Title
Does the Implementation of Reference Pricing Result in Reduced Utilization? Evidence From Inpatient and Outpatient Procedures

Journal
Medical Care Research and Review, 79(1)

ISSN
1077-5587

Authors
Elser, Holly
Lin, Wei
Catalano, Ralph A
et al.

Publication Date
2022-02-01

DOI
10.1177/1077558720971117

Peer reviewed

HHS Public Access

Author manuscript

Med Care Res Rev. Author manuscript; available in PMC 2023 July 05.

Published in final edited form as:

Med Care Res Rev. 2022 February ; 79(1): 58–68. doi:10.1177/1077558720971117.

Does the Implementation of Reference Pricing Result in Reduced Utilization? Evidence From Inpatient and Outpatient Procedures

Holly Elser[1], **Wei Lin**[1], **Ralph A. Catalano**[1], **Timothy T. Brown**[1]

[1]University of California, Berkeley, CA, USA

Abstract

Reference pricing (RP) is an insurance design that can be used to incentivize patients to use low-price settings. While RP is not intended to affect overall utilization, it could unintentionally reduce utilization. We examined whether utilization was reduced when a large employer adopted RP for selected elective surgeries, including inpatient joint replacement surgery and outpatient cataract surgery, colonoscopy, and arthroscopic surgery. Data included a treatment group subject to RP implementation and a comparison group that was not. We applied autoregressive integrated moving average analysis as comparison-population interrupted time-series analysis to determine whether there were procedure reductions following RP implementation. We find no evidence of short-term decreases (within 3 months of RP implementation). However, we find very modest declines of approximately 14 (20%) fewer arthroscopic knee surgeries 6 months after RP implementation and 129 (17.2%) fewer colonoscopies 8 months after RP implementation. There were no declines in the other procedures examined.

Keywords

reference pricing; service utilization; interrupted time series; ARIMA

Introduction

Contemporary health care markets are characterized by substantial variation in prices for surgical procedures, laboratory tests, and pharmaceuticals (Robinson et al., 2017). For example, prices for many outpatient surgical procedures are considerably higher when performed in hospital outpatient departments (HOPDs) as compared with freestanding ambulatory surgical centers (ASCs), which can be at least partly attributable to both the higher costs inherently associated with hospitals and their stronger bargaining position with insurers (Cassidy, 2014; Robinson, Brown, & Whaley, 2015). Not only are there price

Corresponding Author: Timothy T. Brown, Division of Health Policy and Management, University of California, Berkeley, School of Public Health, 2121 Berkeley Way West, Berkeley, CA 94720, USA. timothy.brown@berkeley.edu.
Holly Elser is also affiliated with Stanford University School of Mediciane, Stanford, CA.

Declaration of Conflicting Interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

differences across classes of facilities, there is also a large amount of price dispersion within classes of facilities, such as hospitals, that occur for a variety of reasons (Robinson et al., 2012). Both types of price dispersion have stimulated interest among insurers and employers hoping to encourage beneficiaries to select low-price facilities.

One approach to encourage beneficiaries to select low-price facilities is reference pricing (RP). With RP, the employer or insurer establishes a reference-based limit on payment, known as the reference price, which provides ordinary insurance coverage if the enrollee chooses a low-price setting but requires beneficiaries to pay an additional amount when choosing a high-price setting. Specifically, when choosing a high-price setting, patients pay their usual coinsurance and deductible, which are applied to the reference price. Patients must then pay 100% of any difference between the allowed charge and the reference price, which will be zero or positive depending on the size of the difference between the reference price and the allowed charge.

RP has been implemented either symmetrically or asymmetrically, depending on the procedure. When RP is implemented symmetrically, it applies to all facilities. An example of this is the application of RP to a surgical procedure that is only performed in hospitals. In this situation, the reference price itself effectively differentiates low-price and high-price facilities (Robinson & Brown, 2013; White et al., 2014). Alternatively, RP may be implemented asymmetrically and only apply to some types of facilities, such as HOPDs, and not to other types of facilities, such as ASCs. This latter approach may be implemented to simplify patient choice when one type of facility has been found to usually charge a higher price than the other type of facility (Robinson, Brown, & Whaley, 2015). RP is most appropriately applied to nonemergency procedures that allow for informed comparison, in settings supporting consumer choice, and in policy contexts in which consumer financial responsibility is acceptable (Robinson et al., 2017).

The California Public Employees’ Retirement System (CalPERS) is self-insured and provides health insurance to approximately 1.4 million employees of the State of California and other public entities, of whom approximately 225,000 are enrolled in Preferred Provider Organizations (PPOs; Robinson, Brown, & Whaley, 2015; Zhang et al., 2017). The RP program was only implemented in these PPOs (Robinson, Brown, & Whaley, 2015; Zhang et al., 2017). CalPERS first applied RP to two inpatient procedures, knee and hip replacement surgeries, in January of 2011, and later extended the RP program to outpatient procedures— including colonoscopies, cataract surgeries, and arthroscopic surgeries—in January of 2012. All procedure types were selected because of substantial dispersion in provider prices and a lack of evidence that high-price providers provided higher quality care. Colonoscopies, cataract surgeries, and arthroscopic surgeries are outpatient procedures that can be performed safely and effectively in ASCs, and all of the services subject to RP are “shoppable” insofar as they are nonemergency procedures whose outcomes are not compromised if a patient delays the procedure for a short period of time to compare providers (Brown et al., 2018; Robinson, Brown, & Whaley, 2015). The CalPERS RP program has been found to shift a substantial portion of patients from high-price to low-price facilities, substantially lowering expenditures, without affecting complication rates (Brodke et al., 2019; Robinson, Brown, Whaley, & Bozic, 2015; Robinson, Brown, Whaley, & Finlayson, 2015).

New Contributions

While there has been a difference-in-differences (DID) analysis of the utilization for hip and knee replacement surgery in a PPO versus health maintenance organization (HMO) population that found no change in utilization due to the introduction of RP (Zhang et al., 2017), this study performs the first formal determination of the effect of RP on utilization of all surgical procedures subject to RP within CalPERS whether subject to the symmetrical or asymmetrical application of RP.

We examine the effect of the implementation of RP on overall procedure utilization among CalPERS enrollees covered by Anthem Blue Cross PPO plans between 2008 and 2014 for hip and knee replacements and between 2009 and 2013 for cataract surgery, colonoscopy, and arthroscopy. We focus specifically on the 3 months immediately following RP, during which time sudden decreases in service utilization would be most plausibly attributable to RP and would be more economically disruptive to insurers, hospitals, and physicians who otherwise expected a higher volume of procedures. We additionally consider the possibility that implementation of RP could lead to delayed decreases in utilization of elective hospital procedures up to 1 year after it was introduced.

Conceptual Framework

The quantity of any given surgical procedure is determined by the demand for and supply of that procedure. The RP program was designed to minimize any impact on overall quantity, and primarily impact the relative distribution of overall quantity across facility types. This included design features intended to minimize factors that may impact consumer choice regarding whether to receive a given surgical procedure and design features intended to minimize any impact on physician choice regarding whether to recommend a given surgical procedure.

With regard to the impact of RP on beneficiaries' choices regarding whether to receive a given surgical procedure, this was minimized primarily by ensuring that the out-of-pocket price for a procedure subject to RP did not change due to the implementation of RP when the procedure was performed at a select group of facilities (which we refer to as low-price facilities). The out-of-pocket price for procedures subject to RP only increased if beneficiaries chose to receive their procedure from the remaining set of facilities (which we refer to as high-price facilities). To the extent that the RP policy resulted in beneficiaries viewing low-price and high-price facilities as sufficiently close substitutes, there should be no significant impact on the overall demand for the relevant surgical procedures, only a switch from high-price facilities to low-price facilities, other things equal.

Two additional RP program features were designed to prevent the lack of an acceptable facility from being a reason for beneficiaries not receiving a procedure, where an acceptable facility was based on distance and/or quality. To prevent distance from affecting consumer choice regarding whether to receive a procedure, the RP program exempted beneficiaries from being subject to RP if low-price facilities were located 30 or more miles from the consumer (Brown & Atal, 2019; Robinson & Brown, 2013). To prevent quality from affecting consumer choice regarding whether to receive a procedure, the RP program exempted beneficiaries from being subject to RP if their physician provided a medically documented reason that the consumer's procedure needed to be performed at a high-price facility (Brown & Atal, 2019; Robinson & Brown, 2013).

That the goals of removing relative distance and relative quality from being relevant factors in consumer choice were largely achieved is supported by the following findings. There was no statistical difference in average distance traveled due to the implementation of RP for colonoscopy (Whaley et al., 2017) and no statistical difference in average complication rates (an objective measure of one important measure of quality) for colonoscopy, arthroscopy, or hip/knee replacement due to the implementation of RP (Brodke et al., 2019; Brown & Atal, 2019; Chi et al., 2014; Robinson, Brown, Whaley, & Bozic, 2015; Robinson, Brown, Whaley, & Finlayson, 2015).

With regard to the supply of surgical procedures, hospitals and ASCs codetermine physician choices regarding whether and where to provide a given surgical procedure through surgery schedules and the amenities (or lack thereof) that hospital and ASCs provide to physicians. Holding these influences constant, the RP policy was designed to minimize any alteration of the incentives faced by physicians regarding whether to recommend a given surgical procedure. This was done by exempting physician fees from RP; only facility fees were affected (Brown & Atal, 2019; Robinson & Brown, 2013). Thus, the RP policy largely minimized altering physician incentives to engage in supplier-induced demand. While physicians with ownership stakes in facilities may still be affected, physicians are more likely to have an ownership stake in a local ACS than in a local hospital (Gabel et al., 2008; Hollingsworth et al., 2010), and ACS facilities were excluded from RP, such that no change in per patient reimbursement occurred at ACS facilities due to the RP policy.

Thus, the major decision for the average patient subject to RP was whether to patronize low-price or high-price facilities, both of which were sufficiently close and had sufficiently similar levels of quality, for a given surgical procedure that was recommended by a physician whose fee for the procedure was not affected by RP. If a patient patronized the low-price facility, they paid their deductible and coinsurance, whereas if they patronized the high-price facility they paid their deductible and coinsurance but also paid 100% of any additional amount above the reference price. In other words, deductibles and coinsurance were the same in both low-price and high-price facilities. Deductibles were not reduced in any setting due to the implementation of RP.

Nevertheless, although not intended to do so, RP may still result in a reduction in the utilization of procedures subject to RP if a sufficient number of beneficiaries chose to defer or cancel procedures subject to RP due to a strong preference for a high-price facility for reasons other than distance or quality and also had sufficiently strong sensitivity to the additional out-of-pocket costs at a high-price facility. A decrease in utilization could also occur if beneficiaries switched out of their PPO insurance plan during the open enrollment period in response to or in anticipation of the effects of RP. However, the only available health plans within CalPERS that did not use RP were HMO plans, which are even more restrictive in their choice of settings for the relevant procedures. Thus, beneficiaries who wished to avoid being subject to formal RP as implemented in the available PPO plans would only be able to do so by choosing an HMO plan that implicitly used a more extreme form of RP in which beneficiaries would face out-of-pocket payments of up to 100% of the price of the procedure if they choose a facility outside of the small set of facilities available to them within their HMO. Even so, given past evidence of consumer backlash against managed care elements in insurance plans, some consumers may still prefer to switch from the PPO plan to an HMO plan when RP was added to the PPO plan.

Similarly, RP may still result in at least some physicians being less likely to recommend a procedure covered by RP if that procedure would not be performed in a facility in which they had an ownership stake due to this facility being designated a high-price facility. Such a physician may simply shift their recommendation for the relevant procedures away from patients whose procedures where subject to RP and toward patients whose procedures where not subject to RP.

The argument above does not yield an unequivocal prediction with regard to whether the impact of RP on the quantity of relevant surgical procedures will be zero or negative. Thus, we examine the null hypothesis that no change in overall utilization for any of the relevant procedures occurred due to the implementation of RP against the one-sided alternative hypothesis that there was a decrease in overall utilization due to the implementation of RP. To test this hypothesis, we assume that no concomitant reforms or events occurred that would influence the utilization of the relevant procedures among CalPERS beneficiaries apart from the implementation of RP. This assumption appears to be reasonable based on our investigation of the existence of such concomitant reforms/events. Note that we also include a two-sided test for comparison purposes.

Apart from concomitant reforms/events, the likely existence of other trends may influence our results and must be accounted for. We thus compare longitudinal surgical volumes of joint replacement, arthroscopy, colonoscopy, and cataract surgery in the treatment group to the same categories of longitudinal surgical volumes in a comparison group. Our treatment group is made up of CalPERS beneficiaries who became subject to RP during the relevant time period, while our comparison group is made up of those covered by the same insurer used by CalPERs but did not become subject to RP during the same time period.

Method

Data and Measures

Study data were composed of monthly medical claims data gathered from January 2008 to December 2014 for hip and knee replacements and from January 2009 to December 2013 for colonoscopy, arthroscopy, and cataract surgery. In January of 2011, all CalPERS enrollees covered by Anthem Blue Cross PPO products became subject to RP for hip and knee replacement surgery, which was applied symmetrically to all patients receiving either procedure in any hospital, not including patients exempted due to a clinical condition requiring treatment in a high-price setting and not including patients living more than 30 miles from a low-price facility. By January of 2012, all CalPERS enrollees covered by

Anthem Blue Cross PPO products were also subject to asymmetric RP for colonoscopy, arthroscopy, and cataract surgery with RP applying to these procedures when performed in a HOPD, and RP not applying to these procedures when performed in an ASC (subject to the same two exemptions listed above).

The study population is composed of two distinct groups of beneficiaries. The treatment group included enrollees in CalPERS self-insured PPO products, which were administered by Anthem Blue Cross. The comparison group were non-CalPERS beneficiaries covered by PPO products issued by the same insurance carrier, but who were not subject to RP. Because Anthem negotiates the actual payment to facilities for all of its membership, prices for procedures performed in any given hospital or ASC are identical for CalPERS enrollees and non-CalPERS enrollees. Our analysis was restricted to patients aged 18 to 64 years who resided in California and received their procedure in California.

Statistical Analysis

Due of data limitations, we do not have data on utilization rates, only utilization volume. We thus analyze this data using comparison-population interrupted time-series analysis, an application of autoregressive integrated moving average analysis (ARIMA). ARIMA models effectively control for all factors that affect both the treatment and comparison populations as well as those that affect only the treatment population and exhibit autocorrelation when the size of the population at risk varies similarly over time in both the test and comparison populations, thereby obviating the need to calculate utilization rates. We thus evaluate our hypothesis for each surgical procedure using comparison-population interrupted time-series analysis.

To further understand our approach, it is helpful to distinguish it from the more commonly used DID approach. Using the difference between an outcome measured in a treatment and comparison population as a dependent variable makes sense when intending to control for what the epidemiologic literature refers to as “generally occurring” (i.e., affecting all populations rather than just the treated—also referred to as “common shocks”) and “inaccessible” (i.e., things we cannot specify because we do not suspect or cannot measure) third variables (Catalano & Serxner, 1987). The DID approach cannot, however, control for “local” (i.e., affects only the treated population), inaccessible third variables that exhibit autocorrelation. In other words, comparison-population interrupted time-series and DID logically converge when the time series of differences in the differenced outcome variables exhibits no autocorrelation, but logically diverge when autocorrelation is present. Put another way, comparison-population interrupted time-series can control for confounders that are autocorrelated, but not other types of violations of the DID common shock assumption. Finally, DID as typically implemented does not help with the problem of inefficient estimation arising from autocorrelation peculiar to the test population. DID does not, in other words, ensure the estimation has met the assumption of normally distributed and independent error terms.[1]

[1]. DID as typically implemented does not include a test of whether the differences between the differenced outcomes in the treatment and comparison populations exhibit autocorrelation. Given that a binary shock cannot, by axiom, induce autocorrelation in a series before the shock occurs, then autocorrelation in the DID will carry into the error terms of the estimating equation. This violates

The comparison-population interrupted time-series analysis approach used here regresses the dependent variable in the treated population on that in the comparison population and thereby has all the benefits of the DID approach (i.e., it controls for generally occurring, inaccessible third variable) and, by detecting and specifying autocorrelation in the error terms of that regression, also controls for locally occurring inaccessible third variables that exhibit autocorrelation (i.e., trends, seasonality, and the tendency to remain elevated or depressed, or to oscillate, after high or low values). By specifying autocorrelation in the error terms, moreover, the time-series approach has the added advantage of producing efficient estimators because the assumption of independent and normal error terms has been met.

Our analysis does not include covariates since specifying covariates would only be appropriate if the covariates affected only the treated population, exhibited no autocorrelation, and were both suspected and measured.[2] We know of no measured covariates that exhibit no trends, cycles, or tendencies to remain elevated or depressed after high or low values that affect only the treated population.

In the comparison-population interrupted time-series analysis used here, increases or decreases in utilization due to behavioral response to the anticipated implementation of RP that persist for at least two consecutive months will be picked up as a level shift. Our test for level shifts thus identifies sequences of outlying values that consistently rise or fall following implementation of RP and at some point, form a new “plateau” but take more than one month to emerge.

This test does not identify sequences of outlying values that form a continuous downward or upward slope that extends to the end of the test period. The test does not identify such continuous upward or downward slopes for two reasons. First, the theory that justifies RP does not predict that procedures will fall to 0 or be performed on the entire enrolled population (i.e., the logical extension of a continuing trend line in either direction). Second, such a trend would dominate the data and our methods would detect it not as an outlying sequence but as autocorrelation and, as such, information from which to build the counterfactual.

We consider the association of RP with the number of inpatient knee and hip replacement surgeries as well as outpatient procedures: knee and shoulder arthroscopies; colonoscopies; and cataract removal surgeries. For hip and knee replacement surgeries, we calculate the number of procedures performed using data from each month from January 2008 through December 2014. For outpatient surgical procedures, we calculate the number of each

the assumption of independence of error terms—a violation not corrected by the typical adjustment for clustering because such adjustments do not exhaust all forms of autocorrelation (e.g., moving averages at orders higher than t-1 or autoregression, including seasonality, at orders higher than $t-2$ or $t-3$). Box–Jenkins detection and ARIMA modeling can, in fact, be thought of as akin to adjustment of episodic clustering.

2. Our analysis does not include covariates since specifying covariates would only be appropriate if the covariates affected only the treated population, exhibit no autocorrelation, and were both suspected and measured. This is because first, an inaccessible covariate cannot, by axiom, appear in either approach to the problem. Second, an accessible “known confounder” thought, a priori, to affect only the test population should be included regardless of whether the test uses DID or comparison-population interrupted time-series analysis. Third, accessible confounders thought, a priori, to affect both populations should not appear as a covariate in either DID or comparison-population interrupted time-series analysis. Including them would violate the logic of the design and risks biasing the test toward a Type I error by shrinking, even a bit, variance in the error term.

procedure performed using data from each month from January 2009 to December 2013. For all outcomes of interest, we create separate time-series of monthly counts for both the treatment and comparison groups.

Our test essentially compares the observed monthly number of procedures to the number expected under the assumption that RP had no effect (i.e., the "counterfactual"). The argument that RP reduces utilization predicts that the observed counts will drop significantly below the counterfactual values at or near the adoption of RP. We arrived at the counterfactual through the following steps:

1. We regressed the monthly number of procedures performed in the RP group on the number performed in the comparison group and obtained the residuals. These residuals represent that portion of the time-series that differs between the two groups.

2. We used Box–Jenkins methods to detect and model autocorrelation in the residuals of the regression performed in Step 1 (Box et al., 2015). Box–Jenkins methods use differencing (i.e., subtracting values at t from those at $t + 1$), as well as moving average and autoregressive parameters to describe autocorrelation in time series and to address nonstationarity in the time-series. These well-described methods appear frequently in physical, social, and health sciences research (De Gooijer & Hyndman, 2006). This includes studies of cost-sharing in health care (Balasubramanian et al., 2015; Haddad & Fournier, 1995; Holder & Blose, 1992; Y.-C. Lee et al., 2006; I.-H. Lee et al., 2012; Moreno-Torres et al., 2011; Nelson et al., 1984; Petrou, 2015; Petrou & Ingleby, 2019; Puig-Junoy et al., 2014; Wolfson et al., 1982; Zechnich et al., 1998).

3. We combined the results of Steps 1 and 2 into a Box-Jenkins "transfer function" that estimates the counterfactual values of the treatment group (i.e., values expected if RP had not been introduced) from the observed values for the comparison group and from autocorrelation unique to the treatment group (Box et al., 2015). The residuals of a transfer function represent the degree to which the observed values of the treatment group differed from the counterfactual values. The general form of our transfer functions was as follows:

$$(1 - \varphi\beta^{p})Y_t = c + \omega X_t + (1 - \varphi\beta^{q})\varepsilon_t$$

where the term p indicates the number of preceding ("lagged") Y values that have to be added to or subtracted from Y in the model in order to capture local periods of growth and decline (i.e., autocorrelation) in the data. The term q represents the number of preceding values for the error term that have to be added to or subtracted from the model.

The backshift operator β is the value of Y at time t-p or of ε at time t-q. The parameter φ is the autoregressive parameter that measures the fraction of Y at time t remembered at time ($t + p$). Autoregressive parameters best express slow regression to expected values. The parameter θ is the moving average parameter, which measures the fraction of ε at time

t remembered at time $t + q$. Moving average parameters best express quick regression to expected values.

Y_t is the number of outpatient procedures performed in the treatment group in month t, c is a constant, and the term X_t represents the number of procedures performed in the comparison group. The coefficient ω expresses the response—if any—of Y to changes in X. The model residuals at time t are captured by ε_t.

1. In the final step in our statistical analysis, we use standard methods to determine if the residuals of the transfer function exhibited a significant drop at or near the month when RP went into effect (Chang et al., 1988). These methods essentially identify outliers that fall significantly (i.e., $p < .01$; single-tailed test) above or below counterfactual values and remain so until the end of the test period. We also use a two-tailed test for sensitivity purposes.

All analyses were conducted using SCA, Version 8.1 (Scientific Computing Association, River Forest, IL). This study was approved by the Institutional Review Board at the University of California.

Results

Participant characteristics in the treatment (CalPERS) and comparison groups are summarized in Table 1. Overall, the two groups were similar across all measured characteristics, although the CalPERs group was older and showed a slight female predominance relative to the comparison group.

For our analysis of hip and knee replacement surgeries, the study population was composed of 2,825 CalPERS beneficiaries who underwent at least one of these two procedures between 2008 and 2014, while the comparison group included 11,859 beneficiaries. For our analysis of outpatient surgeries (cataract removal surgeries, colonoscopies, and arthroscopic surgery of the knee and shoulder), the study population was composed of 50,875 CalPERS beneficiaries and the comparison group included 363,409 beneficiaries. We depict the number of hip and knee replacement surgeries performed each month for the CalPERS and comparison groups in Figure 1. We depict the number of cataract removals, colonoscopies, knee arthroscopies and shoulder arthroscopies in Figure 2. Across procedure types, there is seasonal variation apparent in the comparison group time-series, with more procedures performed at the end of each calendar year. We note no clear increases or decreases in the number of procedures performed in the CalPERS group either immediately before or after RP was implemented. We note no clear trends in the number of procedures performed in the CalPERS group throughout follow-up.

We examined the impact of RP on the utilization of each outpatient procedure type separately, identifying significant decreases (“level shifts”) in the number of procedures performed each month above or below the value expected given the past. We summarize coefficients for our final Box–Jenkins models in Table 2 and all level shifts detected for each procedure type in Table 3.

Hip and Knee Replacement Surgeries

The average number of surgeries performed each month was 34.5 in the CalPERS population and 144.6 in the comparison group. Coefficients for the final Box–Jenkins model are summarized in Table 2. Residuals for the model exhibited no autocorrelation. The estimated coefficient for the first autoregressive parameter ($\varphi = 0.24$) implies that high or low values of Y at month t persist into the month $t + 1$. We find evidence that 30 months after the implementation of RP in July of 2013, approximately 14.8% fewer hip and knee replacement surgeries were performed than expected among CalPERS beneficiaries (−6.7, $p < .01$). This level shift falls outside of our 3-month postimplementation time window and our 1-year postimplementation time window.

Cataract Removal Surgery

Throughout the study period, 3,891 cataract removal surgeries were performed among CalPERS beneficiaries, and 22,760 were performed in the comparison group. The average number of surgeries performed each month was 64.9 in the CalPERS population and 379.3 in the comparison group. Coefficients for the final Box–Jenkins model are summarized in Table 2. Residuals for the model exhibited no autocorrelation. We estimated the magnitude of the first autoregressive parameter as $\varphi = 0.30$, which implies that high or low values of Y at month t persist into the month $t + 1$. We detected no evidence of level shifts in the number of cataract removal surgeries performed from month-to-month in these data (Table 3).

Colonoscopies

Throughout the study period, 39,916 colonoscopies were performed among CalPERS beneficiaries, and 280,910 were performed in the comparison group. The average number of colonoscopies performed each month was 665.3 in the CalPERS population and 4,681.8 in the comparison group. Coefficients for the final Box–Jenkins model are summarized in Table 2. Residuals for the model exhibited no autocorrelation. The first autoregressive parameter ($\varphi = 0.16$) autoregressive parameter into the month $t + 1$. We found evidence that the number of colonos-copies performed decreased by 17.2%, approximately 129 procedures, relative to the expected value 8 months after implementation of RP in September of 2012 (−128.64, $p <.01$) and that the number of colonoscopies decreased again by 10.0%, 21 months after the implementation of RP in October of 2013 (−61.54, $p < .01$). These level shifts both fall outside of our 3-month postimplementation time window, but the level shift at 8 months falls within our 1-year postimplementation time window.

Arthroscopic Surgeries

Arthroscopic Knee Surgeries.—Throughout the study period, 4,769 arthroscopic knee surgeries were performed among CalPERS beneficiaries, and 41,205 were performed in the comparison group. The average number of surgeries performed each month was 79.5 in the CalPERS population and 686.8 in the comparison group. Coefficients for the final Box–Jenkins model are summarized in Table 2. Residuals for the model exhibited no autocorrelation. The estimated coefficient for the first autoregressive parameter ($\varphi = 0.13$) was not statistically significant. We find evidence of two significant shifts in the monthly number of arthroscopic knee surgeries performed. First, we find that 29 months prior to the implementation of RP, in June of 2009, the number of arthroscopic knee surgeries increased by approximately six procedures (18.4%) above its expected value (5.53, $p < .01$). Six months after the implementation of RP in June of 2012, we find that approximately 20.0% or 14 fewer procedures were performed than expected (−13.80, $p < .01$). This falls outside of our 3-month postimplementation time window but within our 1-year postimplementation time window.

Arthroscopic Shoulder Surgeries.—Throughout the study period, 2,299 arthroscopic shoulder surgeries were performed among CalPERS beneficiaries, and 18,534 were performed in the comparison group. The average number of surgeries performed each month was 38.3 in the CalPERS population and 308.9 in the comparison group. Coefficients for the final Box–Jenkins model are summarized in Table 2. Residuals for the model exhibited no autocorrelation. We find no evidence of level shifts in the number of arthroscopic shoulder surgeries performed either before or after the implementation of RP.

Note that using two-tailed tests rather than single-tailed tests does not alter any of the above findings. To determine the *p* value of the two-tailed test simply double the *p* value of the single-tailed test.

Discussion

The implementation of RP may potentially reduce health care utilization if patients perceive care at less expensive facilities to be inferior to care at more expensive facilities and are both unwilling to receive care at a less expensive facility and unwilling to pay a premium to receive care at a more expensive facility. Past research on surgical complication rates —one measure of quality—consistently finds that there are no statistical differences in complication rates due to RP (Brodke et al., 2019; Brown & Atal, 2019; Chi et al., 2014; Robinson, Brown, Whaley, & Bozic, 2015; Robinson, Brown, Whaley, & Finlayson, 2015).

For the present study, we examined the effect of the implementation of RP on surgical procedure utilization among CalPERS beneficiaries. We conducted a comparison-population interrupted time-series analysis, which is a type of ARIMA analysis, to identify whether any level shifts in the number of inpatient and outpatient surgeries occurred in response to the implementation of RP, while simultaneously accounting for any trends, cycles or seasonality in the data, including the autocorrelation patterns in the comparison group. To our knowledge, this is the first comparison-population interrupted time-series analysis examination of whether RP affects the frequency of service utilization among covered beneficiaries.

We find no evidence of significant increases or decreases in the number of procedures in the three months that followed implementation of RP. We do, however, identify two level shifts that occurred within 1 year of the implementation of the RP. We find that 14 (20%) fewer arthroscopic knee surgeries were performed than expected 6 months after RP was implemented in June of 2012, and we find that 129 (17.2%) fewer colonoscopies were performed than expected 8 months after RP was implemented in August of 2012. Importantly, both of these level shifts occurred prior to the open enrollment period for

CalPERS (September 5, 2012 to October 12, 2012) which makes it unlikely that the apparent decreases in procedures are attributable to beneficiaries switching to an alternative insurance product in response to RP. We also note that although statistically significant, the decreases in the number of arthroscopic knee surgeries and colonoscopies were quite modest. This suggests that while RP may have had some delayed impact on utilization of out-patient procedures, these effects were relatively minor.

Limitations

The present study is based on data from California. Because the study population is based in California—which is demographically distinct from other states– our findings will not necessarily generalize to other regions or states. This motivates continued examination of the effects of policies such as RP in varied geographic contexts in the United States. Second, our study data do not allow us to distinguish between changes in unnecessary and necessary care. It is possible that the implementation of policies such as RP will have different effects on utilization of services that are medically necessary versus those that are elective. However, we believe it is unlikely that the effect of RP on necessary and unnecessary service utilization was equal and opposite so as to yield our finding of no important change in service utilization within the relevant time windows.

We know of no concomitant reforms or events that could have influenced the frequency of the outpatient procedures among CalPERS beneficiaries. However, it is possible that our findings are influenced, at least in part, by unknown and unmeasured simultaneous changes to plan design, networks, pricing, quality reporting, and advertising that coincided with the implementation of RP that could affect outpatient procedure rates among CalPERS beneficiaries, the comparison group, or both. Although we suspect it is unlikely that individuals anticipating an elective procedure might switch into a more restrictive HMO plan, we cannot rule out changes in enrollment or the possibility that compositional effects lead to changes in the propensity of the CalPERS population to seek out and undergo elective procedures in the months that followed the implementation of RP.

Finally, we cannot know the number of persons clinicians judged in any month as candidates for the procedures we studied (i.e., the population at risk). The number of persons insured by either CalPERS or within the comparison group does not, moreover, necessarily “trace” the population at risk.

Conclusion

For the present study, we examined the effect of the implementation of RP on the utilization of selected inpatient and outpatient procedures among CalPERS PPO beneficiaries. We used a comparison-population interrupted time-series approach, an application of ARIMA, which identified significant increases or decreases (i.e., level shifts) in the number of procedures performed each month as compared with the expected value. Past research finds evidence that implementation of RP leads to substantial cost-savings, increased use of low-price facilities, and does not alter complication rates (Brodke et al., 2019; Brown & Atal, 2019; Chi et al., 2014; Robinson, Brown, Whaley, & Bozic, 2015; Robinson, Brown, Whaley, & Finlayson, 2015). We find limited evidence for only small decreases in utilization of some outpatient procedures several months after RP was implemented. Our findings motivate continued examination of the effects of RP on service utilization for a range of procedure types in varied settings.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This project was supported by grant number R01 HS022098 from the Agency for Healthcare Research and Quality (AHRQ).

References

Balasubramanian D, Prinja S, & Aggarwal AK (2015). Effect of user charges on secondary level surgical care utilization and out-of-pocket expenditures in Haryana State, India. PLOS ONE, 10(5), Article e0125202. 10.1371/journal.pone.0125202

Box GE, Jenkins GM, Reinsel GC, & Ljung GM (2015). Time series analysis: Forecasting and control. John Wiley.

(2019). Impact of reference pricing on cost and quality in total joint arthroplasty. Journal of Bone & Joint Surgery, 101(24), 2212–2218. 10.2106/JBJS.19.00475

Brown TT, & Atal JP (2019). How robust are reference pricing studies on outpatient medical procedures? Three different preprocessing techniques applied to difference-in differences. Health Economics, 28(2), 280–298. 10.1002/hec.3841 [PubMed: 30450623]

Brown TT, Guo C, & Whaley C (2018). Reference-based benefits for colonoscopy and arthroscopy: Large differences in patient payments across procedures but similar behavioral responses. Medical Care Research and Review, 77(3), 261–273. 10.1177/1077558718793325 [PubMed: 30103654]

Cassidy A (2014, July 24). Site-neutral payments. Health Affairs. http://www.healthaffairs.org/healthpolicybriefs/brief.php?brief_id=121

Catalano R, & Serxner S (1987). Time-series designs of potential interest to epidemiologists. American Journal of Epidemiology, 126(4), 724–731. 10.1093/oxfordjournals.aje.a114712 [PubMed: 3631061]

Chang I, Tiao GC, & Chen C (1988). Estimation of time series parameters in the presence of outliers. Technometrics, 30(2), 193–204. 10.1080/00401706.1988.10488367

Chi WC, Wu S, & DeVries A (2014). Value-based purchasing's effect on quality and costs. Health Affairs, 33(4), Article 723. 10.1377/hlthaff.2014.0145

De Gooijer JG, & Hyndman RJ (2006). 25 Years of time series forecasting. International Journal of Forecasting, 22(3), 443–473. 10.1016/j.ijforecast.2006.01.001

Gabel JR, Fahlman C, Kang R, Wozniak G, Kletke P, & Hay JW (2008). Where do I send thee? Does physician-ownership affect referral patterns to ambulatory surgery centers? Health Affairs, 27(Suppl. 1), w165–w174. 10.1377/hlthaff.27.3.w165 [PubMed: 18349040]

Haddad S, & Fournier P (1995). Quality, cost and utilization of health services in developing countries: A longitudinal study in Zaïre. Social Science & Medicine, 40(6), 743–753. 10.1016/0277-9536(94)00134-F [PubMed: 7747209]

Holder HD, & Blose JO (1992). The reduction of health care costs associated with alcoholism treatment: A 14-year longitudinal study. Journal of Studies on Alcohol, 53(4), 293–302. 10.15288/jsa.1992.53.293 [PubMed: 1619923]

Hollingsworth JM, Ye Z, Strope SA, Krein SL, Hollenbeck AT, & Hollenbeck BK (2010). Physician-ownership of ambulatory surgery centers linked to higher volume of surgeries. Health Affairs, 29(4), 683–689. 10.1377/hlthaff.2008.0567 [PubMed: 20368599]

Lee I-H, Bloor K, Hewitt C, & Maynard A (2012). The effects of new pricing and copayment schemes for pharmaceuticals in South Korea. Health Policy, 104(1), 40–49. 10.1016/j.healthpol.2011.09.003 [PubMed: 22000599]

Lee Y-C, Yang M-C, Huang Y-T, Liu C-H, & Chen S-B (2006). Impacts of cost containment strategies on pharmaceutical expenditures of the National Health Insurance in Taiwan, 1996–2003. PharmacoEconomics, 24(9), 891–902. 10.2165/00019053-200624090-00006 [PubMed: 16942123]

Moreno-Torres I, Puig-Junoy J, & Raya JM (2011). The impact of repeated cost containment policies on pharmaceutical expenditure: Experience in Spain. European Journal of Health Economics, 12(6), 563–573. 10.1007/s10198-010-0271-1

Nelson AA, Reeder CE, & Dickson WM (1984). The effect of a Medicaid drug copayment program on the utilization and cost of prescription services. Medical Care, 22(8), 724–736. 10.1097/00005650-198408000-00004 [PubMed: 6433121]

Petrou P (2015). An interrupted time-series analysis to assess impact of introduction of co-payment on emergency room visits in Cyprus. Applied Health Economics and Health Policy, 13(5), 515–523. 10.1007/s40258-015-0169-2 [PubMed: 25894739]

Petrou P, & Ingleby D (2019). Co-payments for emergency department visits: A quasiexperimental study. Public Health, 169(1), 50–58. 10.1016/j.puhe.2018.12.014 [PubMed: 30818104]

Puig-Junoy J, Rodríguez-Feijoó S, & Lopez-Valcarcel BG (2014). Paying for formerly free medicines in Spain after 1 year of co-payment: Changes in the number of dispensed prescriptions. Applied Health Economics and Health Policy, 12(3), 279–287. 10.1007/s40258-014-0097-6 [PubMed: 24696429]

Robinson JC, & Brown TT (2013). Increases in consumer cost sharing redirect patient volumes and reduce hospital prices for orthopedic surgery. Health Affairs, 32(8), 1392–1397. 10.1377/hlthaff.2013.0188 [PubMed: 23918483]

Robinson JC, Brown TT, & Whaley C (2015). Reference-based benefit design changes consumers' choices and employers' payments for ambulatory surgery. Health Affairs, 34(3), 415–422. 10.1377/hlthaff.2014.1198 [PubMed: 25732491]

Robinson JC, Brown TT, & Whaley C (2017). Reference pricing changes the "choice architecture" of health care for consumers. Health Affairs, 36(3), 524–530. 10.1377/hlthaff.2016.1256 [PubMed: 28264955]

Robinson JC, Brown TT, Whaley C, & Bozic KJ (2015). Consumer choice between hospital-based and freestanding facilities for arthroscopy. Journal of Bone & Joint Surgery, 97(18), 1473–1481. 10.2106/JBJS.O.00240

Robinson JC, Brown TT, Whaley C, & Finlayson E (2015). Association of reference payment for colonoscopy with consumer choices, insurer spending, and procedural complications. JAMA Internal Medicine, 175(11), 1783–1789. 10.1001/jamainternmed.2015.4588 [PubMed: 26348851]

Robinson JC, Pozen A, Tseng S, & Bozic KJ (2012). Variability in costs associated with total hip and knee replacement implants. Journal of Bone & Joint Surgery, 94(18), 1693–1698. 10.2106/JBJS.K.00355

Whaley CM, Guo C, & Brown TT (2017). The moral hazard effects of consumer responses to targeted cost-sharing. Journal of Health Economics, 56(December), 201–221. 10.1016/j.jhealeco.2017.09.012 [PubMed: 29111500]

White C, Reschovsky JD, & Bond AM (2014). Understanding differences between high- and low-price hospitals: Implications for efforts to rein in costs. Health Affairs, 33(2), 324–331. 10.1377/hlthaff.2013.0747 [PubMed: 24476706]

Wolfson J, Kapadia AS, Decker M, Sear AM, & Roht LH (1982). Effects of cost-sharing on users of a state's health service program. Medical Care, 20(12), 1178–1187. 10.1097/00005650-198212000-00004 [PubMed: 6217386]

Zechnich AD, Greenlick M, Haxby D, & Mullooly J (1998). Elimination of over-the-counter medication coverage in the Oregon Medicaid population: The Impact on program costs and drug use. Medical Care, 36(8), 1283–1294. 10.1097/00005650-199808000-00015 [PubMed: 9708599]

Zhang H, Cowling DW, & Facer M (2017). Comparing the effects of reference pricing and centers-of-excellence approaches to value-based benefit design. Health Affairs, 36(12), 2094–2101. 10.1377/hlthaff.2017.0563 [PubMed: 29200355]

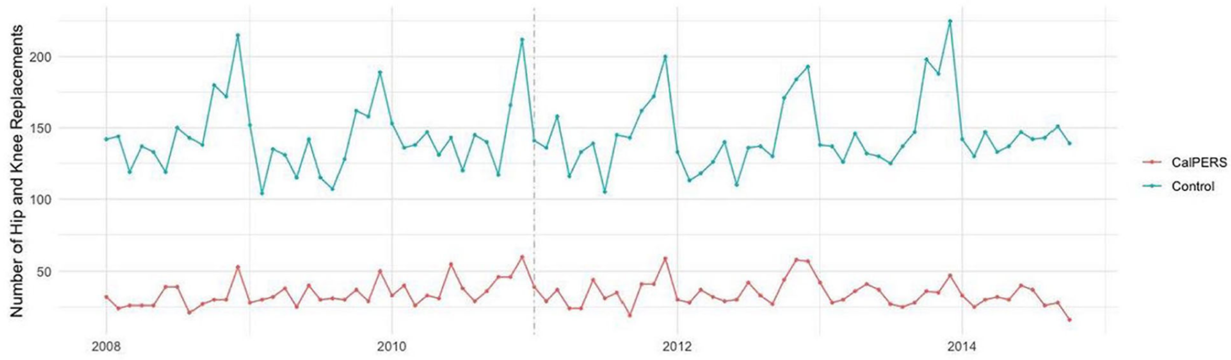

Figure 1.

Number of hip and knee replacements performed each month from January 2008 through December 2014 among CalPERS beneficiaries and a comparison group of non-CalPERS beneficiaries who were not subject to RP but were covered by PPO products issued by the same insurance carrier.

Note. Hip and knee replacements were subject to RP beginning January 1, 2011 depicted with the dashed vertical line. CalPERS = California Public Employees' Retirement System; RP = reference pricing; PPO = Preferred Provider Organization.

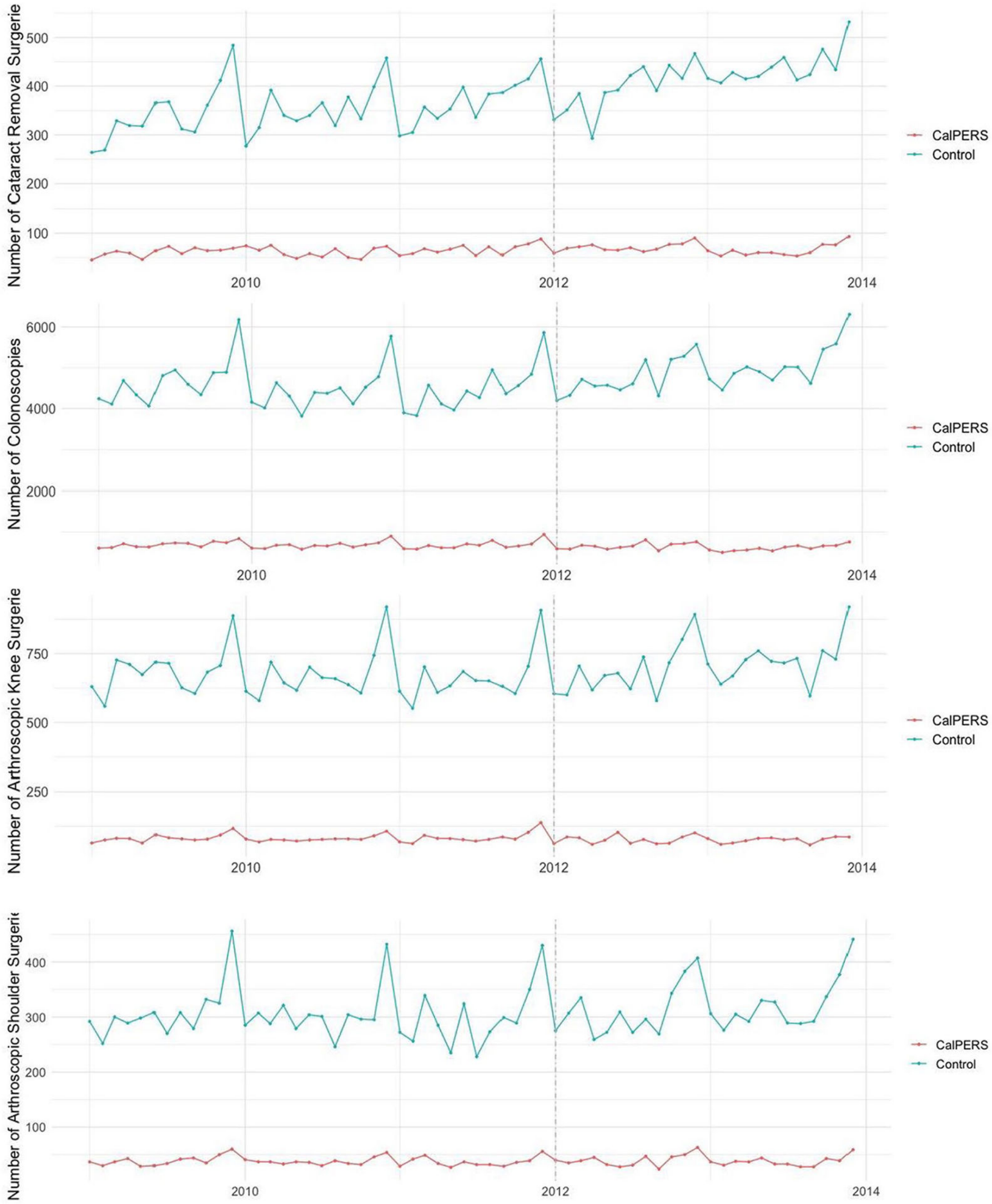

Figure 2.

Number of cataract removal surgeries, colonoscopies, knee arthroscopies, and shoulder arthroscopies performed each month from January 2009 through December 2014 among CalPERS beneficiaries and a comparison group of non-CalPERS beneficiaries who were not subject to RP but were covered by PPO products issued by the same insurance carrier.

Note. Cataract removal surgeries, colonoscopies, knee arthroscopies, and shoulder arthroscopies were subject to RP beginning January 1, 2012 depicted with the dashed vertical line. CalPERS = California Public Employees' Retirement System; RP = reference pricing; PPO = Preferred Provider Organization.

Table 1.

Characteristics of CalPERS Beneficiaries and the Comparison Group.

	CalPERS beneficiaries	**Comparison group**
Hip and knee replacement		
Numbers of beneficiaries	2,825	11,859
Average total price (*SD*)	$28,706.43 (16,864.38)	$29,705.19 (18,430.56)
Mean age (*SD*)	57.9 (5.8)	56.7 (6.3)
Proportion female (*SD*)	0.60 (0.49)	0.52 (0.50)
Charlson comorbidity index (*SD*)	0.32 (0.63)	0.27 (0.65)
Cataract removal surgery		
Numbers of beneficiaries	3,891	22,760
Average total price (*SD*)	$3,509.81 (3,340.21)	$3,329.52 (3,719.08)
Mean age (*SD*)	59.1 (5.1)	57.7 (6.3)
Proportion female (*SD*)	0.60 (0.50)	0.54 (0.49)
Charlson comorbidity index (*SD*)	0.11 (0.38)	0.11 (0.38)
Colonoscopy		
Numbers of beneficiaries	39,916	280,910
Average total price (*SD*)	$1,955.94 (2,638.94)	$2,123.75 (3,448.11)
Mean age (*SD*)	54.7 (7.7)	52.7 (8.7)
Proportion female (SD)	0.56 (0.50)	0.53 (0.50)
Charlson comorbidity index (*SD*)	0.11 (0.40)	0.10 (0.39)
Arthroscopic surgery		
Numbers of beneficiaries	7,068	59,739
Average total price (*SD*)	$7,948.34 (9,894.85)	$7,852.30 (10,901.47)
Mean age (*SD*)	51.1 (11.5)	47.2 (12.5)
Proportion female (*SD*)	0.49 (0.50)	0.42 (0.49)
Charlson comorbidity index (*SD*)	0.07 (0.31)	0.06 (0.27)

Note. CalPERS = California Public Employees' Retirement System.

Table 2.

Transfer Function Parameters for Procedures Among Insured Persons Subjected to RP (n = 60 Months).[a,b,c]

	Constant			ω for comparison group			Autoregressive parameters[d]		
	C	***SE***	***p***	X_t	***SE***	***p***	**φ**	***SE***	***p***
Cataract removal	31.53	8.66	.023	0.09	0.02	.016	0.30	0.13	.078
Colonoscopy	−42.56	39.87	.33	0.16	0.01	<.00001	0.42	0.12	.018
Knee arthroscopy	−19.72	8.55	.06	0.14	0.01	.00002	0.05	0.13	.75
Shoulder arthroscopy	−2.02	4.93	.70	0.13	0.01	.001	—	—	—
Hip and knee replacement	1.53	4.93	.77	0.024	0.03	.002	0.23	0.11	.11

Note. RP = reference pricing; SE = standard error.

[a] RP was implemented in January of 2011 for hip and knee replacements and in January of 2012 for all other outpatient procedure types (cataract removal surgeries, colonoscopies, knee and shoulder arthroscopy).

[b] We identified level shifts using methods described by Chang et al. (1988) in which the number of procedures performed in a given month increased or decreased significantly (i.e., $p < .01$; single-tailed test) relative to values expected given the past. To compute the p value for a two-tailed test, simply double the p value of the single-tailed test in the table.

[c] No moving parameters were identified for any of the time-series included in our analysis.

[d] All autoregressive parameters are for ($t - 1$)

Table 3.

Detected Level Shifts in the Number of Hip and Knee Replacement Surgeries, Cataract Removal Surgeries, Colonoscopies, and Arthroscopic Surgeries of the Knee and Shoulder Performed Among CalPERS Beneficiaries.

Procedure type	Month of level shift	Time elapsed since implementation of RP[a]	Estimate (t value)[b]
Hip and knee replacement surgeries	July 2013	30 Months after	−6.70 (−3.02)
Cataract removal surgeries	—	—	—
Colonoscopies	August 2012	8 Months after	−128.64 (−8.75)
	October 2013	21 Months after	−61.54 (−2.19)
Knee arthroscopy	June 2009	29 Months prior	5.53 (3.99)
	June 2012	6 Months after	−13.80 (−5.59)
Shoulder arthroscopy	—	—	—

Note. CalPERS = California Public Employees' Retirement System; RP = reference pricing.

[a] RP was implemented in January of 2011 for hip and knee replacements and in January of 2012 for all other outpatient procedure types (cataract removal surgeries, colonoscopies, knee and shoulder arthroscopy).

[b] We identified level shifts using methods described by Chiang et al. (1988) in which the number of procedures performed in a given month increased or decreased significantly (i.e., $p < .01$; single-tailed test) relative to values expected given the past. To compute the p value for a two-tailed test, simply double the p-value of the single-tailed test in the table.