Saving Lives or Saving Money?
Understanding the Dual Nature of Physician Preferences†

Alice Chen* and Darius Lakdawalla**

July, 2015

Abstract

A longstanding literature has highlighted the tension between the altruism of physicians and their desire for profit. This paper develops new implications for how these forces drive pricing and utilization outcomes in healthcare markets. Altruism dictates that providers reduce utilization in response to higher prices, but profit-maximization does the opposite. Rational physicians will behave more altruistically towards poorer, vulnerable patients, and when the financial costs of altruism are lower. These insights help explain the observed heterogeneity in pricing dynamics across different healthcare markets. We empirically test the implications of our model by utilizing two exogenous shocks in Medicare price setting policies. Our results demonstrate that uniform policy changes in reimbursement or patient cost-sharing may not generate the intended responses on quantity.

* University of Southern California, Sol Price School of Public Policy. alicejc@price.usc.edu.
** University of Southern California, School of Pharmacy and Sol Price School of Public Policy; and NBER. dlakdawa@usc.edu.

† We gratefully acknowledge comments from participants at the BU-Harvard-MIT Health Seminar, the USC CESR-Schaeffer Brown Bag, and the Midwest Health Economics Conference.

I. Introduction

Economists have long emphasized the peculiarities of healthcare markets, compared to other markets for goods and services. Since at least Kenneth Arrow's pioneering paper on the subject, economists have recognized two features in particular: the altruism of healthcare providers towards their patients, and the reliance of patients on their physicians for information and guidance (Arrow 1963). Less attention has been paid to the market pricing and utilization implications of these well-known insights into physician behavior .

Altruism encourages physicians to represent the interests of their patients. For example, an altruistic physician will tend to economize on the use of scarce inputs and attempt to maximize the utility of patients subject to their own resource constraints. However, the informational advantage of physicians creates a classic agency problem that physicians might exploit to pursue their own interests instead of their patients' interests. (See, for example, Blomqvist, 1991; Dranove and White 1987; Emanual and Emanual, 1992; Mooney and Ryan, 1993; Lu, 1999; and Zweifel and Breyer, 1997).

Figure 1 provides some initial insight into the importance of this issue. The figure depicts the histogram of price elasticities within Medicare services that experienced an approximately 50% increase in annual physician reimbursement rates . Such large and sudden changes are unlikely to reflect changes in demand, but instead more likely to reflect movements along a demand curve. Price increases coincided with increased quantity in half the cases depicted, but with decreased quantity in the other half.[1] This pattern is difficult to explain when relying on either agency problems or altruism alone. Physicians exploiting agency problems would drive utilization higher in the event of higher reimbursement. Physicians behaving altruistically would drive utilization lower in the event of higher prices, part of which are borne by patients. What is needed is a unified approach that considers how altruism and agency problems interact to drive healthcare

[1]The large majority of the procedures depicted in Figure 1 are major or minor procedures. Very few are lab test, imaging, or evaluation and management services.

pricing and utilization. Significant policy questions are at stake, since price is often viewed as an important lever for influencing behavior. In some market segments, for example, higher physician reimbursements can be expected to curb utilization, while in others, the opposite effect obtains.

In this paper, we study how altruism and agency problems compete to influence price and utilization, along with the positive and normative implications of this competition. We rely on well-established models of physician behavior, but apply these to problems of pricing and utilization that have not been studied through the lens of physician preferences. From a positive standpoint, we show that exogenous price changes may increase or decrease quantity supplied. When higher prices lower quantity, we say dynamics are primarily "patient-driven," and when the opposite is true, we say they are primarily "physician driven." Moreover, the degree to which markets are patient-driven or physician-driven endogenously depends on physician incentives. Specifically, pricing is more likely to be patient-driven when patients are poorer, and when healthcare provision is less profitable. In other words, physician altruism is more likely to win out when the value of behaving altruistically is higher and the cost is lower.

From a policy standpoint, patient-driven behavior limits the potential for overuse of healthcare resources, while physician-driven behavior exacerbates it. Thus, we expect less overuse, from the consumer's perspective, when consumers are poorer, patient cost-sharing is higher, input prices are lower, and profitability is lower. Increases in patient wealth, therefore, are expected to increase "waste" in healthcare, as are expansions in the availability of insurance.

Empirically, we test the conjectures of our model by using two exogenous policy shocks to Medicare payments: the 1997 consolidation of geographic payment regions and the 1999 change in estimation of practice expenses. Our results indicate that the size and sign of the own-price elasticity does vary when there is joint decision making between patients and physicians. We also show that procedures are more likely to follow patient-driven pricing behavior when patient income is lower, patient cost-sharing is higher, and the physician's price-cost margin is lower. We use these findings illustrate why uniform changes in payment or cost-sharing may not generate the intended responses in quantity.

Our main contribution is offering a theory that can identify when quantity varies positively versus negatively with price. The literature thus far has offered piecemeal explanations of the observed heterogeneity in response to price changes. Some empirical studies observe that higher reimbursements will lead to increased utilization, and the accompanying theory relies on physicians being profit maximizers (Clemens and Gottlieb, 2003; Gruber et al., 1999; and Jacobson et al., 2006). Other empirical studies show that there is a negative relationship between price and quantity (Escarce, 1993; Nguyen and Derrick, 1997; Rice, 1983; and Yip, 1998). Theories used to explain a negative price-quantity relationship include models of physician induced demand and non-fee-for-service reimbursement schemes. For example, Farley (1986) discusses implications of the target-income model. Ellis and McGuire (1986) demonstrate that having a prospective-payment system can lead to too few services being provided if physicians undervalue the benefits of patients relative to hospital profits, and Choné and Ma (2011) and Glied and Zivin (2002) discuss how managed care can restrict quantity. Finally, some studies find a low responsiveness between quantity and price, and they conclude that there is uncertainty in a physician's objective function (Holohan, 1977; Hurley et al., 1990; and Hurley and Labelle, 1995).

We unify these findings by offering a simple modification to the existing theory. Unlike prior studies such as Ellis and McGuire (1986), Ellis and McGuire (1990), and Liu and Ma (2013), our model allows physicians to care about patient health and patient spending. This extension generates new insights on when services are likely to be patient-driven versus physician-driven. Our theory highlights that certain demand-side policies may be just as effective as supply-side policies in controlling costs.[2] This work relates to Dickstein (2014), who empirically quantifies the contributions of patient and physician incentives to prescription drug utilization.

[2] While policymakers have traditionally focused on controlling Medicare expenditures by altering Medicare payments, demand-side policies, such as changes to patient cost-sharing and supplemental insurance, are currently being debated (Gruber, 2013; National Commission on Fiscal Responsibility and Reform, 2010; and Zuckerman et al., 2010).

The rest of the paper is organized as follows. In Section 2, we propose a theoretical framework for the joint decision-making between patients and physicians, and we derive the normative implications from our model. In section 3, we discuss the empirical approach for testing conjectures derived from our model. In Section 4, we present the empirical results, and Section 5 concludes.

2. Theoretical framework

Physician altruism and joint patient-physician decision making create unique relationships among pricing, utilization, and other economic forces. We demonstrate these points in a simple and standard theoretical model that traces back to Becker (1957), and has been used by a number of health economists to study physician behavior (e.g., Ellis and McGuire 1986; Ellis and McGuire 1990; McGuire 2000; McGuire and Pauly, 1991)

2.1. Simple illustration

For pedagogical purposes, we first illustrate in a very simple, perfectly competitive model how physician and patient decisions interact. Here, we presume that physicians earn zero economic profits, and patients bear the full cost of healthcare.

Suppose health is produced using a good or procedure X, according to $F(X)$, where $F_{XX} < 0$. We also suppose that this good is initially health-improving, but eventually health-reducing if overused. In this context, suppose for simplicity, that a fully informed representative patient maximizes the value of health net of the cost of production. This results in the following household production function for health:

$$\max_X vF(X) - p_X X$$

It is straightforward to show in this context that the derived demand for X is falling in price p_X, as in $\frac{\partial x}{\partial p_x} = \frac{1}{vF_{xx}} < 0$.

Now, however, suppose that the representative patient is not fully informed but instead receives care from a physician, who bears cost $c(X)$, where $c_{XX} \geq 0$. The latter assumption rules out cost complementarities between the procedure and its substitute.

The physician maximizes a weighted average of patient well-being and her own income, as in:

$$\max_X (1-\alpha)[p_X X + -c(X)] + \alpha[vF(X) - p_X X]$$

The parameter α is an index of altruism. With relatively minor modifications, it can also be thought of equivalently as the patient's relative bargaining leverage in a Nash-bargaining problem between patient's and physicians.

Observe that the physician's objective function can be rewritten as:

$$\max_X \alpha vF(X) - (1-\alpha)c(X) + (1-2\alpha)[p_X X]$$

This has the following first-order condition:

$$\alpha vF_X - (1-\alpha)c_X + (1-2\alpha)p_X = 0$$

Define D as the second derivative for this maximization problem. This allows us to write the comparative static of the problem as:

$$\frac{\partial X}{\partial p_X} = \frac{(2\alpha - 1)}{D}$$

If the problem is strictly concave at the optimum, then $D < 0$. As a result, if $\alpha > \frac{1}{2}$, the own-price elasticity is negative, because the physician is sufficiently altruistic to weight patient preferences enough that her decision problem resembles that of the fully informed patient. We call these "patient-driven pricing dynamics." If, on the other hand, $\alpha < \frac{1}{2}$, the opposite dynamics prevail: the own-price elasticity is positive. We call these "physician-driven pricing dynamics."[3]

2.2 General model

The derivation above assumed physicians and patients are risk-neutral over consumption. It also abstracted from the existence of health insurance. To generalize the simple model, suppose the representative patient derives utility $u(I + vF(X) - \pi - \sigma(X; p_X))$, where I is income, π is an ex ante insurance premium, and σ represents patient out-of-pocket expenditures. Here and elsewhere, we abstract from effects of physician

[3] This model can be easily extended to identify the effects of a substitute good or procedure. When the own-price elasticity is positive, the cross-price elasticity is negative, and vice versa.

decision-making on the insurance premium. This assumption sacrifices little generality in a public insurance scheme or when studying a relatively small set of procedures.

Now suppose physicians derive utility from a weighted average of patient utility and their own utility over consumption, $z(\cdot)$, where u and z are weakly concave utility functions. Assume the utility functions satisfy the assumptions of monotonicity, risk-aversion, and weak prudence, as in $z' > 0$, $z'' < 0$, and $z''' \geq 0$ (Felder & Mayrhofer, 2011). The generalized physician objective function can then be written as:

$$\max_X (1-\alpha)z(p_X X - c(X)) + \alpha u(I + vF(X) - \pi - \sigma(X; p_X))$$

The first-order conditions now become:

$$\alpha u' * (vF_X - \sigma_X) + (1-\alpha)z' * (p_X - c_X) = 0$$

The optimality conditions are weighted averages of physician profit-maximization and patient utility-maximization.

To simplify the analysis, we follow the convention adopted in much of the insurance literature and abstract from the direct income effects associated with patient out-of-pocket payments (Lakdawalla & Sood, 2013). This amounts to holding u' fixed when prices change. The comparative static now become:

$$\frac{\partial X}{\partial P_X} = \frac{\left(\alpha u' \sigma_{Xp_X} - (1-\alpha)z' - (1-\alpha)z''(p_X - c_X)X\right)}{D}$$

This comparative static suggests a simple empirical test for the presence of physician altruism. Absent altruism, own-price elasticities will always be positive. To see this, observe from the physician's optimality condition that $p_X = c_X$ in the absence of physician altruism. Therefore, without altruism, it follows that $\frac{\partial X}{\partial p_X} = -\frac{(1-\alpha)z'}{D} > 0$. If price increases lead to quantity reductions in real-world data, this necessarily signals the presence of altruism.

To develop further the implications of the comparative statics, note that pricing dynamics are patient-driven if $\alpha u' \sigma_{Xp_X} > (1-\alpha)z' + (1-\alpha)z''(p_X - c_X)X$. We now show that the right-hand side of this inequality is always positive and decreasing in physician consumption. To do so, we require the relatively weak assumption that $vF_X \geq \sigma_X$ at the optimum. That is, patients will not accept an equilibrium in which their own marginal out-of-pocket cost exceeds their own private marginal benefit. The asymmetry of information means this is not a trivial assumption, but – at least for insured consumers -- it would only be violated in fairly extreme cases of overuse. This assumption, coupled with the first-order condition for X, implies that $p_X \leq c_X$. Thus, $(1-\alpha)z''(p_X X - c_X X)$ has (weakly) the same sign as $(1-\alpha)z'$. This result, coupled with the prudence assumption ($z''' \geq 0$), implies that increases in income will reduce the term $(1-\alpha)z' + (1-\alpha)z''(p_X X - c_X X)$.

With this condition in place, we can conclude that pricing dynamics are more likely to be patient-driven if:

1. Physician altruism is higher – i.e., α is higher;
2. Patient income is lower – i.e., I is lower and u' is higher;
3. Patient health care spending is higher – i.e., $\pi + \sigma$ is higher, and u' higher;
4. Physician income is higher – which implies that z' and $z''(p_X - c_X)$ are lower;
5. The physician's price-cost margin, $p_X - c_X$, is lower.

Intuitively, pricing is more likely to be patient-driven if: physicians care more about their patients (#1); patients are more sensitive to spending growth (#2 and #3); physicians are richer and willing to pay more to purchase patient welfare (#4); and the opportunity cost to physicians of boosting utilization is lower (#5).

One final corollary is worth noting. Pricing can only be patient-driven if patients bear some portion of the financial cost. When cost-sharing is zero, the inequality $\alpha u' \sigma_{X p_X} > (1-\alpha)z' + (1-\alpha)z''(p_X - c_X)X$ is never satisfied. Intuitively, patients only care about price increases when they bear some of the financial burden.

2.3. Policy implications

The simple model above has several implications for patient welfare. Even in the presence of physician altruism, Pareto-efficiency continues to be characterized by the standard input efficiency conditions, $vF_X = c_X$. Thus, we can characterize the degree of inefficient overuse by quantifying $c_X - vF_X$. By inspecting the first-order conditions for physician decisionmaking, we can derive:

$$c_X - vF_X = \frac{\alpha u'}{\alpha u' + (1-\alpha)z'} \overbrace{(c_X - \sigma_X)}^{Moral\ hazard} + \frac{(1-\alpha)z'}{\alpha u' + (1-\alpha)z'} \overbrace{(p_X - vF_X)}^{Over-reimbursement}$$

There are two sources of input overuse. The first is moral hazard, which obtains when social marginal costs exceed out-of-pocket costs. The second is over-reimbursement, which obtains when physician reimbursements exceed the marginal value of care to patients. (If marginal value of care exceeds pricing, this tends toward underuse instead.) The overall degree of input inefficiency is the weighted average of these two sources, with the weights given by the relative importance of patient versus physician consumption. If physicians are perfectly altruistic, the over-reimbursement effect vanishes. On the other hand, if they are perfectly self-interested, the moral hazard effect vanishes. In addition, note that increases in physician consumption levels place more weight on the moral hazard effect, because richer physicians place more value on their patients' consumption than their own.

The relative importance of physician versus patient consumption has implications for which policy levers are most efficient at reducing distortions. If the degree of altruism is high, reimbursements reforms aimed at patients will be relatively more effective. If low, on the other hand, reforms aimed at physician reimbursement will be correspondingly more effective. Put differently, policymakers should focus more on moral hazard in patient-driven markets, but on physician reimbursement in physician-driven markets. More formally, holding all patient and physician incentives constant, reimbursement reforms that compress $p_X - vF_X$ will contribute less to efficiency when $\alpha u' > (1-\alpha)z'$, and vice-versa.

Our analysis also has implications for global reimbursement reforms that affect many markets or procedures at once. The effect of price on quantity may be positive or negative. Thus, uniform reimbursement changes – either global increases or global decreases in price – may have unintended consequences that depend on the mix of patient-driven versus physician-driven markets or procedures. Targeted reforms that change reimbursement for some markets, but not for others, might be dramatically more effective. We return to this point in the empirical analysis.

3. Empirical Analysis

The theoretical analysis generates at least five testable implications:

1. Both the size and even the sign of the price elasticity may vary when there is joint decision making between patients and physicians.
2. When patient income is lower, price elasticities are more likely to reflect patient-driven pricing behavior.
3. When patient cost-sharing is higher, price elasticities are more likely to reflect patient-driven pricing behavior.
4. When the physician's price-cost margin, $p_X - c_X$, is lower, price elasticities are more likely to reflect patient-driven pricing behavior.
5. Physician payment reforms have a larger effect in market segments where pricing is physician-driven than elsewhere.

3.1. Data

To test these implications, we rely on data from 1993 to 2002 from the Center for Medicaid and Medicare (CMS) Medicare Carrier Claims File (CCF) and the Medicare Current Beneficiary Survey (MCBS). The CCF data contains the fee-for-service Physician/Supplier Part B claims for a random 5% sampling of Medicare enrollees. For each service provided, we have information on the price, including the co-pay, deductible, physician submitted charge, and Medicare allowed amount.[4] All prices are converted to 2010 dollars using the Current Price Index for medical expenditures. The CCF also provides information on patient diagnoses and basic demographics, such as age, race, and gender. We define a market-area using the Dartmouth Atlas' Hospital Referral Region (HRR) and collapse the data to the HCPCS-HRR-year level.[5]

The MCBS data consists of a smaller, but still nationally representative, sample of 12,100 Medicare beneficiaries. By combining patient surveys with administrative payment files, the MCBS data provides a richer set of covariates. This feature allows us to consider additional patient characteristics, such as income and education. Due to MCBS' small

[4] The submitted charge is the amount physicians bill Medicare. The allowed amount is what Medicare actually pays for the procedure, which is described in Section 3.2.
[5] An alternative to the HRR level is to use either the pre-1997 CMS geographic regions or counties to identify market areas. Results are similar when using both alternative measures.

sample size, we rely on price and quantity data from CCF and use the MCBS only as a supplement the CCF data.

3.2. Medicare Payments and Policy Shocks

For each HCPCS, CMS calculates a payment based on three factors: (1) a relative value unit (RVU), (2) a geographic adjustment factor (GAF), and (3) a conversion factor (CF).[6] RVUs are procedure specific, and they reflect differences in the time, skill, training, and costs required to perform different procedures. GAFs are region-specific, so they account for geographic variation in the cost of providing services.[7] Finally, the CF is a nationally uniform adjustment factor that converts RVUs into a dollar amount. This factor is updated annually by CMS according to a formula specified by statute, but Congress can and has overridden the statutorily defined formula.[8]

To measure price elasticities, we need to identify payment changes within a market that are independent of patient demand, technological change, and supply. First, we consider changes to the overall Medicare payment rate, which will include variation from RVUs, GAFs, and the CF. Since GAFs are set across several different markets and CF is one number set nationally, these two components of Medicare pricing are likely exogenous to the dynamics within any one given market. However, variation in RVUs may not be exogenous within a market over time. At least once every five years, about 138 physicians from the Specialty Society Relative Value Committee (RUC) and its advisory committee convene to re-evaluate and assign RVUs. Their main objective is to adjust the work component of RVUs to reflect procedural differences in physician time, skill, and training. If adjustments in RVUs are systematically correlated with demand for a procedure, then price elasticity estimates based on RVU variation may be biased.

---

[6] The exact formula for calculating Medicare payments is given by

$$Pay = [RVU_W GPCI_W + RVU_{PE} GPCI_{PE} + RVU_{MP} GPCI_{MP}] \times CF,$$

where W indexes the work component, PE indexes the practice expense component, and MP indexes the malpractice expense component. GPCI represents the geographic practice cost indices, and CF is the conversion factor.

[7] GAF is a weighted sum of the work, practice expense, and malpractice GPCIs. Details can be found in MaCurdy et al. (2012).

[8] The CF in 2013 was $36.61 per RVU. Congress has overridden this formula in 1998, 2009, and 2011.

While changes in work RVUs may be non-random in theory, the practical case for bias is less clear. The assignment of relative weight is complex and political with battle lines and alliances drawn between specialties (Eaton, 2010). Deliberations are complicated by the fact that the size of the Medicare payment pie is fixed. As such, the final weights have been viewed as being somewhat arbitrary. For example, after the first major review of RVUs, the Health Care Financing Administration (HCFA) received "voluminous identical comments from family practitioners stating that [the HCFA had…] used an arbitrary method for revising the work RVUs" (Department of Health and Human Services, 1996).[9]

Nonetheless, to address potential endogeneity, we rely on policy shocks in Medicare pricing as instruments for the overall Medicare payment rate. The first major policy shock occurred in 1997 when HCFA consolidated the number of geographic payment regions from 210 distinct payment regions to only 89 distinct regions in 1997. Discussed in Clemens and Gottlieb (2014), this consolidation generated differential price shocks across county groupings within a state. While some states were unaffected by this policy, in about 26 states, the variation in reimbursement rates across counties was either significantly reduced or eliminated because multiple regions were collapsed into one single payment area.

While the first policy shocks affected payments across geographic areas, the second major policy shock created differential changes across services. Prior to 1997, practice expense RVUs (PE-RVUs) were measured using prevailing charges. However, Section 121 of the Social Security Amendments of 1994and the Balanced Budget Act of 1997 mandated that PE-RVUs be determined by relative costs, instead of prevailing charges. Phased in over a four-year period from 1999 to 2002, the modified PE-RVU calculations, also better differentiated between the costs of performing a procedure in a facility setting—such as a

[9] Between 199 to 2002, work RVUs experienced two major reviews which became effective in 1997 and 2002. The change in average work RVU is depicted in Appendix Figure 1.

hospital, skilled nursing facility, or ambulatory surgical center—and a non-facility setting, such an office or clinic.[10]

We exploit these two policy shocks to generate exogenous variation in Medicare reimbursements that is arguably unrelated to the local demand for and supply of services. Variation in these components over time is depicted in Figure 2. Using data from Federal Register reports, plot (a) depicts the change in GAF among counties that were affected by the 1997 consolidation versus those that were unaffected. It is clear that much of the pre-1997 differentiation across counties was eliminated post-1997. Plot (b) shows the change in average facility and non-facility PE-RVUs across HCPCS over time. While the transition from charge- to resource-based estimations was phased in over a four-year period, the differentiation between facility and non-facility RVUs created a large drop in average PE-RVUs over time. As Appendix Figure 1 depicts, much of the observed drop in PE-RVUs in 1999 comes from changes in the non-facility estimates.

3.3. Empirical Approach

In our baseline specification, we estimate the following equation for each HCPCS:

$$\log(\mathrm{Q}_{\mathrm{iht}}) = \beta^{\mathrm{i}} \log(\mathrm{P}_{\mathrm{iht}}) + \Gamma^{\mathrm{i}}\mathrm{X}_{\mathrm{iht}} + \gamma_{\mathrm{h}}^{\mathrm{i}} + \eta_{\mathrm{t}}^{\mathrm{i}} + \xi_{\mathrm{h}}^{\mathrm{i}}\mathrm{t} + \epsilon_{\mathrm{iht}}. \quad (1)$$

$\mathrm{Q}_{\mathrm{ict}}$ is the count of claims recorded for HCPCS i in HRR h in year t. $\mathrm{P}_{\mathrm{ict}}$ measures the allowed Medicare payment for the service. $\mathrm{X}_{\mathrm{ict}}$ are county-specific determinants of quantity that change over time, including the Charlson Comorbidity Index (CCI) calculated according to Quan et al. (2005), beneficiary age, Black and Hispanic dummies, and gender. γ_h are market fixed effects, η_t are year fixed effects, $\xi_{\mathrm{h}}^{\mathrm{i}}\mathrm{t}$ is a market by year time trend, and ϵ_{iht} is an idiosyncratic error term. Robust standard errors are clustered by HRR. Assuming that variation in Medicare payments for a specific HCPCS within a given market over time is plausibly exogenous to other unobserved changes in local health demand and supply, the β^i estimate denotes the own-price elasticity of HCPCS i.

[10] Prior to 1999, the non-facility PE-RVU was simply 50% if the facility PE-RVU (Maxwell and Zuckerman, 2007).

This assumption of exogeneity fails when, for example, changes in Medicare payments are correlated with the likelihood of performing a given procedure. Given the political nature of RVU changes, more popular procedures may draw a higher Medicare payment increase. Alternatively, the exogeneity assumption fails when changes in payments reflect changes in cost of performing a given procedure. Although CMS uses the decennial census to determine certain indices, such as employee wage indices, it also uses the most recent retrospective data to determine other indices, such as office rental expenses. If costs are serially correlated, then changes in overall payment may be correlated with changes in costs. Furthermore, CMS updates RVUs based on comments submitted by physicians, health care workers, and professional associations and societies, increasing the likelihood of payment changes being correlated with other local supply factors (Federal Register).

We relax our assumption of exogeneity by relying on the two policy shocks as instruments. In our main specification, we consider both the 1997 geographic shock and 1999 PE-RVU procedure-specific shock as an instrument for observed Medicare payments. Specifically, our first stage identifies the predictability of PE-RVU and GAF changes on overall Medicare payment changes within a market while controlling for the covariates specified in Equation (1). Because the PE-RVU policy shock differentially changes reimbursements for services performed in facility versus non-facility settings, we use the pre-1999 facility to non-facility ratio to generate a PE-RVU instrument that is independent of where a provider decides to perform a given procedure. In other words, the PE-RVU instrument for HCPCS i in HRR h in year t is given by:

$$\widehat{PE-RVU}_{iht} = \begin{cases} s_{iht} * PERVU_{it}^{f} + (1 - s_{iht}) * PERVU_{It}^{nf} & if\ t < 1999 \\ \overline{s_{\iota h}} * PERVU_{it}^{f} + (1 - \overline{s_{\iota h}}) * PERVU_{It}^{nf} & if\ t \geq 1999 \end{cases}$$

where the f and nf superscripts denote facility and non-facility components, respectively. s_{iht} is the share of services performed in a facility setting for a given HCPCS-HRR-year. For post-1999 policy years, we use the time-invariant share $\overline{s_{\iota h}}$ of services performed in a facility setting using data from 1996 to 1998. The GAF instrument is simply the GAF for a given HRR-year. The second stage uses the instrumented variation to estimate price elasticities.

Because Medicare payments are based on both PE-RVUs and GAFs, these instruments will be highly correlated with Medicare payments. These two policy shocks are also conditionally independent of other sources of change in quantity, strengthening the case for instrument validity. We discuss the other changes in Medicare payments in Appendix A.

4. Results

4.1. Prediction 1: Heterogeneity in Elasticities

4.1.a Elasticities by Service

First, we show that the size and sign of price elasticities may vary. Ordering HCPCS by their price elasticities, we plot the price elasticities estimated via OLS in Figure 3a and 2SLS with both instruments in Figure 3b. Only estimates that are statistically significant at the 5% level are shown. Both subplots clearly indicate that there are two types of HCPCS: (1) patient-driven HCPCS with negative price elasticities, and (2) physician-driven HCPCS with positive price elasticities. Furthermore, Appendix Figure B1 shows the price elasticities estimated from an IV approach using either PE-RVU or GAF as the sole instruments. These plots also indicate that there are patient- and physician-driven HCPCS.

When comparing Figure 3a with Figure 3b, it is evident that the OLS estimates tend to be more negative than the 2SLS estimates. This is consistent with a story where higher costs, which are unaccounted for in the regressions, decrease utilization. It is also consistent with RUC showing preferential payment increases for less common procedures, perhaps because those services were considered to be undervalued. Despite the differences between OLS and 2SLS, we cannot reject OLS as a valid approach. We test the endogeneity of the Medicare payment variable by examining the difference of two Sargan-Hansen statistics: one where payments are treated as endogenous (i.e., 2SLS) and another where payments are treated as exogenous (i.e, OLS).[11] Unlike the Durbin-Wu-Hausman test, this

[11] Under homeoskedasticity, this test is numerically equivalent to a Hausman test (Hayashi, 2000).

statistic is robust to violations of homeskedacitiy (Sargan, 1958; Hansen, 1982). Panel B of Table 1 lists the summary statistics for the p-value of the endogeneity test. Because p-values are large, we cannot reject the use of OLS in favor of IV. Also in Table 1 are summary statistics for the first stage F-statistics (Panel A) and p-values of tests on the over-identifying restriction (Panel C).

Of the statistically significant 2SLS estimates, the first stage F-statistic is higher than 10 in only about 30% of the HCPCS estimates. To check if the observed heterogeneity in price elasticities is driven by weak instruments, we consider two tests. First, we perform the conditional likelihood ratio test, which is robust to weak instruments and dominates the Anderson-Rubin test (Andrews et al., 2006). Figure 4(a) shows the confidence sets for HCPCS where the point estimate is contained in the set and the set does not include 0. The second test we consider is to examine the distribution of elasticities among the HCPCS with first stage F-statistics greater than 10, shown in Figure 4(b). Despite the fact that the GAF and PE-RVU are weak instruments for some services, both plots indicate that the existence of patient- and physician-driven HCPCS is independent of bias from weak instruments.

To test whether we should use one instrument or two, we perform a Sargan-Hansen test where the joint null hypothesis is that the instruments are valid and that the excluded instruments are correctly excluded from the estimated equation. Because the p-values are large, we find that in most cases, the over-identifying restrictions are not rejected. As a result, we rely on both instruments in subsequent instrumental variables regression estimates.

We examine the types of procedures that compose each type of HCPCS by tabulating the Berenson-Eggers Type of Service (BETOS) codes by type of HCPCS in Figure 5. The BETOS coding system covers all HCPCS codes, and each HCPCS code is assigned to only one BETOS codes. As Figure 5a demonstrates, major procedures, which are most often surgical in nature, tend to be patient-driven. On the other hand, imaging services and evaluation and management (E&M) services tend to be marginally more physician-driven. Imaging services include x-rays, computerized tomography (CT) scans, magnetic resonance imaging (MRI) scans, and ultrasound diagnostic testing. Evaluation and management services include physician office and hospital visits.

Despite the fact that there are both patient- and physician-driven HCPCS in each of these categories, the quantity of use across the different services are very different. Figure 5b illustrates the number of patient- versus physician-driven procedures performed in each category. Most notably, the share of imaging and E&M procedures is only marginally physician-driven, but the use of the physician-driven HCPCS in these categories greatly outweighs the use of patient-driven HCPCS in these categories. Furthermore, the share of major procedures that are patient-driven is high, but the difference in usage of patient- versus physician-driven major procedures is only marginal. The fact that utilization is disproportionately higher among physician-driven services provides suggestive evidence of our conjecture in Section 2.3 that patient-driven HCPCS can limit over-use.

Table 2 provides more detail regarding the types of services that are physician- versus patient-driven and their usages. We examine the 24 categories of care created by the two-digit BETOS codes. The five most patient-driven categories of care have usage fractions that relatively comparable to the share of services: about one-third of major procedures are physician-driven HCPCS, and these HCPCS account for about one-third of procedures performed. On the other hand, approximately 50-70% of tests, office visits, consultations, and oncology procedures are physician-driven HCPCS, but they account for more than 90% of procedures performed.

4.1.b Elasticities by Value

To shed further light on whether patient-driven procedures can reduce overuse, we examine the elasticities of services which have been identified as wasteful. We identify low-value services by following Schwartz et al. (2014), who draw upon the American Board of Internal Medicine Foundation's Choosing Wisely initiative, the US Preventive Services Task Force "D" recommendations, the National Institute for Health and Care Excellence "do not do" recommendations, the Canadian Agency for Drugs and Technologies in Health

health technology assessments, and peer-reviewed medical literature to identify services that provide little to no clinical benefit on average or in specific clinical scenarios.[12]

The results are shown in Table 3. Among the diagnostic and preventive screening tests, services tend to be patient-driven. However, among imaging and cardiovascular testing, low-value procedures tend to be physician driven. 11 of the 13 procedures have positive elasticities, and 7 of those are statistically significant. These findings suggest that inefficient overuse tends to follow physician-driven pricing behavior.

4.1.c Elasticities by Market

Another method of identifying the heterogeneity in price elasticities is to consider market-level elasticities across all services. With the large heterogeneity in reimbursements, cost, and utilization across geographic regions, it can be illuminating to consider what market areas are on average physician- versus patient-driven. Using data at the HRR-HCPCS-year level, we estimate elasticities for each HRR using a model that is similar to Equation (1). Instead of including HRR fixed effects, we include HCPCS fixed effects. Robust standard errors are estimated. To account for the fact that certain HCPCS are performed more than others, we weight observations by the national count of times each HCPCS is performed.

Figure 5 depicts which HRRs are patient- versus physician-driven. The darker shaded areas show markets where the estimates are statistically significant at the 10% level, and the lighter areas show markets where the estimates are not statistically significant. While many estimates are not statistically significant, the sign of the elasticities suggest that in most HRRs, increasing price will increase quantity. In scattered HRRs across the US, increasing price will reduce quantity. We highlight differences between patient- and physician-driven counties in Sections 4.2 to 4.4.

[12] Detailed methodology for identifying these procedures can be found in Schwartz et al. (2014) Supplementary Online Content. For sufficient estimation power, we focus on those procedures that are observed in at least 1,000 HRR-year units from 1993-2010.

4.2. Prediction 2: Patient Income

Next, we test the conjecture that HCPCS are patient driven when patient income is lower. To evaluate the effects of patient income, we rely on data from MCBS and consider socioeconomic status more broadly. Panel A of Table 1 shows the sample means when dividing the MCBS samples between patient- and physician-driven HCPCS. The means are weighted by the number of times each HCPCS is performed, which accounts for the relative importance of each HCPCS. Panel A demonstrates that on average, patient-driven HCPCS are correlated with patients with lower incomes, fewer years of schooling, and a smaller likelihood of having employer-sponsored insurance. These conclusions hold true regardless of whether we split the sample using the OLS or 2SLS price elasticity estimates, and the difference in means between patient- and physician-driven HCPCS are statistically different at the 5% level.

To further test these predictions, we show that price changes have differential effects at the patient-level. We collapse the data to the patient-year level and estimate

$$\log(RVU_{it}) = \beta \log(\mathrm{P_{it}}) + \delta \mathrm{C_{it}} + \alpha[\log(\mathrm{P_{it}}) \times \mathrm{C_{it}}] + \Gamma \mathrm{X_{it}} + \eta_{\mathrm{i}} + \delta_t + \epsilon_{\mathrm{it}}. \qquad (2)$$

Instead of a count of procedures performed, we now consider $\log(RVU_{it})$ for patient i in year t because the intensity level may differ across procedures.[13] For example, the total RVU measure takes into account the fact that performing a cardiac catheterization is not the same as performing a routine office visit. Our main independent variable is the interaction between log physician reimbursements and C_{it}, which is a measure of patient income, patient cost-sharing, or physician profitability. We additionally control for patient (η_i), year fixed effects (δ_t), and patient characteristics that vary over time (X_i), including CCI and age. We estimate Equation (2) using both OLS and 2SLS with the PE-RVU and GAF instruments.

To understand whether our prediction also holds across markets, we also collapse the data at the HRR-year level and estimate Equation (2), where i indexes HRRs instead of patients. The results are shown in Table 3. Columns (1) and (2) show the OLS and 2SLS estimates

[13] Estimates using quantity counts are similar in sign to the ones shown in Table 3.

using patient-year level data, and Columns (3) and (4) show the OLS and 2SLS estimates using HRR-year level data.

Panel A uses MCBS data and examines log of patient income as the interacted variable with log price. Following MacKinnon and Magee (1990), we use the inverse hyperbolic sine transform to address the problem of fitting an earnings distribution with a long left-tail. Furthermore, this transform allows us to address those with zero earnings, which is very prevalent in the Medicare setting. Because there is insufficient variation in income for a given person over time, we omit person fixed effects in Panel A. All columns show that the interacted coefficient is positive, and most of the columns are statistically significant. The positive interaction term indicates that when patient or market-area income increases, elasticities are more likely to be higher, or more physician-driven. This supports our conjecture that patient-driven HCPCS are more likely when patient income is lower.

4.3. Prediction 3: Patient Cost Sharing

Third, we test the prediction that HCPCS are patient-driven when patient cost sharing is higher. Shown in Panel B of Table 2, patient-driven HCPCS are correlated with higher out-of-pocket (OOP) payments and higher coinsurance payments. The difference in deductible payments is not statistically different. This finding is not surprising because the deductible—set at $100 per year in 2003— is not HCPCS dependent. On the other hand, coinsurance payments are set at 20% of the Medicare specified-fee and therefore vary by HCPCS. Out-of-pocket costs are defined as the sum of the deductible and coinsurance.

In Panel B of Table 3, we use CCF to estimate Equation (2) using patient-level data. Due to the large number of patients, we take a random 20% sample of the patient population to perform the analysis. We find that the interaction term between log price and the fraction of payments which are OOP is negative, which indicates that when patients are responsible for a larger share of the physician payment, the procedure tends to have a more negative elasticity, or behave more like a patient-driven HCPCS. Both the summary statistics and the patient-level regression results support the conjecture that patient-driven HCPCS are associated with higher patient cost-sharing.

4.4. Prediction 4: Physician Price-Cost Margin

Fourth, we test the conjecture that HCPCS are patient-driven when the physician's price-cost margin is lower. This is equivalent to testing that HCPCS are physician-driven when the physician's price-cost or profit margin is higher. Because we do not have data on costs, we construct two proxies to measure profitability using the allowed amount—which is what Medicare pays physicians—and the submitted amount—which is what physicians say they should be paid. The ratio between the allowed and submitted charge should indicate the percent of a physician's charges that are covered by CMS. Alternatively, the difference between the submitted and allowed charges should indicate the shortfall or the remaining cost to physicians that they must "cover" themselves because Medicare reimburses less than their proposed charges.

The first row of Panel C of Table 2 shows that physician-driven HCPCS are associated with procedures where Medicare covers a larger share of their requested payment. The second row shows that physician-driven HCPCS are associated with procedures where physicians incur a smaller cost from performing the procedure. Panel C of Table 3, which looks at the interaction between log prices and log physician profitability, also shows that when Medicare covers a larger share of a physician's requested payment, HCPCS tend to be more positive, i.e., more physician driven.

One may argue that submitted charges are biased by measurement error as the charges physicians submit have no bearing on the actual payment received. While this concern may be valid, changes in the submitted charge for a given procedure over time are less likely biased. Thus, we approach this conjecture by administering another test. For each HCPCS, we calculate two elasticities: one that uses profitability changes above the median and another that uses changes below the median. Results are shown in Appendix Table B1. We find that when changes in physician profitability are larger, HCPCS have a 0.06 to 0.09 higher probability of being physician-driven.

4.5. Prediction 5: Policy Implication

One of our normative implications is that physician payment reforms – such as reductions in reimbursement -- will have larger effects among physician-driven HCPCS. To empirically assess this hypothesis, we rely on the 1999 PE-RVU payment change as our physician payment reform policy. We first establish whether HCPCS are physician- or patient-driven by estimating a price-elasticity using pre-policy data from 1993 to 1998. Because our instrument tends to be weak at the HCPCS-level, this first regression is estimated following Clemens and Gottlieb (2014):

$$\log\left(Q_{ht}^{i}\right) = \beta^{i}\Delta GAF_{h} * 1(t \geq 1997) + \Gamma^{\mathrm{i}}X_{ht}^{i} \ + \eta_{t} + \delta_{h} + \epsilon_{iht}, \qquad (3)$$

where ΔGAF is calculated using the change in GAF from 1996 to 1997. Then, we run a second regression at the HCPCS-year level that examines whether the post-1999 PE-RVU shock led to larger quantity changes for physician-driven HCPCS. Specifically, the second regression utilizes data from 1998-2002, and we estimate separately for physician-driven HCPCS ($\hat{\beta} > 0$) and patient-driven HCPCS ($\hat{\beta} < 0$):

$$\log(RVU_{it}) = \beta \log(P_{it}) + \Gamma X_{it} + \eta_{t} + \delta_{i} + \epsilon_{it}, \qquad (4)$$

We utilize a seemingly unrelated regression framework and bootstrap for standard errors.

It is important to note that coinsurance rates are 20% of Medicare reimbursement rates, so it is difficult to separate responses due to a change in physician payments from those due to a change in patient cost-sharing. Our conjectures indicate that physician-driven HCPCS are positively correlated with profitability and negatively correlated with patient cost-sharing. Thus, it is important to be able to separate the effects of payment increases on profitability from the effect of cost-sharing. Focusing on the Qualified Medicare Beneficiaries (QMBs) can allow us to separate out these two factors. QMBs are Medicare beneficiaries who are also eligible for Medicaid, and they are not responsible for paying either the Medicare deductible or Part B. Thus, changes in quantity among this population will only reflect responses to a physician payment change.[14]

[14] Also in 1997, the Balanced Budget Act reduced QMB cost-sharing rates. Post-1997, states were only required to cover cost-sharing rates up to the Medicaid reimbursement rate, instead of the Medicare reimbursement rate (Mitchell and Haber, 2003). This reduced the payments that physicians received, but it did not affect the zero cost-sharing policy among QMBs.

Results are shown in Table 6. In Panel A, we consider all Medicare beneficiaries. In Panels (B) and (C), we consider dual- and non-dual eligibles. In all three panels, the response to a price change is larger for physician-driven HCPCS. However, the difference in quantity response between physician- and patient-driven HCPCS is most statistically significant among the dual-eligible population.[15]

One drawback of the data is that it does not allow us to differentiate between QMBs and other dual-eligibles. For example, Service Limited Medicare Beneficiaries (SLMBs) are not responsible for their deductible, but they are still required to pay the copay. Because we cannot isolate QMBs from other dual-eligibles, the difference between Columns (2) and (3) will be understated. Note that column (1) is not directly comparable to the other columns because the identification of patient- versus physician-driven HCPSC is specific to the sample population.

4.6. Implications for Medicare price changes

Our analysis suggests that changes in physician reimbursement rates may not always have the intended effects. To further illustrate this point, we consider four types of Medicare payment changes and show their corresponding responses in total quantity.

First, consider a uniform 10% decrease or increase in Medicare reimbursements for all services. We use the IV-estimated elasticities from Equation (1) to calculate the change in total quantity when prices change by 10%. Depicted in Scenario 1 of Figure 7, the percent change in total quantity is either a 34% decrease or increase. This finding is in line with the idea that on average, increasing Medicare reimbursements increases use, and decreasing Medicare reimbursements decreases use.

However, now consider two other policies: a 10% decrease in payments for physician-driven HCPCS and a 10% increase in payments for patient-driven HCPCS. Physician- and

[15] Estimates with an OLS second stage model are shown in Appendix Table B3. The results are similar in this table.

patient-driven HCPCS are again identified using the IV-estimated elasticities from Equation (1). These policies lead to a reduction in total quantity by 52% and 17% respectively. This second scenario illustrates two important points. First, it is possible to generate a reduction in quantity by increasing the payment for certain services, namely the patient-driven services. Second, the magnitude of change can be much larger when the payment policy targets specific procedures.

An alternative policy is to target markets, instead of services. Scenario 3 indicates that a targeted 10% decrease in payments in physician-driven markets and a 10% increase in payments in patient-driven markets can lead to overall reductions in care. Scenario 4 additionally increases income across all markets. Changing income can magnify the total cost reduction stemming from a targeted change in prices across markets.

5. Conclusion

In this paper, we present a model of joint physician and patient decision-making. By examining how altruism interacts with profit-maximizing incentives, our model demonstrates that the quantity response to a price change can vary in not only magnitude, but also direction. We identify when HCPCS are likely to follow physician-driven pricing behavior versus patient-driven pricing behavior. Specifically, patient-driven behavior is more common when patient income is low, patient health care spending is high, and when the physician price-cost margin is low. We provide empirical evidence in support of these conjectures. The theory suggests two remaining implications that could be tested in future work: patient-driven behavior is more common when physician altruism is high and when physician income is high.

Our model also offers an important policy implication: physician reimbursement reforms that move reimbursements closer to the social value of inputs used will be more effective in reducing social inefficiency when pricing is physician-driven. While we do not structurally estimate the degree of social inefficiency in our data, we provide empirical evidence that suggests physician reimbursement reforms have a larger effect on physician-driven HCPCS.

The health economics literature has long recognized the tension between physician altruism and physician profit-maximization. Economists have developed an elegant and tractable model accounting for this tension. We exploit these tools to generate novel testable predictions about pricing and utilization behavior in healthcare markets. Our analysis demonstrates that the unique preferences and objectives of physicians creates pricing dynamics in healthcare that depart from those in other product markets.

These implications seem consistent with the data and provide useful guidance for policymakers and researchers. First, physicians are systematically more "altruistic" – in the sense of pursuing patient interests – when treating more vulnerable and disadvantaged patients. Second, heterogeneity in the effect of reimbursement changes is to be expected, and can be exploited to increase the effectiveness of reimbursement reforms. Reimbursement reductions might be useful tools for containing costs when physicians are largely profit-maximizing, but they may be counterproductive when they are more altruistic. Being able to differentiate when a service or market is physician- vs. patient-driven will allow policy makers to more effectively target supply- and demand-side incentives.

References

Andrews, D., M. Moreira, and J. Stock. "Optimal Two-Sided Instrumental Variables Regression." *Econometrica* 74(3): 116-132.

Arrow, K.J. (1963). "Uncertainty and the Welfare Economics of Medical Care." *American Economic Review* 53: 941-973.

Becker, G. (1957). The Economics of Discrimination. Chicago: The University of Chicago Press, 2nd edition.

Blomqvist, A. (1991). "The Doctor as Double Agent: Information Asymmetry, Health Insurance and Medical Care." *Journal of Health Economics* 10:411-432.

Clemens, J. and Gottlieb, J. (2014). Do Physicians' Financial Incentives Affect Medical

Treatment and Patient Health? *American Economic Review*, 104(4): 1320-49.

Clemens, J. and Gottlieb, J. (2013). Bargaining in the Shadow of a Giant: Medicare's Influence on Private Payment Systems. NBER Working Paper No. 19503.

Chone, P. and Ma, C. (2011). Optimal Health Care Contract Under Physician Agency. *Annals of Economics and Statistics* 101/102: 229-256.

Dickstein, M. (2014). Physician vs. Patient Incentives in Prescription Drug Choice. *Draft.*

Dranove, D. (1988). Demand Inducement and the Physician-Patient Relationship. *Economic Inquiry*, 26(2): 281-296.

Dranove, D. and W.D. White (1987). "Agency and the Organization of Health Care Delivery." *Inquiry* 24:405-415.

Eaton, J. Little-known AMA Groups has Outsized Influence on Medicare Payments. *The Center for Public Integrity*. <http://www.aafp.org/fpm/2008/0200/p36.html>

Ellis, R. and McGuire, T. (1985). Provider Behavior Under Prospective Reimbursement: Cost Sharing and Supply. *Journal of Health Economics* 5:129-151.

Ellis, R. and McGuire, T. Optimal Payment Systems for Health Services. *Journal of Health Economics* 9:375-396.

Emanual, E. and Emanual, L. (1992). Four Models of the Physician-Patient Relationship. *Journal of the American Medical Association*, 267(16), 2221-2226.

Department of Health and Human Services. "Medicare Program; Revisions to Payment Policies and Five-Year Review of and Adjustments to the Relative Value Units Under the Physician Fee Schedule for Calendar Year 1997. *Federal Register* 61(227): Report 59490.

Farley, P.J. (1986). 'Theories of the Price and Quantity of Physician Services." *Journal of Health Economics* 315-333.

Felder, S., & Mayrhofer, T. (2011). *Medical decision making : a health economic primer*. Heidelberg: Springer.

Gaynor, M. (19940. "Issues in the Industrial Organization of the Market for Physician Services." *The Journal of Economics & Management Strategy* 211-255.

Glied, S. and Zivin, J.G. (2002). How do Doctors Behave When Some (But Not All) of Their Patients are in Managed Care? *Journal of Health Economics* 21:337-53.

Gruber, J. (2013). Proposal 3: Restructuring Cost Sharing and Supplemental Insurance for Medicare. *Brookings: The Hamilton Project*.

Hadley, J. and R. Lee (1978). "Toward a Physician Payment Policy: Evidence From the Economic Stabilization Program." *Policy Sciences* 10:105-120.

Hansen, L.P. (1982). Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica* 50: 1029-1054.

Hayashi, F. (2000). Econometrics. Princeton, NJ: Princeton University Press.

Lakdawalla, D. and Sood, N. (2013). Health Insurance as a Two-Part Pricing Contract. *Journal of Public Economics, 102*, 1–12.

Liu, T. and Ma, C. (2013). Health Insurance, Treatment Plan, and Delegation to Altruistic Physician. *Journal of Economic Behavior & Organization* 85:79-96.

Lu, M. (1999). "Separating the 'True Effect' from 'Gaming' in Incentive-based Contracts in Health Care." *Journal of Economics and Management Strategy* 8:383-432.

MaCurdy T., Sood, J., DeLeire, T., DeVaro, J., Bounds, M., Pham, D., and Chia, A. (2012). Geographic Adjustment of Medicare Payments to Physicians: Evaluation of IOM Recommendations. Acumen, LLC: Burlingame, CA.

MacKinnon, J.G. and L. Magee (1990). Transforming the Dependent Variable in Regression Models. International Economic Review, 31(2): 315-339.

Maxwell, S. and Zuckerman, S. Impact of Resource-Based Practice Expenses on Medicare Physician Volume. Health Care Financing Review 29(2): 65-79.

McGuire, T. (2000). Physician Agency. In: *Handbook of Health Economics, Volume I.* Culyer, A.J. and Newhouse, J.P. editors. Elsevier Science: 462-536.

Mitchell, J.B. and Haber S.G. (2003). State Pyament Limitations on Medicare Cost-Sharing: Impacts on Dually Eligible Beneficiaries and Their Profiders. *Health Economics Research, Inc*. CMS Contract No. 500-95-0058.

Mooney, G. and M. Ryan (1993). "Agency in Health Care: Getting Beyond First Principles." *Journal of Health Economics* 12:125-135.

Pauly, M. (1980). Doctors and their Workshops: Economic Models of Physician Behavior." *National Bureau of Economic Research Monograph*.

Quan, H., Sundararajan, V., Halfon, P., Fong, A., Burnand, B., Luthi, J., Saunders, L., Beck, C., Feasby, T. (2005). Coding Algorithms for Defining Comorbidities in

ICD-9-CM and ICD-10 Administrative Data. *Medical Care*, 43(11): 1130-9.

Sargan, J.D. (1958). The Estimation of Economic Relationships Using Instrumental Variables. *Econometrica* 26: 393-415.

Schwartz, A.L., Lnadon, E. Elshaug, A., Chernew, M., McWilliams, J. (2014). Measuring Low-value Care in Medicare. *JAMA Internal Medicine* 174(7):1067-76.

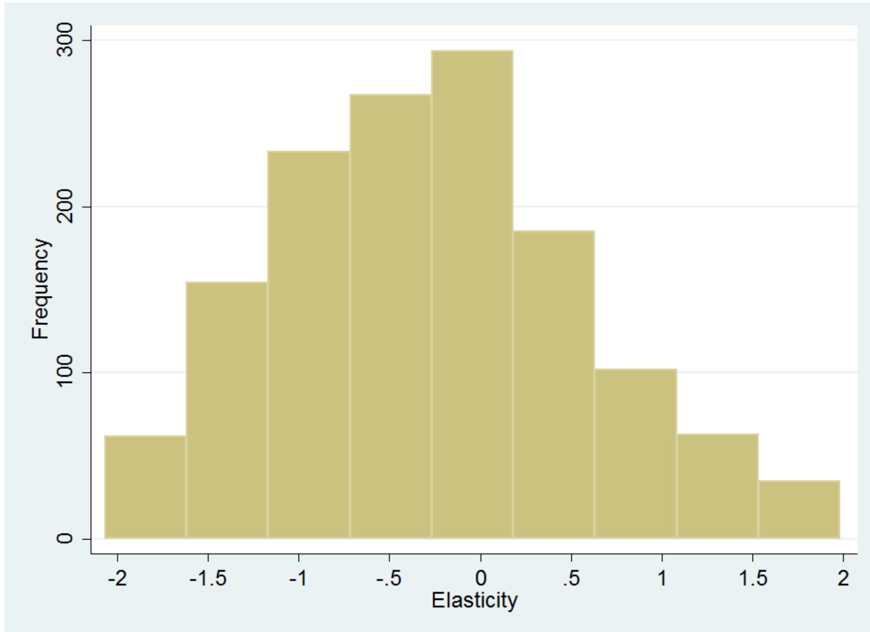
The National Commission on Fiscal Responsibility and Reform. (2010). The Moment of Truth. Accessed February 26, 2015. <http://www.fiscalcommission.gov/sites/fiscalcommission.gov/files/documents/TheMomentofTruth12_1_2010.pdf>.

U.S. Government Accountability Office (2005). *Medicare Physician Fees Geographic Adjustment Indices are Valid in Design, But Data and Methods Need Refinement.* Pub No. GAO-05-119. Washington DC.

Zwefel, P. and F. Breyer (1997). Health Economics. Oxford University Press: New York.

Zuckerman, S., Shang, B., and Waidman, T. (2010). Reforming Beneficiary Cost Sharing to Improve Medicare Performance. *Inquiry*, 47(3): 215-225.

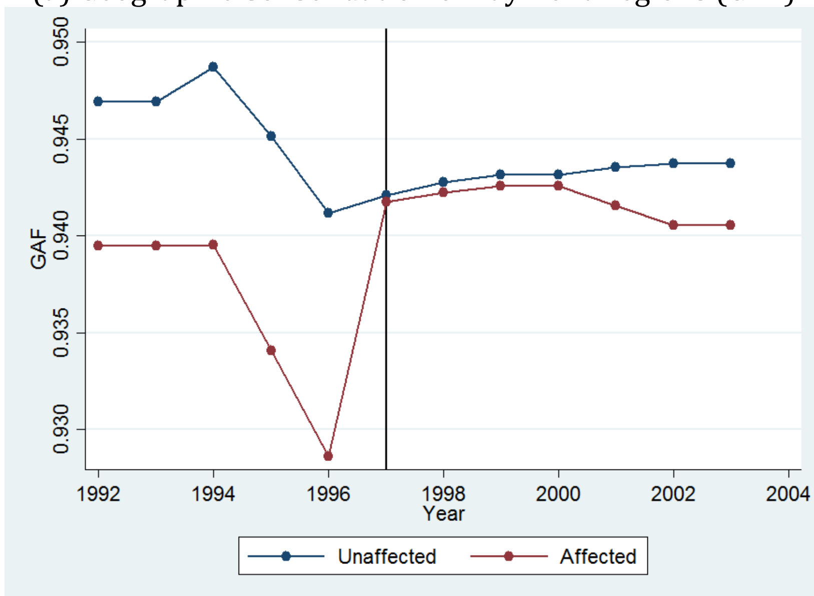
Figure 1: Histogram of Elasticities When Prices Increase Significantly



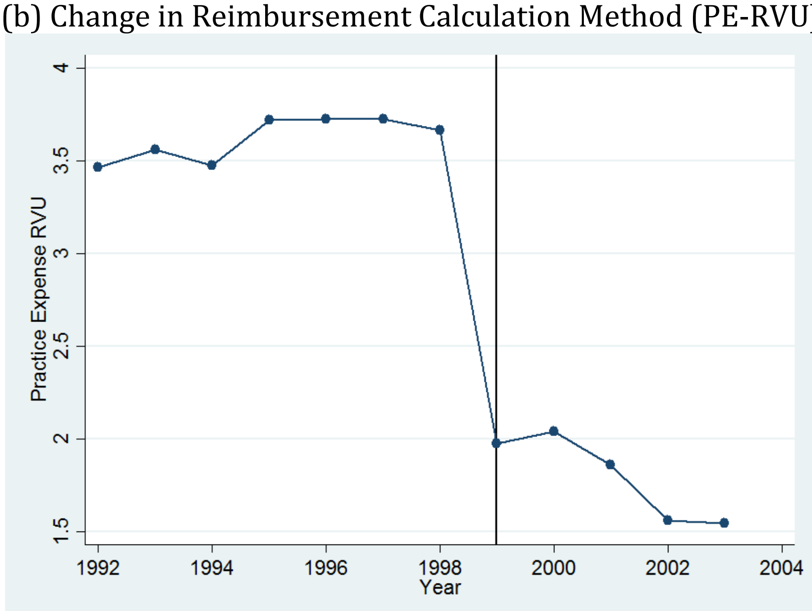
Notes: Data from CMS Medicare 5% claims, 1992-2003. This figure shows the elasticities (calculated simply as the annual percent change in quantity divided by the annual percent change in price) for HCPCS with annual physician payment increases ranging from 45% to 55%. It is evident that quantity increases for about half of the HCPCS, while quantity falls for the other half. The long right tail has been truncated.

Figure 2: Shocks in Components of Medicare Payments Over Time

(a) Geographic Consolidation of Payment Regions (GAF)



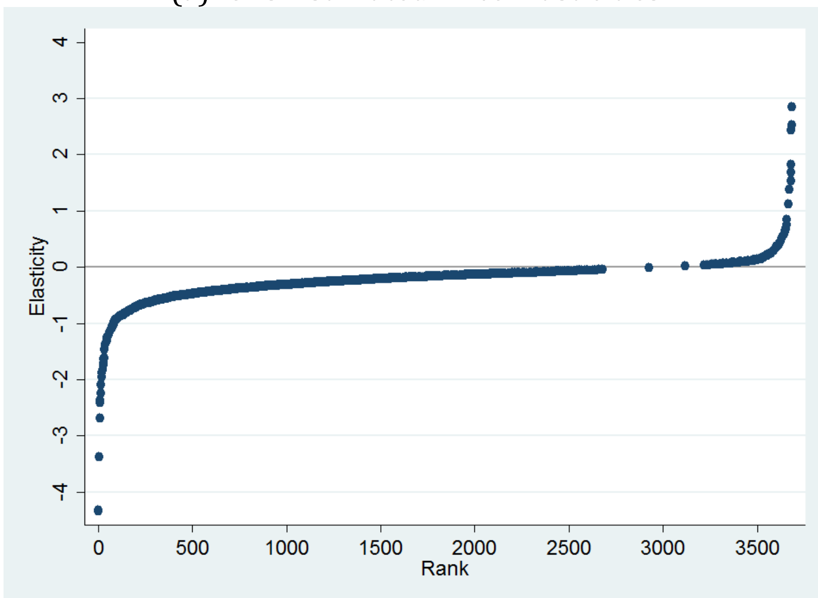
(b) Change in Reimbursement Calculation Method (PE-RVU)



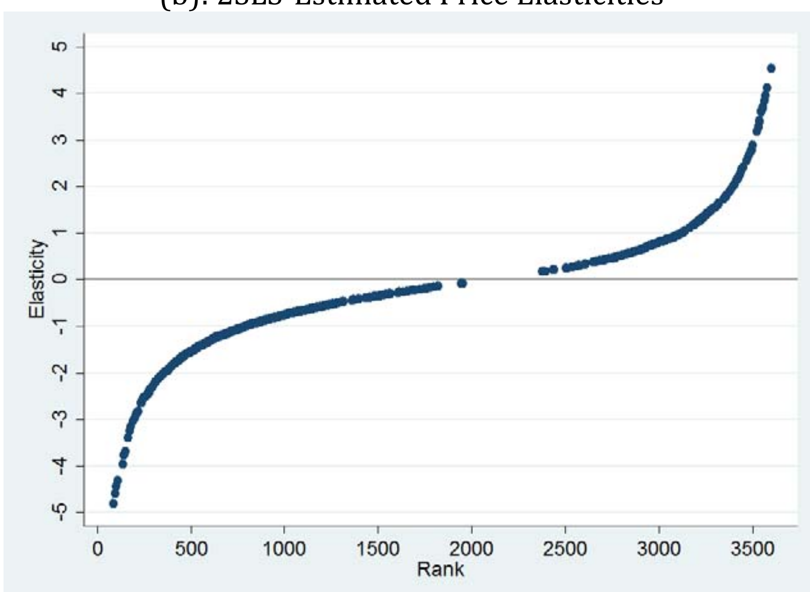
Notes: Data from the Federal Register 1992-2003. The sample is limited to HCPS observed in all years. Plot (a) shows the average GAF across counties that were or were not affected by the 1997 consolidation of payment regions from 210 to 89 payment regions. Plot (b) depicts the change in average of facility and non-facility PE-RVUs across HCPCS. In 1999, HCFA more accurately priced non-facility services and phased in a new methodology of calculating PE-RVUs.

Figure 3: Estimated Elasticities by HCPCS

(a): OLS-Estimated Price Elasticities



(b): 2SLS-Estimated Price Elasticities



Notes: Data comes from the CCF. Each dot comes from a separate regression of Equation (1); elasticities are ordered and plotted. For both plots, only the HCPCS with statistically significant price elasticities at the 5% level are shown. In plot (b), the instruments are PE-RVU and GAF. See Appendix Figure B2 for price elasticities estimated using an individual instrument.

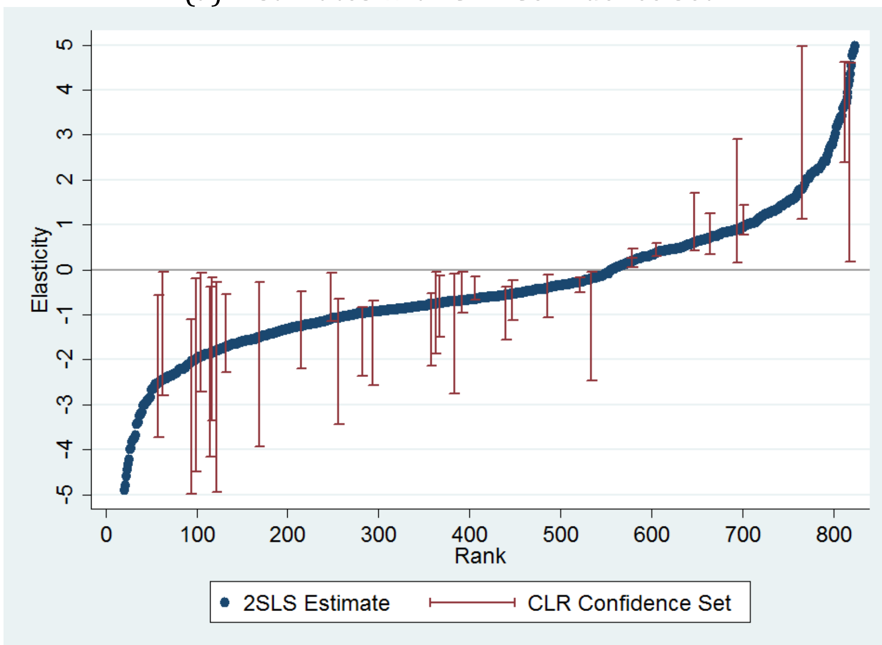
Table 1: Summary of IV Related Statistics

Instruments	(1)	(2)	(3)
	Panel A: First Stage F-Statistics		
	25th pct	Median	75th pct
PE-RVU	2.479	5.277	12.467
GAF	0.699	2.637	6.070
PE-RVU + GAF	1.427	5.096	14.187
	Panel B: P-value of Endogeneity Test		
	Fraction p-value<0.10	Fraction p-value<0.05	Fraction p-value<0.01
PE-RVU	0.059	0.035	0.010
GAF	0.074	0.049	0.018
PE-RVU + GAF	0.096	0.069	0.024
	Panel C: P-Value of Hansen J-Statistic		
	Fraction p-value<0.10	Fraction p-value<0.05	Fraction p-value<0.01
PE-RVU + GAF	0.035	0.016	0.004
No. of Regressions		839	

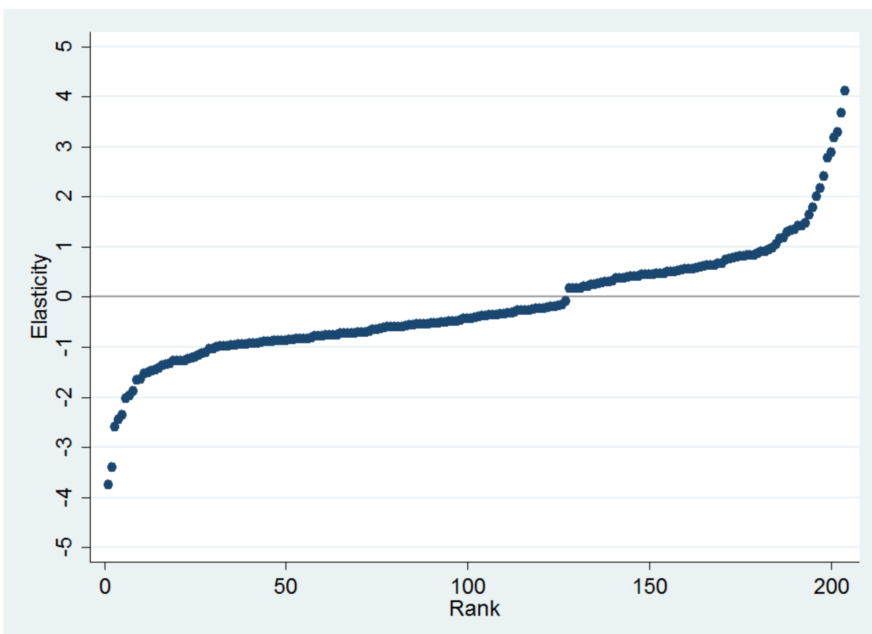
Notes: This table shows the IV related summary statistics used to estimate the statistically significant elasticities shown in Figure 3(b) and Appendix Figure B1. Panel A shows the first-stage F-statistics for when using PE-RVU, GAF, or both PE-RVU and GAF as instruments. One outlier with F-stat>10,000 was dropped to avoid inflating the mean and standard deviation of F-stats. Panel B shows the distribution of p-values for the test that the Medicare price variable used in OLS is endogenous. Panel C shows the Hansen J-statistic for the test for the validity of using both instruments, instead of one or the other.

Figure 4: Statistically Significant 2SLS Estimates

(a): Estimates with CLR Confidence Set



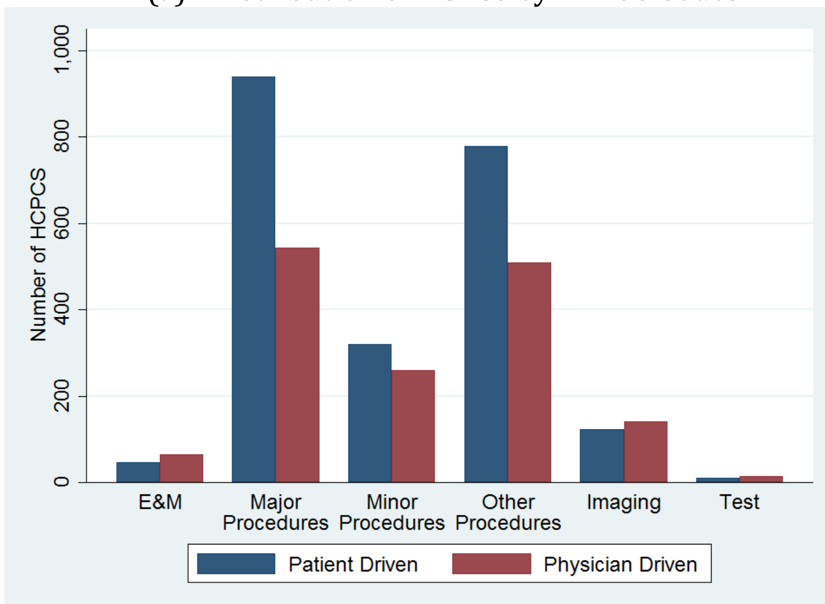
(b): Estimates with First Stage Fstat>10



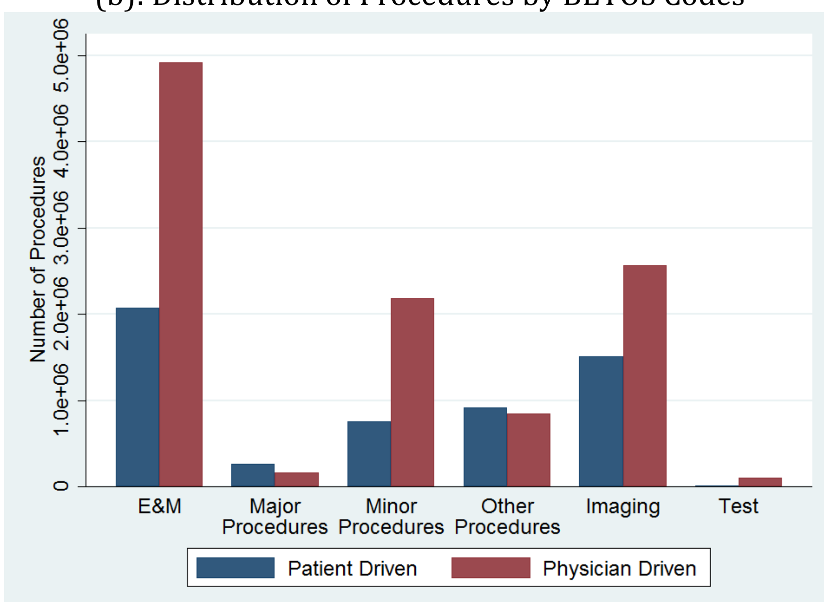
Notes: Data comes from the CCF. HCPCS with statistically significant price elasticities at the 5% level are shown. In plot (a), estimates with β in the Conditional Likelihood Ratio (CLR) and those with CLR not crossing the x-axis are depicted with the error bars. In plot (b), only HCPCS with a first stage F-stat>10 are shown.

Figure 5: Description of Patient- versus Physician-Driven HCPCS

(a) : Distribution of HCPCS by BETOS Codes



(b): Distribution of Procedures by BETOS Codes



Notes: Data from CCF. Patient- and physician-driven HCPCS are identified using the 2SLS estimates shown in Figure 2. Plot (a) shows the number of patient- versus physician-driven HCPCS by BETOS category. E&M represents evaluation and management procedures. Plot (b) uses the 2003 CCF data, and it depicts the number of patient- versus physician-driven procedures performed by BETOS category.

Table 2: Most Patient- and Physician-Driven Categories of Care

	Share Physician-Driven		Elasticity Estimate	
	Procedures (1)	HCPCS (2)	OLS (3)	IV (4)
Five Least Physician-Driven				
Hospital visits (M2)	0.274	0.409	-0.217***	-1.628***
Major eye procedures (P4)	0.336	0.368	-0.086***	-0.046***
Major other procedures (P1)	0.356	0.372	-0.178***	-0.296***
Cardiovascular procedures (P2)	0.363	0.393	-0.080***	-0.179***
Ambulatory procedures (P5)	0.394	0.391	-0.061***	-0.656***
Five Most Physician-Driven				
Tests (T2)	0.929	0.583	0.113***	-0.508***
Office visits (M1)	0.941	0.679	-0.027**	0.186
Consultations (M6)	0.955	0.777	0.624***	0.965***
Oncology (P7)	0.978	0.5	0.090***	0.637***
Emergency room visits (M3)	1	1	0.367***	0.970***

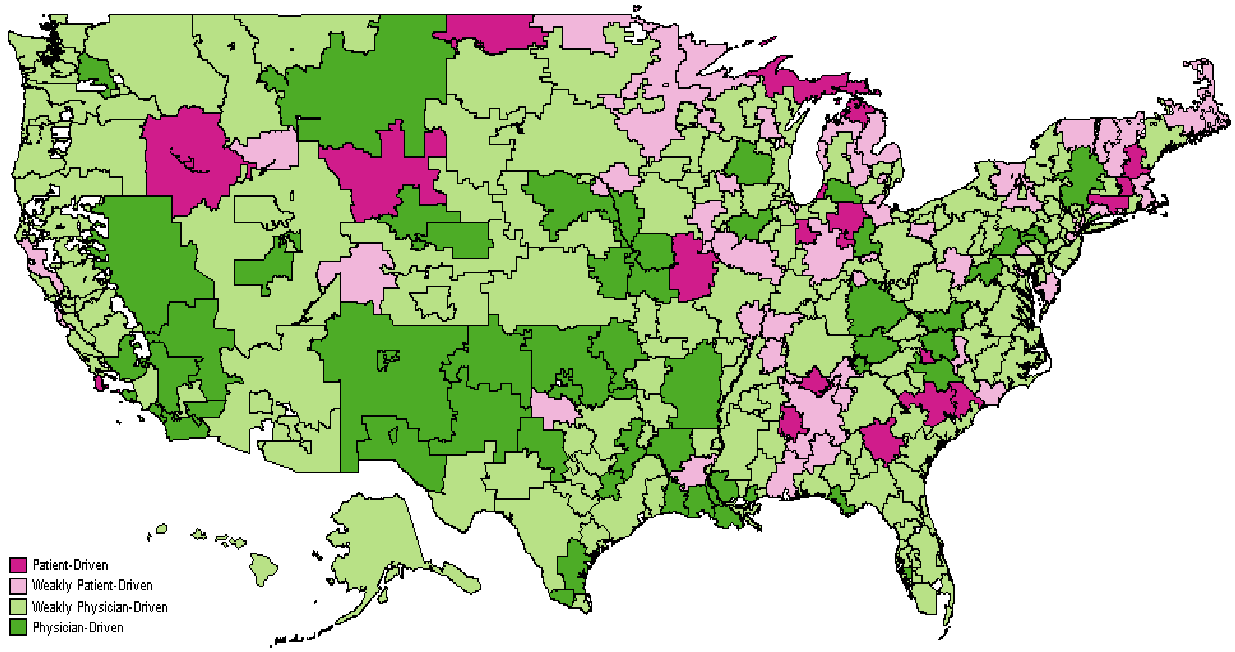
Notes: Data from CCF. We rank the 24 two-digit BETOS codes by the share of procedures which are patient- versus physician-driven. The top (bottom) five categories of care correspond to the most patient- (physician-) driven service groups. Columns (1) and (2) show the physician-driven share of procedures performed in 2003 and the physician-driven share of HCPCS. The number of procedures in each category range from 135,896 to 2,837,881 procedures. The number of HCPCS in each category range from 5 to 848. Columns (3) and (4) show the elasticity estimates vis OLS and 2SLS approaches for procedures in each category.

Table 3: Elasticities for Low-Value Care

Dep. Variable: Log(Q)	Coefficient	Standard deviation	# of Obs.
	A. Diagnostic and preventative testing		
Colorectal cancer screening	-0.201***	(0.046)	3,008
Preoperative			
Chest radiography	-0.0394**	(0.0194)	3,003
Echocardiography	0.139***	(0.0455)	2,952
Stress testing	-0.131**	(0.0567)	2,848
	B. Imaging		
CT of sinus	0.360***	(0.107)	2,873
Imaging evaluation	0.117**	(0.0521)	3,003
Imaging headache	0.260***	(0.0546)	3,003
Electroencephalogram	0.515	(0.402)	2,947
Imaging back	0.192***	(0.0397)	3,004
	C. Cardiovascular testing and others		
Asymptomatic screening	0.817***	(0.196)	3,002
Syncope screening	1.155***	(0.176)	3,003
Coronary stress test	0.668***	(0.0705)	3,003
Coronary balloon test	0.123	(0.262)	2,980
Coronary renal stent	0.0725	(0.0737)	1,690
Endarterectomy	-0.0462	(0.626)	2,412
Arthroscopic surgery	0.355	(0.257)	2,438

Notes: Data from CCF at the service-year-HRR level. Low-value services identified using eTable 1 from Schwartz et al. (2014). We focus on procedures with at least 1,000 observations over time. Each row comes from a separate 2SLS regression. Corresponding OLS regressions are shown in Appendix Table B2.

Figure 6: Patient- and Physician-Driven HRRs



Notes: Data from CCF. For each HRR, we calculate the average price elasticity using data at the HCPCS-year level and a 2SLS model. We include HCPCS fixed effects, year fixed effects, BETOS by year trend, CCI, age, and gender and race dummies. HCPCS are weighted by the national usage. Green areas represent physician-driven HRRs. Pink areas represent patient-driven HCPCS. The lighter shades indicate HRRs where the price elasticity estimate is not statistically significant at the 10% level.

Table 4: Summary Statistics, by Sign of Own-Price HCPCS Elasticity

	OLS		2SLS	
	Patient-Driven (1)	Physician-Driven (2)	Patient-Driven (3)	Physician-Driven (4)
	Panel A: Patient Socioeconomic Status			
Gross income ($1000s)	21.28	23.06***	21.15	23.35***
Employer coverage	0.33	0.35***	0.33	0.36***
Less than high school	0.43	0.39***	0.37	0.38***
HS graduate	0.16	0.17***	0.43	0.17***
Some college	0.13	0.14***	0.16	0.14***
College grad or more	0.13	0.15***	0.13	0.15***
Black	0.11	0.094***	0.11	0.092***
Hispanic	0.043	0.042	0.043	0.041*
	Panel B: Patient Cost-Sharing			
OOP ($)	51.83	20.45***	48.38	31.01***
Coinsurance ($)	49.87	18.41***	46.45	28.98***
Deductible ($)	1.96	2.04	1.94	2.03
	Panel C: Profitability			
Percent Reimbursed (%)	55	62***	54	63***
Shortfall ($)	419.19	143.40***	387.61	237.26***
No. of Obs. (MCBS)	675	929	884	794
No. of Obs. (CCF)	538	742	562	718

Notes: Data from 1993-2002 at HCPCS level. Data for Panels A is from MCBS. Data from Panels B and C are from CCF. Summary statistics are weighted by number of observations per HCPCS. Columns (1) and (2), or Columns (3) and (4) are statistically different at the * 10% level, ** 5% level, *** 1% level. In Panel A, the insurance coverage and education variables are measures of the fraction of patients with each characteristic. In Panel B, *Any OOP* is the fraction of patients who had OOP>0. *Fraction OOP* is the average fraction of total payments attributed to out of pocket costs. In Panel C, *Fraction Reimbursed* is calculated by the share of payments CMS allows relative to the physician submitted charge (i.e., Allowed/Submitted). The *Shortfall* is the amount providers bill CMS minus the actual CMS payment (i.e., Submitted-Allowed).

Table 5: Relationship Between Elasticities and Patient/HRR Characteristics

† Dep. Var.: Log(RVU) $\mu = 4.09$ or 2.79	Patient-Year Level Data		HRR-Year Level Data	
	OLS (1)	2SLS (2)	OLS (3)	2SLS (4)
	Panel A: Patient Socioeconomic Status			
Log(Price)	1.089***	1.766***	0.429***	1.420***
	(0.0093)	(0.0130)	(0.0547)	(0.107)
Log(Price) x	0.986***	1.256***	1.807*	3.474***
IHS(Income)	(0.163)	(0.174)	(0.065)	(0.114)
First stage F-stat	---	16,157	---	127.5
R-squared	0.274	0.173	0.897	0.067
	Panel B: Patient Cost Sharing			
Log(Price)	1.154***	2.982***	1.080***	9.000***
	(0.001)	(0.009)	(0.454)	(1.329)
Log(Price) x	0.0738***	-0.179***	-0.245	-2.090***
Log(OOP)	(0.0003)	(0.001)	(0.181)	(0.509)
First stage F-stat	---	678,596	---	5.34
R-squared	0.622	0.204	0.876	0.037
	Panel C: Physician Profitability			
Log(Price)	0.829***	1.607***	-0.407***	2.348***
	(0.0004)	(0.001)	(0.306)	(1.315)
Log(Price) x	0.096***	0.063***	1.061***	-1.863
Percent Reimbursed	(0.0003)	(0.0003)	(0.252)	(1.470)
First stage F-stat	---	179,110	---	37.62
R-squared	0.566	0.150	0.878	0.316
Person FE, Year FE?	Y	Y	---	---
HRR FE, Year FE?	---	---	Y	Y
No. of Obs (MCBS)	83,222	83,222	2,192	2,192
No. of Obs (CCF)	13,944,454	13,944,454	3,012	3,012

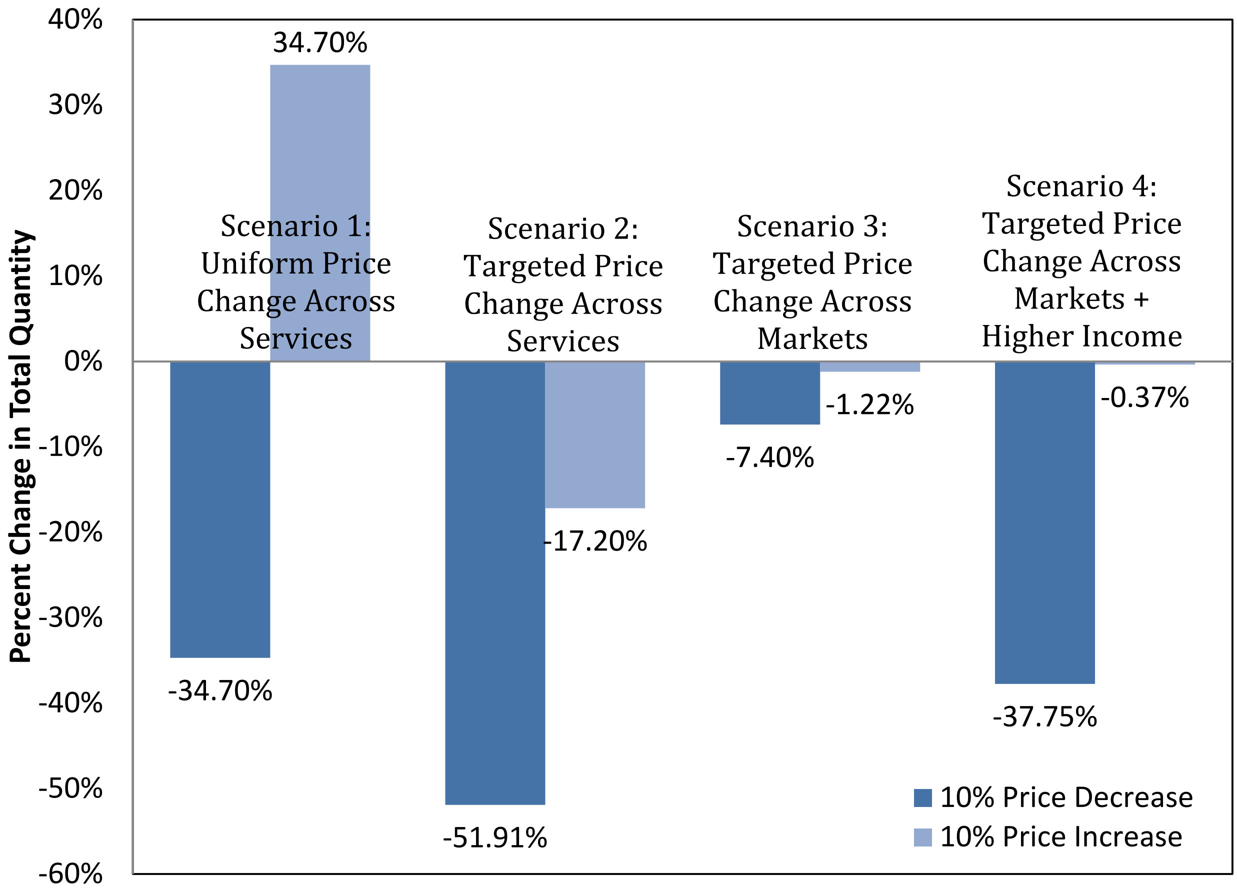
Notes: Each panel and column represents a separate regression. Data for Panel A is from MCBS. Data for Panels B and C are from CCF. The analysis in Columns (1) and (2) are at the patient-year level, and the analysis in Columns (3) and (4) are at the HRR-year level. † First reported mean is from MCBS; second reported mean is for CCF. The dependent variable is log(total RVU). All regressions include the relevant characteristic (income, cost-sharing, or profitability). All columns control for patient's CCI, age, and year fixed effects. For Panel A, all columns additionally control for HRR fixed effects. For Panel B, Columns (1) and (2) additionally control for person FE while Columns (3) and (4) additionally control for HRR fixed effects. Robust standard errors are shown in parentheses. * 10% level, ** 5% level, *** 1% level.

Table 6: Differential Effect of a Physician Reimbursement Reform

	Dependent Var: Log(RVU)		
	Physician-Driven HCPCS (1)	Patient-Driven HCPCS (2)	X^2 and P-value for H_0: (1)=(2) (3)
	Panel A. All Beneficiaries		
Log(Price)	1.755***	1.000***	2.41
	(0.142)	(0.320)	0.125
	Panel B. Dual Eligibles		
Log(Price)	1.630***	0.796***	5.07
	(0.150)	(0.245)	0.024
	Panel C. Non-Dual Eligibles		
Log(Price)	1.784***	1.217***	3.89
	(0.138)	(0.400)	0.067
First Stage F-stat	[60.3, 89.5]	[6.3, 17.1]	---
No. of Obs.	3,396	735	---

Notes: The physician-and patient-driven HCPCS are determined using CCF data from 1993 to 1998 and the GAF policy change. Each cell contains data from a separate regression using CCF data from 1998 to 2002. The dependent variable is Log(Total RVU) and independent varaibles include CCI, age, race, and gender dummies, year, and HCPCS fixed effects. Bootstrapped errors shown in parentheses. Column (3) shows the two-sided chi-squared and p-values for the hypothesis test that the elasticity estimates in Columns (1) and (2) are the same.

Figure 7: Effects of Counterfactual Price Changes on Quantity



Notes: Scenario 1 shows the percent change in total RVUs performed when 2000 prices uniformly decrease by 10% (dark blue) or uniformly increase by 10% (light blue). Scenario 2 shows the percent change in total RVU when 2000 prices decrease by 10% for only the physician-driven HCPCS (dark blue) or increase by 10% for only the patient-driven HCPCS (light blue). Scenarios 3 and 4 show the targeted and uniform changes by HRRs, instead of HCPCS. Scenario 4 adds a 10% income increase across all HRRs. The IV-elasticity estimates, as shown in Figure 3b, are used to calculate the percent change in total RVU.

Appendix A

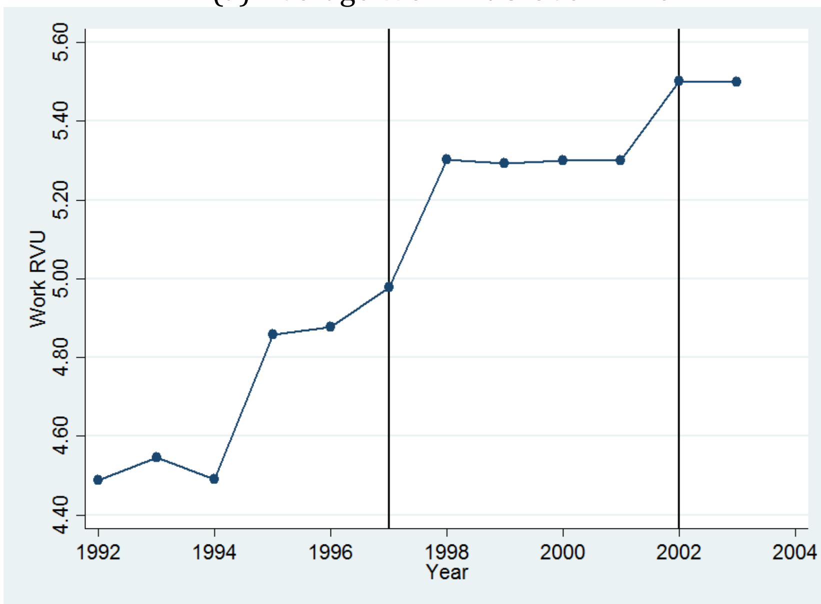
In this section, we discuss the policy changes that affected the remaining Medicare components. As discussed in Section 3.2, we use the 1997 GAF and 1999 PE-RVU policy shocks as instruments for Medicare prices. The remaining variation in Medicare payments come from variation in the work RVU, malpractice RVU, and CF. On average, work accounts for 52% of total physician payments, practice expenses represent 44%, and liability insurance represents 4% (US Government Accountability Office, 2005). Because the malpractice component accounts for a small share of payments, we do not focus on that component.

During this time period, work RVUs experienced two major reviews which became effective in 1997 and 2002. Plot (a) of Figure A1 shows the average work RVU over time for HCPCS that are observed in each year of the study period. After the RUC committee met to re-assess work RVUs, we see clear jumps in the RVU. However, with competing political pressures and physician incentives, it is unlikely that RUC committee changes are exogenous to local demand and supply factors.

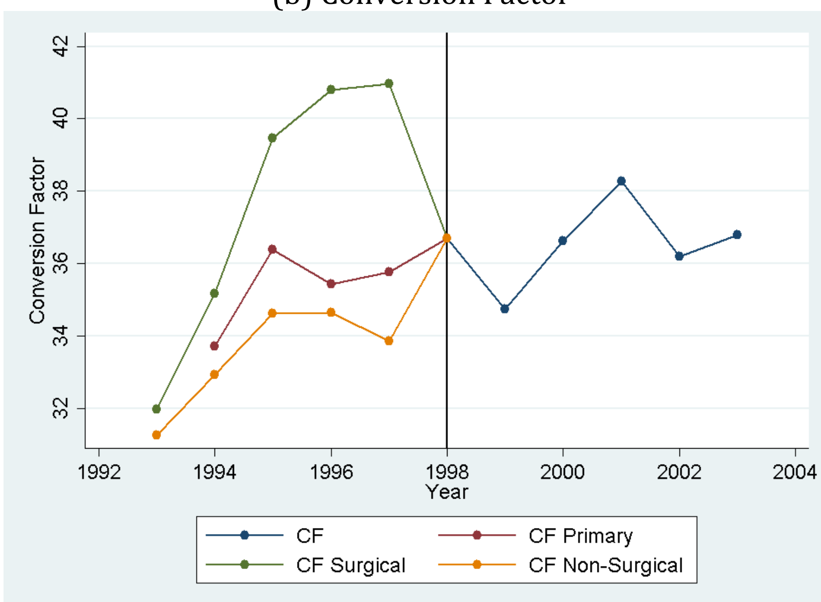
The CF also experienced a major change during the study period. Prior to 1998, there were three different CFs: one for surgery, primary care, and non-surgical services. The CF for surgical procedures led to surgeons earning a 17% bonus payment relative to all other procedures. This generated political discontent and led to a budget-neutral merger of CFs in 1998 (Clemens and Gottlieb, 2013). Plot (b) shows the CFs over time. After 1998, the CF for surgical procedures fell by about 11%, whereas the CF for non-surgical procedures increased by about 6%. We do not use this policy shock as another instrument for two reasons. First, CFs are constant across all geographic regions and all procedures, so their explanatory power for payment changes within in market area for a given HCPCS is weak. Second, the shock in CF payments occurs mainly for surgical procedures, while changes in CF for non-surgical and primary care procedures are much less pronounced.

Figure A1: Remaining Variation in Medicare Payments

(a) Average Work RVU Over Time

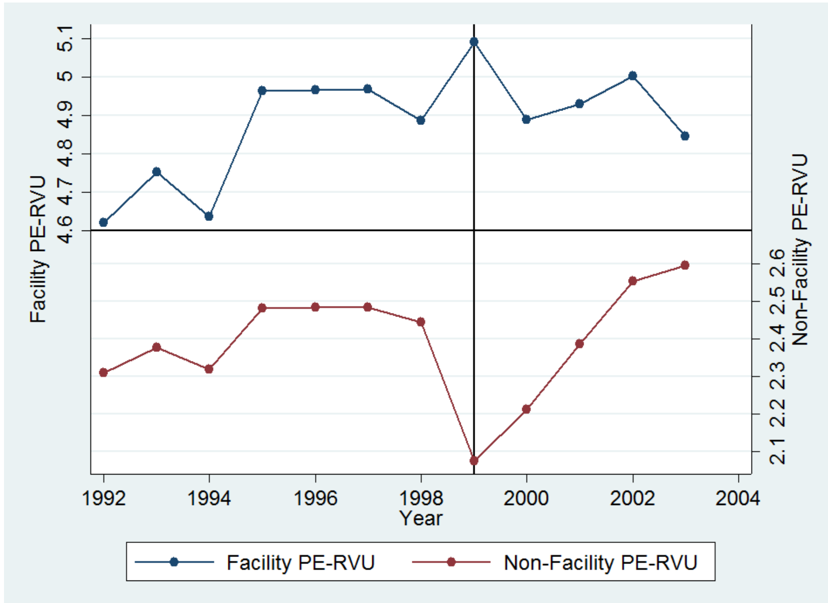


(b) Conversion Factor



Notes: Data from Federal Register 1992-2003. Plot (a) show the change in work-RVUs. Evident from the graph are the two major reviews by the RUC committee in 1997 and 2002. The sample is restricted to HCPCS observed in all years. Plot (b) shows the change from three CFs (primary care, surgical, and non-surgical) to a single budget-neutral CF in 1998.

Figure A2: Practice Expense RVU, by Facility Over Time



Notes: Data from the Federal Register 1992-2003. The top line shows changes in the facility PE-RVU. The bottom line shows changes in the non-facility PE-RVU. Sample restricted to HCPCS observed in all years.

Appendix B

In this section, we show additional results for the paper. In Table B1, we administer another test that physician-driven HCPCS are associated with higher price-cost margins. For each HCPCS, we calculate two elasticities: one that uses profitability changes above the median, and another that uses changes below the median. Elasticities measured from larger changes in profitability should have a higher probability of being positive. Table B1 shows that when the allowed and submitted charges become less negative—in other words, when the physician incurs a lower cost of administering a procedure—HCPCS have a 0.04 to 0.14 higher probability of being physician-driven. This test provides evidence in support of Conjecture #3.

The remaining tables and figures in this Appendix show results when using either PE-RVU or GAF policy shocks as the sole instruments. Figure B1 shows estimated elasticities using either instrument; both subplots indicate both positive and negative elasticities.

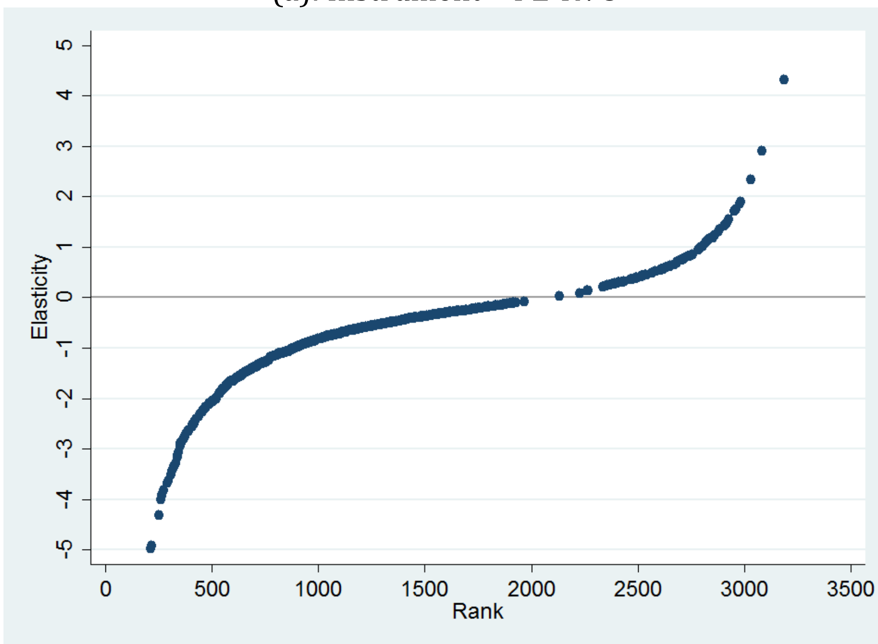
Table B1: Probability of Being Physician-Driven, by Changes in "Profitability"

	OLS		2SLS	
	$\Delta\pi$ Below Median (1)	$\Delta\pi$ Above Median (2)	$\Delta\pi$ Below Median (3)	$\Delta\pi$ Above Median (5)
1(Physician-Driven)	0.525	0.589*	0.777	0.639***

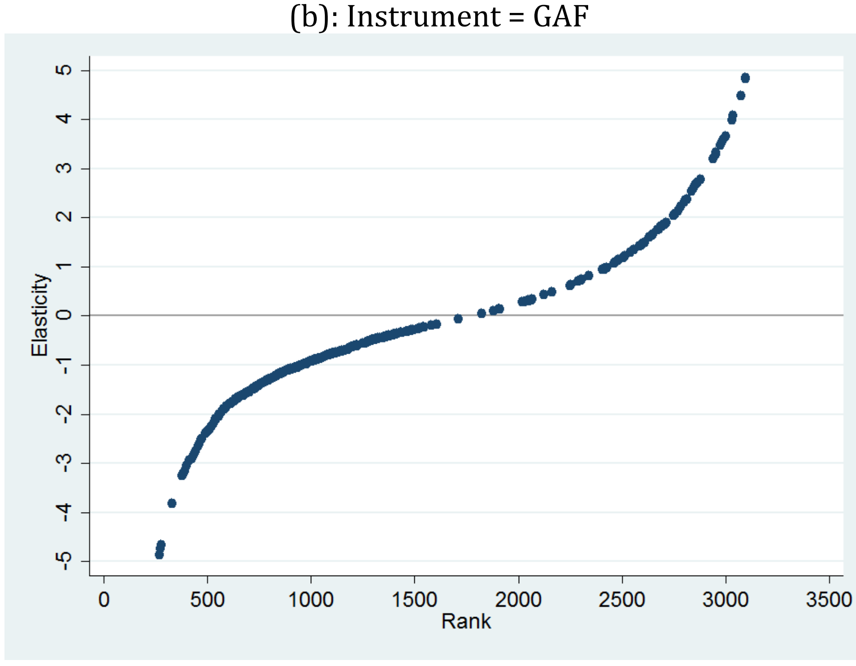
Notes: Column (1) and (3) shows the probability that the own-price elasticity, calculated using changes in annual profitability that are below the median, is positive. Columns (2) and (4) show the probability that the elasticity, calculated using changes in profitability above the median, is positive. Above- and below- median are identified according to the data for each HCCPS-HRR. The means in columns (1) and (2) or (3) and (4) are statistically different at the ** 5% level or * 10% level. Profitability is measured using the "allowed-submitted" measure.

Appendix Figure B1: IV-Estimated Price Elasticities

(a): Instrument = PE-RVU



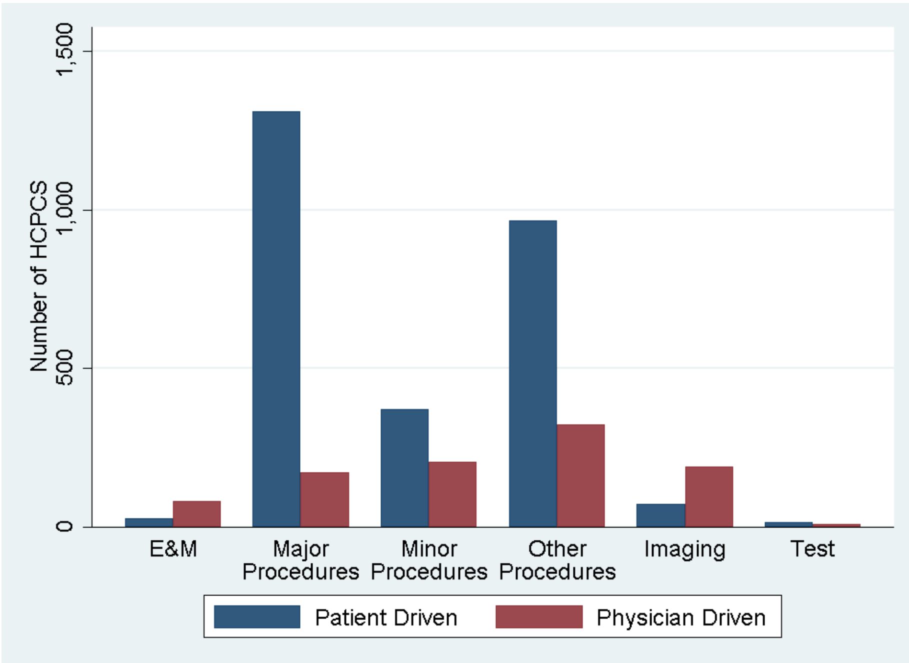
(b): Instrument = GAF



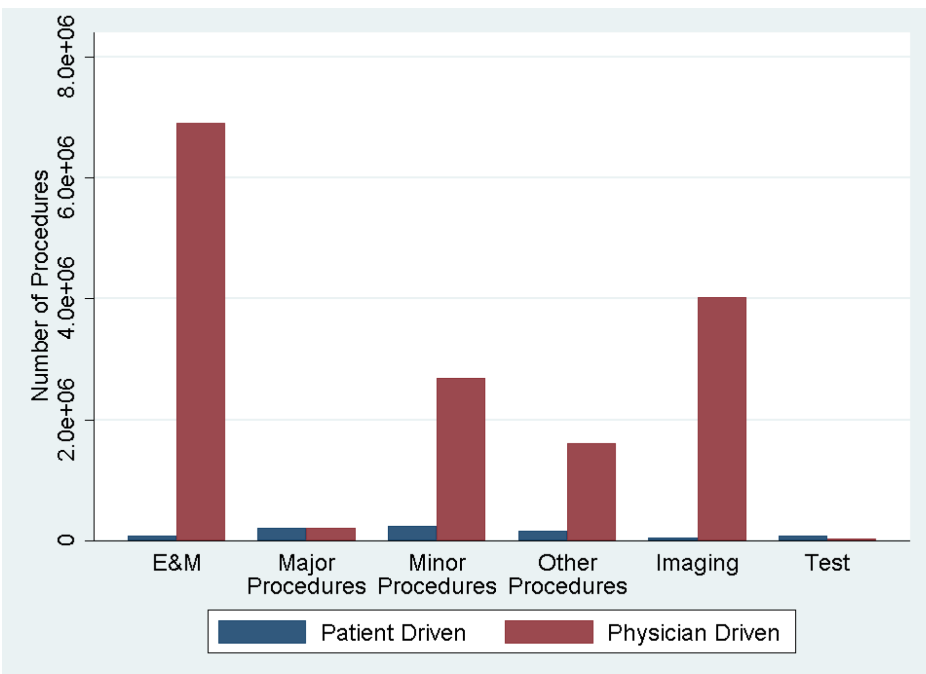
Notes: Each plot shows IV-estimated elasticities where the instrument is either PE-RVU, GAF, or CF. Only statistically significant elasticities at the 5% level are shown. Elasticities are truncated at +/-5.

Appendix Figure B2: Patient- and Physician-Driven Services Using OLS Estimates

(a): By BETOS Category



(b): By Place of Service



Notes: These plots are equivalent to Figure 4, except the patient- and physician-driven HCPCS are defined using OLS estimates.

Appendix Table B1: Elasticities for Low-Value Care Using OLS

Dep. Variable: Log(Q)	Coefficient	Standard deviation	# of Obs.
	A. Diagnostic and preventative testing		
Colorectal cancer screening	-0.165***	(0.0245)	3,008
Preoperative			
Chest radiography	0.260***	(0.0372)	2,674
Echocardiography	0.00258	(0.0488)	1,108
Stress testing	0.0681***	(0.0143)	3,003
	Imaging		
CT of sinus	0.232***	(0.0305)	2,954
Imaging evaluation	-0.0296	(0.0336)	2,865
Imaging headache	0.306***	(0.0451)	2,873
Electroencephalogram	0.168***	(0.0324)	3,003
Imaging back	0.246***	(0.0302)	3,003
	0.271***	(0.0639)	2,951
	Cardiovascular testing and others		
Asymptomatic screening	0.215***	(0.0213)	3,004
Syncope screening	0.249***	(0.0401)	3,002
Coronary stress test	0.404***	(0.0337)	3,003
Coronary balloon test	0.492***	(0.0346)	3,003
Coronary renal stent	-0.0244	(0.0556)	2,980
Endarterectomy	0.0395	(0.0326)	1,728
Arthroscopic surgery	-0.156*	(0.0897)	2,416

Notes: Data from CCF at the service-year-HRR level. See notes to Table 3 which shows the corresponding IV regressions.

Appendix Table B2: Descriptions of Low Value Care

Abbreviated Description	**Full Description**
1. Colorectal cancer screening	1. Colorectal cancer screening for adults older than age 85
2. Preoperative chest radiography	2. Preoperative chest radiography for patient undergoing select surgeries
3. Preoperative echocardiography	3. Preoperative echocardiography for patient undergoing select surgeries
4. Preoperative stress testing	4. Preoperative stress testing for patient undergoing select surgeries
5. CT of sinus	5. CT of the sinuses for uncomplicated acute rhinosinusitis
6. Imaging evaluation	6. Head imaging in the evaluation of syncope
7. Imaging headache	7. Head imaging for uncomplicated headache
8. Electroencephalogram	8. Electroencephalogram for headaches
9. Imaging back	9. Back imaging for patients with non-specific low back pain
10. Asymptomatic screening	10. Screening for carotid artery disease in asymptomatic adults
11. Syncope screening	11. Screening for carotid artery disease for syncope
12. Coronary stress test	12. Stress testing for stable coronary disease in patients with ischemic heart disease
13. Coronary balloon test	13. Percutaneous coronary intervention with balloon angioplasty or stent placement for stable coronary disease
14. Coronary renal stent	14. Renal artery angioplasty or stenting for patients with hypertension
15. Endarterectomy	15. Carotid endarterectomy in asymptomatic patients
16. Arthroscopic surgery	16. Arthroscopic surgery for knee osteoarthritis

Appendix Table B3: Differential Effect of a Physician Reimbursement Reform, OLS

	Dependent Var: Log(RVU)		
	Physician-Driven HCPCS (1)	Patient-Driven HCPCS (2)	X^2 and P-value for H_0: (1)=(2) (3)
	Panel A. All Beneficiaries		
Log(Price)	0.496***	-0.0019	21.12
	(0.105)	(0.0273)	0
	Panel B. Dual Eligibles		
Log(Price)	0.389***	-0.0200	9.90
	(0.105)	(0.0766)	0.0017
	Panel C. Non-Dual Eligibles		
Log(Price)	0.542***	-0.0148	26.81
	(0.103)	(0.0293)	0
No. of Obs.	3,351	723	----

Notes: Data from CCF. Standard errors are bootstrapped. OLS estimates are shown. 2SLS counterpart shown in Table 6.