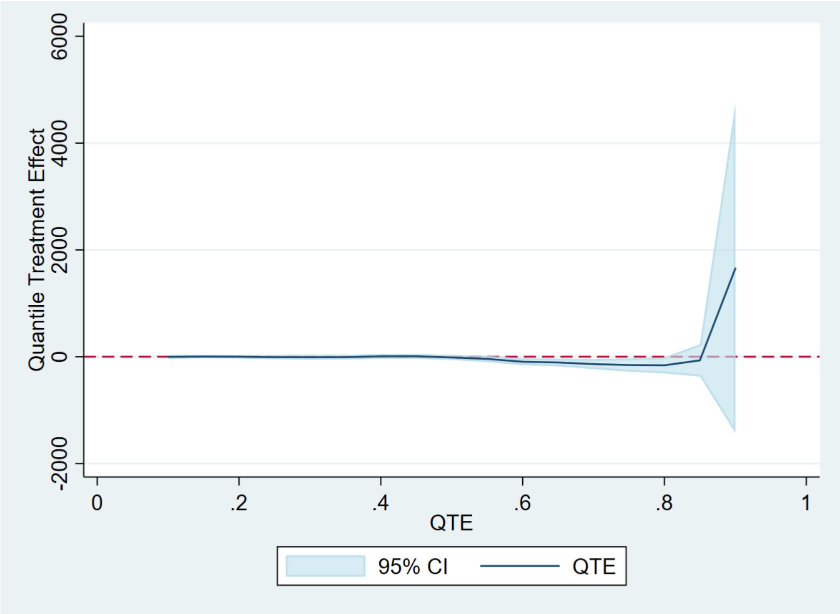
	Pre-Reference Pricing				Post-Reference Pricing			
	CalPERS		Non-CalPERS		CalPERS		Non-CalPERS	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Arthroscopy								
Insurer Payment	$4,794	$5,737	$4,383	$5,537	$4,801	$6,255	$4,862	$6,291
Consumer Payment	$2,866	$7,640	$2,763	$7,765	$3,600	$10,135	$4,005	$11,511
Share of Procedures at ASC	61.9%	-	67.3%	-	69.2%	-	67.1%	-
Patient Age	51.0	11.4	47.1	12.4	51.3	11.7	47.3	12.6
Share of Male Patients	50.8%	-	58.6%	-	50.4%	-	58.0%	-
Patient Charlson Index	0.1	0.3	0.1	0.3	0.1	0.3	0.1	0.3
Share of Knee Procedures	68.3%	-	68.9%	-	66.2%	-	69.0%	-
Observations	4,317		35,238		2,751		24,501	
Colonoscopy								
Insurer Payment	$1,359	$1,282	$1,027	$1,298	$1,357	$1,377	$1,317	$1,473
Consumer Payment	$504	$1,940	$875	$2,445	$751	$3,003	$1,114	$4,077
ASC Use	66.6%	-	71.9%	-	75.0%	-	72.4%	-
Patient Age	54.7	7.5	52.6	8.7	54.8	7.9	52.8	8.8
Share of Male Patients	44.5%	-	46.6%	-	43.2%	-	46.8%	-
Patient Charlson Index	0.1	0.4	0.1	0.4	0.1	0.4	0.1	0.4
Screening Colonoscopy	47.6%	-	44.9%	-	47.7%	-	45.5%	-
Interventional Colonoscopy	51.9%	-	53.6%	-	56.6%	-	57.4%	-
Observations	24,698		163,004		15,192		117,584	
Cataract Surgery								
Insurer Payment	$947	$2,415	$981	$2,656	$1,446	$2,962	$1,134	$3,298
Consumer Payment	$2,557	$2,213	$2,275	$2,307	$2,072	$1,898	$2,293	$2,540
ASC Use	73.4%	-	75.2%	-	81.1%	-	79.4%	-
Patient Age	59.0	5.0	57.6	6.3	59.3	5.2	57.8	6.4
Share of Male Patients	41.1%	-	47.0%	-	37.7%	-	45.1%	-
Patient Charlson Index	0.1	0.4	0.1	0.4	0.1	0.4	0.1	0.4
Observations	2,268		12,773		1,623		9,973	
Hip and Knee Procedures								
Insurer Payment	$30,932	$21,738	$28,913	$22,384	$24,584	$10,162	$28,489	$16,926
Consumer Payment	$2,114	$5,122	$2,170	$7,150	$4,251	$8,343	$2,271	$3,690
VBPD Use	48.7%	-	54.7%	-	65.8%	-	52.5%	-
Patient Age	57.8	5.9	56.7	6.3	58.0	5.9	56.9	6.2
Share of Male Patients	40.1%	-	48.3%	-	40.7%	-	47.2%	-
Patient Charlson Index	0.9	1.3	0.8	1.4	0.9	1.4	0.7	1.3
Share of Hip Procedures	40.2%	-	49.5%	-	43.8%	-	51.1%	-
Observations	753		2,879		1,577		6,650	

Figure 1. Quantile Treatment Effects Estimated Using Changes in Changes



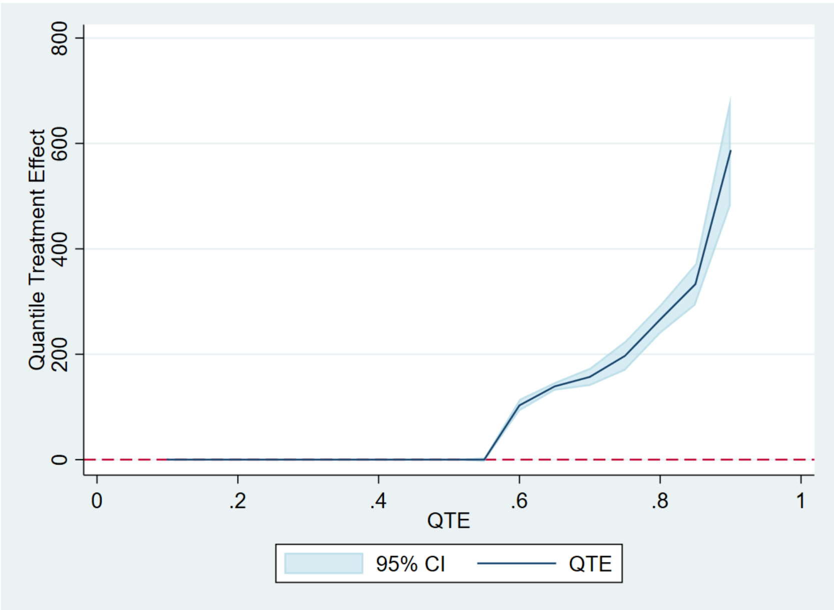
Panel A: Stylized illustration of quantile treatment effects (QTE).
Panel B: $F_t(y)$ and $G_t(y)$ are stylized outcome distributions for treated and comparison groups, respectively, before (t=0) and after (t=1) reference pricing implementation. Changes in changes (CIC): (A – B) – (C – D).

Figure 2. Patient Payment (Level $): Arthroscopy



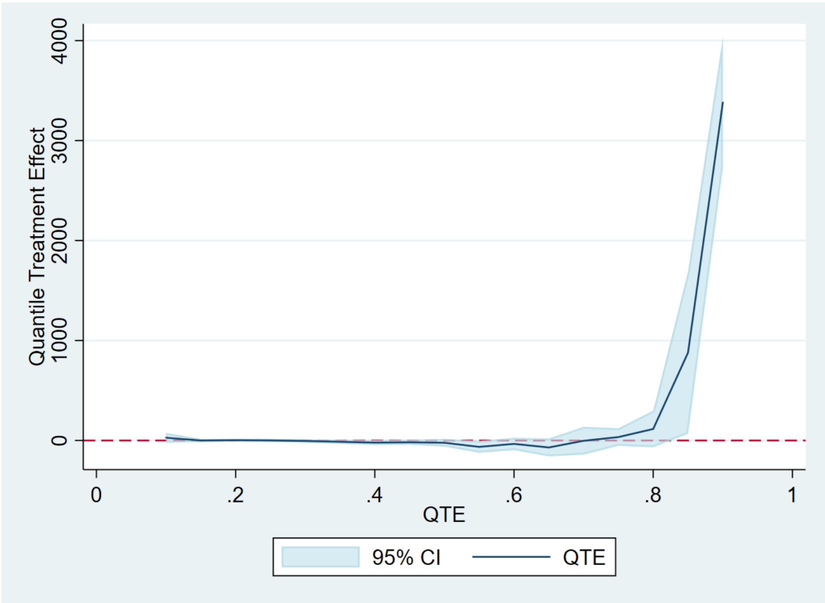
CI: 95% confidence interval.

Figure 3. Patient Payment (Level $): Colonoscopy



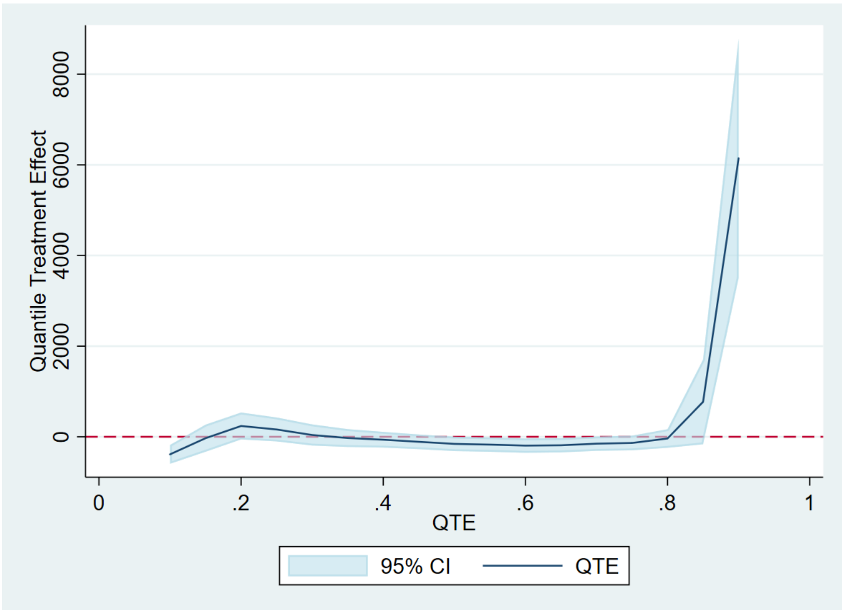
CI: 95% confidence interval.

Figure 4. Patient Payment (Level $): Cataract Surgery



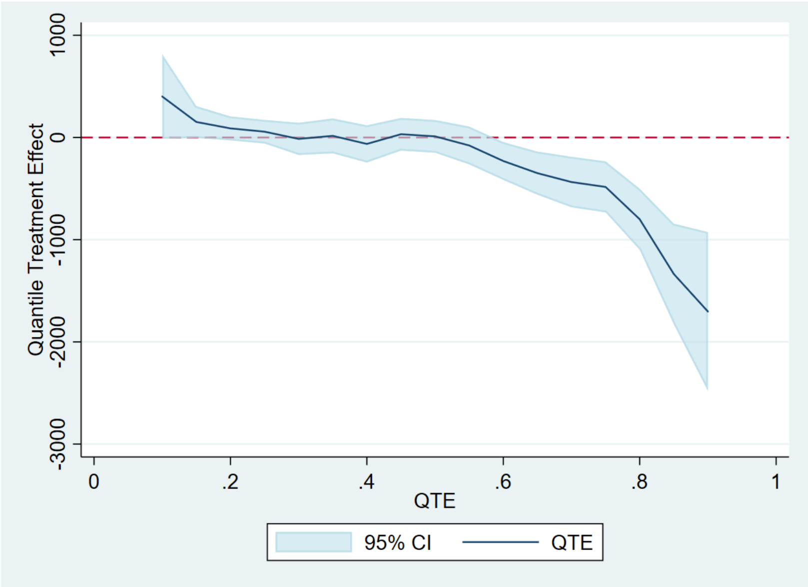
CI: 95% confidence interval.

Figure 5. Patient Payment (Level $): Total Joint Replacement



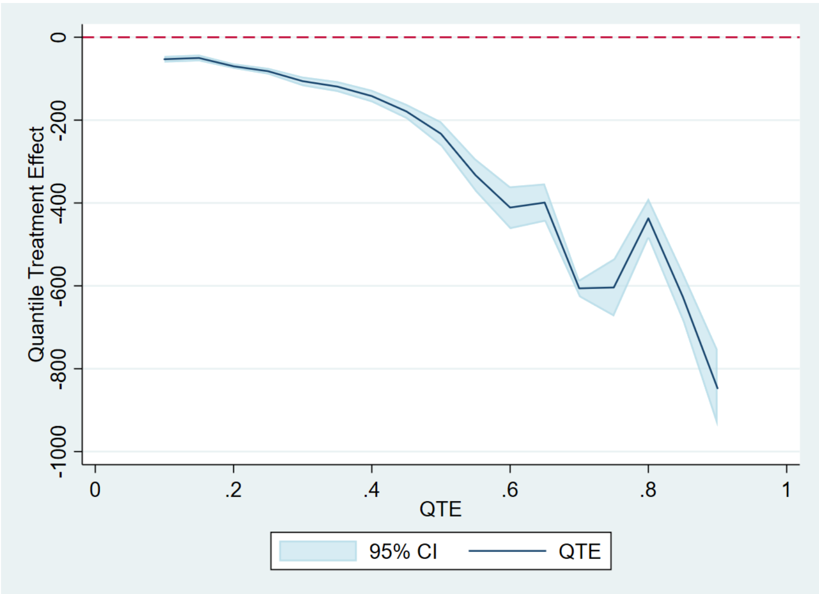
CI: 95% confidence interval.

Figure 6. Insurer Payment (Level $): Arthroscopy



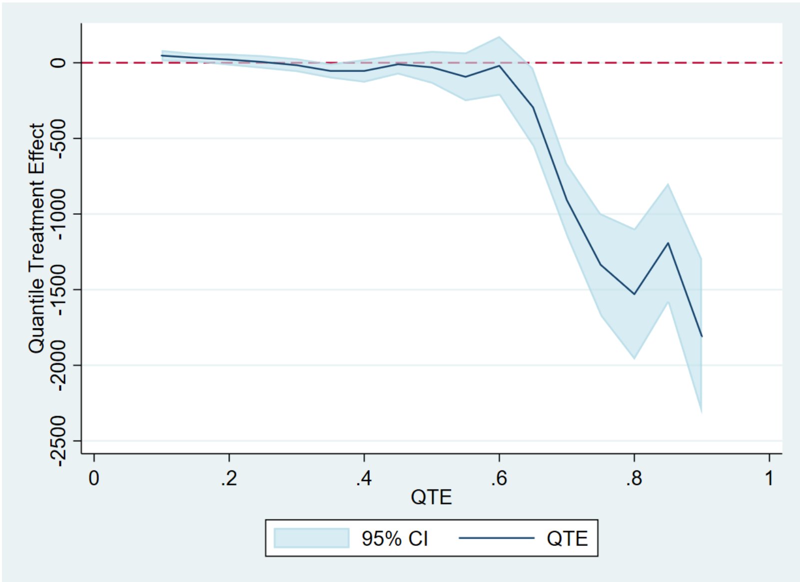
CI: 95% confidence interval.

Figure 7. Insurer Payment (Level $): Colonoscopy



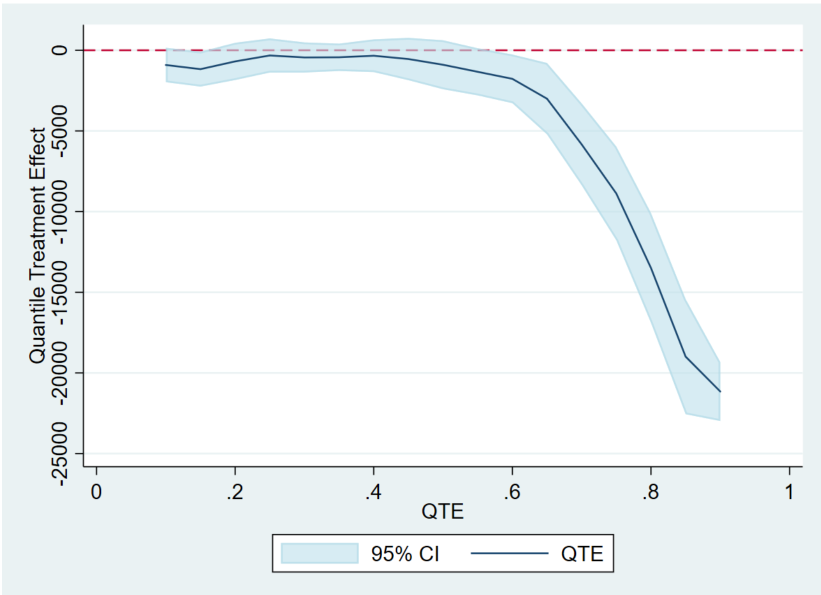
CI: 95% confidence interval.

Figure 8. Insurer Payment (Level $): Cataract Surgery



CI: 95% confidence interval.

Figure 9. Insurer Payment (Level $): Total Joint Replacement



CI: 95% confidence interval.

Supplemental Appendix

Appendix Table 1. Patient Payment (Levels in $s): Arthroscopy

Quantile	QTE	95% Confidence Interval	
0.10	-3	-39	34
0.15	3	-29	36
0.20	0	-33	33
0.25	-8	-47	31
0.30	-9	-56	38
0.35	-8	-52	36
0.40	7	-39	53
0.45	8	-39	54
0.50	-14	-61	33
0.55	-41	-100	18
0.60	-94	-159	-29
0.65	-108	-181	-36
0.70	-138	-233	-43
0.75	-156	-279	-32
0.80	-161	-312	-11
0.85	-68	-372	235
0.90	1654	-1428	4736

QTE: Quantile treatment effect

Appendix Table 2. Patient Payment (Levels in $s): Colonoscopy

Quantile	QTE	95% Confidence Interval	
0.10	0	0	0
0.15	0	0	0
0.20	0	0	0
0.25	0	0	0
0.30	0	0	0
0.35	0	0	0
0.40	0	0	0
0.45	0	0	0
0.50	0	0	0
0.55	0	-4	4
0.60	103	92	115
0.65	139	131	147
0.70	157	140	174
0.75	197	169	225
0.80	266	239	294
0.85	333	293	372
0.90	586	482	690

QTE: Quantile treatment effect

Appendix Table 3. Patient Payment (Levels in $s): Cataract Surgery

Quantile	QTE	95% Confidence Interval	
0.10	27	-23	76
0.15	0	-16	16
0.20	3	-10	16
0.25	1	-16	17
0.30	-5	-22	13
0.35	-13	-34	7
0.40	-21	-45	3
0.45	-18	-43	7
0.50	-23	-61	15
0.55	-64	-123	-4
0.60	-34	-96	27
0.65	-70	-159	20
0.70	-3	-141	136
0.75	34	-53	122
0.80	115	-68	299
0.85	879	68	1690
0.90	3381	2728	4035

QTE: Quantile treatment effect

Appendix Table 4. Patient Payment (Levels in $s): Total Joint Replacement

Quantile	QTE	95% Confidence Interval	
0.10	-387	-595	-180
0.15	-31	-329	268
0.20	239	-57	536
0.25	162	-101	426
0.30	40	-192	273
0.35	-27	-224	169
0.40	-64	-237	109
0.45	-110	-271	51
0.50	-156	-314	3
0.55	-172	-330	-14
0.60	-194	-352	-36
0.65	-187	-343	-30
0.70	-151	-309	6
0.75	-136	-297	25
0.80	-37	-242	168
0.85	771	-164	1705
0.90	6146	3507	8785

QTE: Quantile treatment effect

Appendix Table 5. Insurer Payment (Levels in $s): Colonoscopy

Quantile	QTE	95% Confidence Interval	
0.10	-53	-61	-45
0.15	-50	-58	-42
0.20	-70	-76	-63
0.25	-82	-90	-74
0.30	-106	-118	-95
0.35	-119	-132	-106
0.40	-142	-157	-127
0.45	-179	-197	-161
0.50	-233	-263	-203
0.55	-333	-373	-293
0.60	-411	-463	-360
0.65	-399	-445	-353
0.70	-606	-627	-585
0.75	-604	-674	-534
0.80	-437	-487	-388
0.85	-627	-687	-567
0.90	-847	-940	-753

QTE: Quantile treatment effect

Appendix Table 6. Insurer Payment (Levels in $s): Arthroscopy

Quantile	QTE	95% Confidence Interval	
0.10	401	-13	816
0.15	152	-4	307
0.20	89	-28	206
0.25	57	-59	172
0.30	-14	-171	144
0.35	16	-155	186
0.40	-63	-245	119
0.45	32	-128	191
0.50	11	-149	171
0.55	-78	-263	107
0.60	-230	-415	-45
0.65	-349	-560	-138
0.70	-436	-683	-189
0.75	-484	-733	-234
0.80	-800	-1100	-501
0.85	-1336	-1828	-843
0.90	-1703	-2480	-926

QTE: Quantile treatment effect

Appendix Table 7. Insurer Payment (Levels in $s): Cataract Surgery

Quantile	QTE	95% Confidence Interval	
0.10	47	10	84
0.15	33	4	63
0.20	21	-18	60
0.25	5	-39	49
0.30	-16	-61	29
0.35	-54	-103	-5
0.40	-54	-131	23
0.45	-10	-77	56
0.50	-30	-138	78
0.55	-93	-254	68
0.60	-20	-217	177
0.65	-295	-555	-35
0.70	-908	-1152	-664
0.75	-1335	-1675	-995
0.80	-1530	-1964	-1096
0.85	-1192	-1590	-794
0.90	-1809	-2324	-1295

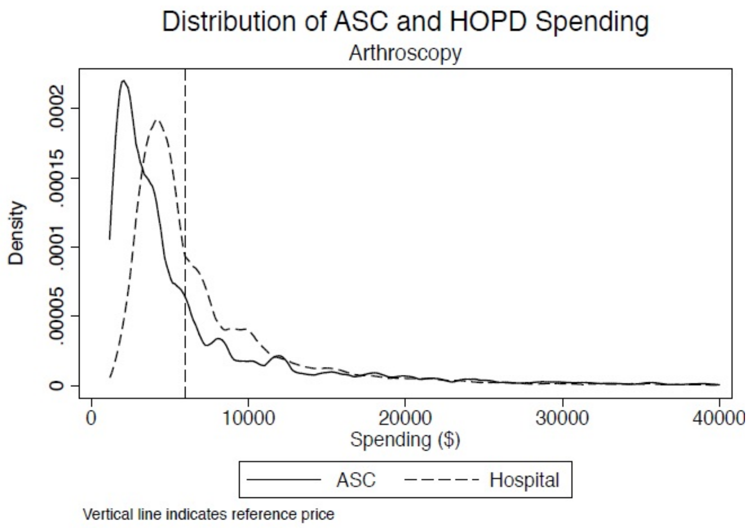
QTE: Quantile treatment effect

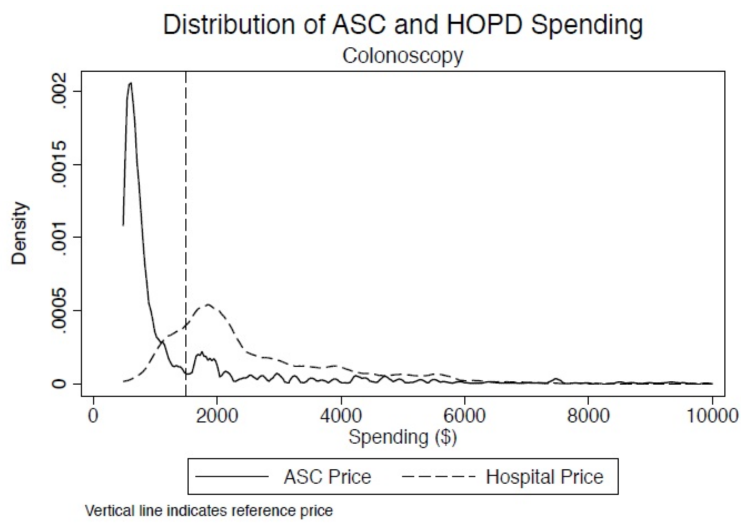
Appendix Table 8. Insurer Payment (Levels in $s): Total Joint Replacement

Quantile	QTE	95% Confidence Interval	
0.10	-911	-1979	158
0.15	-1170	-2250	-90
0.20	-690	-1843	464
0.25	-322	-1379	735
0.30	-444	-1376	488
0.35	-435	-1283	413
0.40	-337	-1351	676
0.45	-542	-1853	769
0.50	-897	-2414	621
0.55	-1334	-2796	128
0.60	-1769	-3277	-261
0.65	-3002	-5215	-788
0.70	-5832	-8371	-3294
0.75	-8876	-11783	-5970
0.80	-13492	-16897	-10087
0.85	-18990	-22557	-15424
0.90	-21151	-22979	-19324

QTE: Quantile treatment effect

Appendix Fig 1. Pre-Reference Pricing Distribution of Spending Across ASCs and Hospitals (HOPD)





This figure plots the empirical probability density functions of ASC and Hospital spending in the pre-reference pricing implementation period (2009-2011).