Preliminary and incomplete; please do not circulate. Comments are very welcome

Voluntary Regulation: Evidence from Medicare Payment Reform*

Liran Einav† Amy Finkelstein‡ Yunan Ji§ Neale Mahoney¶

October 8, 2019

Abstract

Government programs are often offered on an optional basis to market participants. We explore the economics of such voluntary regulation in the context of a Medicare payment reform, in which one medical provider receives a single ("bundled") payment for a sequence of related healthcare services, instead of separate service-specific payments. The program was originally implemented as a 5-year randomized trial, with mandatory participation by hospitals assigned to the new payment model, but after two years participation was unexpectedly made voluntary for half of these hospitals. Using detailed claim-level data we document that voluntary participation is more likely for hospitals who can increase revenue without changing behavior ("selection on levels") and for hospitals that had large changes in behavior when participation was mandatory ("selection on slopes"). To assess outcomes under counterfactual regimes, we estimate a simple model of responsiveness to and selection into the program. We find that the current voluntary regime generates inefficient transfers to hospitals and reduces social welfare compared to the status quo, but that alternative (feasible) designs could be welfare improving. Our analysis highlights key design elements to consider under voluntary regulation.

*We thank Parag Pathak, Jonathan Skinner, and seminar participants at Chicago Booth, Georgetown University, Hebrew University, Ohio State University, Tel Aviv University, University of California San Diego, and Wharton for helpful comments. We gratefully acknowledge support from J-PAL North America's Health Care Delivery Initiative (Finkelstein and Mahoney), the National Institute of Aging grant P01AG019783-15, the Laura and John Arnold Foundation (Einav, Finkelstein and Mahoney), the Becker Friedman Institute at the University of Chicago (Mahoney) and the National Science Foundation SES-1730466 (Mahoney).

†Stanford University and NBER. Email: leinav@stanford.edu

‡MIT and NBER. Email: afink@mit.edu

§Harvard University. Email: ji@g.harvard.edu

¶Chicago Booth and NBER. Email: neale.mahoney@gmail.com

1 Introduction

Government intervention is designed to move market actors away from market equilibrium. Yet, many government programs, across a variety of contexts, allow these actors to voluntarily decide whether they would like to participate in the government program.[1] There are a number of reasons why voluntary programs are popular. From a political perspective, voluntary programs may face less opposition from industry or consumer lobbies, since their members only need to sign up if they benefit. Voluntary programs may also be more palatable to those with an ideological aversion to government mandates and a preference for regulatory "nudges" (Thaler and Sunstein, 2003).

The key economic benefit of voluntary programs – i.e., "choose your own incentives" – is that they might generate favorable selection. If those who enroll have private information about their net benefits from changing behavior, then the resulting "selection on slopes" – also known as selection on gains or Roy selection (Heckman and Honore, 1990) – might result in selection into the program by those with the highest net social benefits. However, if voluntary programs attract participants who, without changing their behavior at all, can simply receive a higher government transfer, the resulting "selection on levels" could lead to higher program costs without the desired change in behavior. Thus, the extent to which voluntary policies are more or less socially desirable depends critically on the nature and extent of selection into the program.

We explore these tradeoffs empirically in the context of voluntary regulation within the U.S. Traditional Medicare program. Over the last decade, Medicare has rapidly increased the use of alternative payment models (APMs) by which Medicare reimburses healthcare providers; examples include Accountable Care Organizations, primary care coordination models, and bundled payment models. By 2016, over 30% of Traditional Medicare spending was based on these APMs (Shatto, 2016). Although provider participation has been voluntary in all of these payment models, with the partial exception of the specific payment model we study in this paper (GAO, 2018), there is an ongoing and active debate over whether these programs should be made mandatory (Gronniger et al., 2017; King, 2019; Frakt, 2019; Levy, Bagley and Rajkumar, 2018).

[1] Examples include the provision of environmental services by landowners in developing countries (Jack and Jayachandran, 2019), the acceptance of public vouchers by private schools (DeAngelis, Burke and Wolf, 2018), or the choice given to electricity consumers of whether to be charged on a constant or a time-varying basis (Ida, Ito and Tanaka, in preparation).

We focus on the Medicare bundled payment program for hip and knee replacement, known as Comprehensive Care for Joint Replacement (CJR). Hip and knee replacement is a large category, with almost half a million procedures in 2014, which accounted for \$10.7 billion in Traditional Medicare spending. Under bundled payments, Medicare makes a single payment to the hospital for all services related to the episode of care, including the initial hospital stay and subsequent care by other medical providers during the recovery period, rather that making separate payments that are based on the provider and the amount of care. The idea is that by making the hospital the residual claimant on all costs related to the entire episode of care, it will internalize the incentives to provide care efficiently, including coordination with downstream providers.

CJR was initially designed as a mandatory participation 5-year randomized trial, which started in Apri 2016. Randomization was conducted at the Metropolitical Statistics Area (MSA) level. In the 67 treatment MSAs, hospitals were paid under the bundled payment program. In the 104 control MSAs, hospitals were paid under the status quote Fee-for-Service (FFS) system. However, toward the end of the second year of the program, Medicare unexpectedly announced that participation would be made voluntary in half the treated MSAs (Centers for Medicare & Medicaid Services, 2017), and about three-quarters of the affected hospitals subsequently opted out. This unusual set of circumstances provides a rare opportunity to assess a voluntary program while observing behavior under the program even for participants who eventually choose not to participate.

We begin by providing descriptive evidence on the mandatory and voluntary regimes. In the mandatory regime, we closely follow prior analyses and find that bundled payments caused, on average, a modest reduction in Medicare spending, driven predominantly by reduced discharges to post-acute care (PAC) facilities (Finkelstein et al., 2018; The Lewin Group, 2018; Barnett et al., 2019; Haas et al., 2019). We then use the voluntary regime to examine the nature of selection into the program. Consistent with selection on levels, we find that hospitals with lower baseline spending, who would receive larger payments holding behavior fixed, are more likely to opt into the program. Consistent with selection on slopes, we also find that hospitals that achieved larger reductions in spending when the program was mandatory are also more likely to opt in when it becomes voluntary.

Motivated by these patterns, we specify and estimate a simple model of responsiveness to and selection into the bundled payment program, and use the model to analyze the social cost of the observed voluntary bundled payment model, and to assess selection and social costs under alternative bundled payment designs. In the model, hospitals are characterized by a hospital-specific baseline expenditure "level" and a hospital-specific response to the program ("slope"). Under a voluntary regime, the selection decision depends on a hospital-specific "target price" – the bundled payment the hospital receives under the program – as well as on the hospital's level and slope parameters. The random assignment in years 1-2 of the program, when participation was mandatory, identifies the levels and slopes, and the voluntary decision in year 3 identifies the selection equation.

We estimate average episode spending under the status quo FFS regime to be about $24,000 – on average, only slightly ($50) lower than the target price – and average spending reductions caused by bundled payments of about $500 per episode. The estimated impact of the program is consistent with prior reduced form estimates (Finkelstein et al., 2018; The Lewin Group, 2018; Barnett et al., 2019; Haas et al., 2019). These averages, however, mask substantial heterogeneity across hospitals in both levels and slopes; heterogeneity is particularly large in levels, where the standard deviation in baseline spending across hospitals is about $4,000. Observed target prices do not come close to capturing this heterogeneity, thus making selection on levels the key driver of the participation decision.

We use the model and its estimates to compare outcomes and social welfare under the observed voluntary and mandatory programs, as well as to assess behavior and social welfare under alternative bundled payment designs. We define social welfare as the sum of consumer surplus and producer (hospital) profits, minus the social cost of public funds. The latter is defined as government (i.e. Medicare) expenditures multiplied by the shadow cost of public funds. This (standard) cost of public funds' generates the key tradeoff in designing a voluntary bundled payment model: higher target prices will induce more hospitals to participate and increase productive efficiency, but will involve higher government spending, which is socially costly. Producer surplus and government spending can be calculated directly from the data and estimated model parameters. We assume that consumer surplus is not affected by the payment regimel this is consistent with the reduced form evidence from the randomized trial that care quality, patient mix, and patient volume did not change with bundled payments (Finkelstein et al., 2018; The Lewin Group, 2018; Barnett et al., 2019; Haas et al., 2019). We make the (conservative) assumption of a shadow cost of public funds of 0.15.

We estimate that, relative to the FFS status quo, the voluntary bundled payment regime lowers social surplus by $149 per episode; by contrast, the mandatory bundled payment regime raises social surplus by $242 per episode. The reduction in social welfare from the voluntary regime is due to substantial selection on levels, which generates inefficient transfers from the government to hospitals. Average baseline spending of hospitals who opt into the voluntary regime is XX lower than those who do not, and the magnitude of the favorable selection (on "slopes") is too small to offset it.

We also show how improved targeting could change the net social welfare impact of the voluntary regime, by reducing the transfers and aligning better participation incentives. Specifically, our findings suggest that a perfectly targeted voluntary bundled payment regime would generate $73 per episode in social surplus relative to the status quo, compared to the $-149 per episode social loss created by the observed voluntary bundled payment program. While perfect targeting may be difficult to achieve in practice, we also report results from more feasible designs and show that improved targeting – either by exploiting information more efficiently or narrowing the definition of the bundle – can come close to generating the social gains from perfect targeting.

Our paper relates to several distinct literatures. Most narrowly, it contributes to the literature on the impact of Medicare bundled payment programs. This includes several recent evaluations of the first two years of the randomized mandatory participation, bundled payment program for hip and knee replacement that we study here (Finkelstein et al., 2018; The Lewin Group, 2018; Barnett et al., 2019; Haas et al., 2019), as well as evaluations of the much larger number of voluntary participation bundled payment programs for a variety of conditions including coronary bypass, prenatal care, cancer, and hip and knee replacement.[2] It is well-understood that non-random selection into voluntary models can bias the estimated impact of the program (Gronniger et al., 2017; Levy, Bagley and Rajkumar, 2018). Our focus here, however, is how voluntary participation affects the actual impact of the program, rather than the estimated impact.

More broadly, our emphasis of the potential for selection not only on levels but also on slope

[2]See Cromwell, Dayhoff and Thoumaian (1997), Carroll et al. (2018), Newcomer et al. (2014), Doran and Zabinski (2015), Dummit et al. (2016), Froemke et al. (2015), Navathe et al. (2017)

relates to work in labor economics on selection on gains (Heckman and Honore, 1990), as well as to the papers discussed above analyzing voluntary regulation in settings as diverse as consumer electricity, landowner deforestation, and education vouchers (see foonote 1). Within health economics, our paper relates most closely to work on so-called “selection on moral hazard” – i.e. consumer selection of health insurance plans based not only on levels but on slopes (Einav et al., 2013, 2016; Shepard, 2016); it also contributes to a growing literature on the impact of financial incentives on healthcare provider behavior and healthcare spending (e.g., Cutler, 1995; Clemens and Gottlieb, 2014; Ho and Pakes, 2014; Einav, Finkelstein and Mahoney, 2018; Eliason et al., 2018)

The rest of the paper proceeds as follows. Section 2 provides background on our setting. Section 3 describes the data and present reduced form, descriptive evidence of the the impact of bundled payment under mandatory participation – both on average as well as heterogeneity across hospitals – as well as the nature of hospital selection once the program became voluntary. Section 4 presents a stylized model of selection into a voluntary bundled payment program. Section 5 presents the econometric specification of the model and describes its identification and estimation. Section 6 presents our main results. The last section concludes.

2 Setting

2.1 Medicare Bundled Payment Programs

Medicare is U.S. public health insurance program for the elderly and the disabled. We focus on the Traditional Medicare program, which provides coverage to about two-thirds of enrollees. Traditional Medicare (hereafter "Medicare") has 38.7 million enrollees and annual expenditures of \$377 billion, as of 2017 (CMS, 2019).

Throughout most of its history, Medicare has paid providers on a cost-plus basis referred to as Fee-for-Service (FFS), in which providers are reimbursed based on claims submitted for services. For instance, for a patient undergoing hip replacement, Medicare might make separate payments to the hospital for the initial hospital stay, the surgeon for performing the procedure, and the skilled nursing facility for post-acute care, as well as additional payments to the surgeon for each post-operative visit, or for renting a wheelchair during the recovery period. Moreover, within most of these categories, the payment would depend on the specific services provided.[3]

[3]One exception to this system is hospital reimbursements, which adopted the Prospective Payment System (PPS)

Over the last decade, Medicare has responded to concerns that the FFS system may encourage excessive healthcare use by attempting to shift providers towards Alternative Payment Models (APMs), such as Accountable Care Organizations (ACOs), bundled payments, and primary care coordination models. By 2016, over 30% of Traditional Medicare spending was based on these APMs (Shatto, 2016).

Our focus is on bundled payments, which represent a middle ground between FFS and capitated models, such as ACOs, in which providers are paid a fixed per capita amount per annum. Under bundled payments, Medicare makes a single payment for all services related to a clearly-defined episode of care. Episodes typically start with an acute-care hospital stay (e.g., for hip replacement surgery) and include most subsequent care during the recovery period. The payments are sometimes adjusted to reflect predictable variation in patient health or in costs in the local medical market. The contracts may also be structured to limit risk exposure for the hospital.

Proponents of bundled payments argue that by providing a single, fixed reimbursement, bundled payments will improve coordination of care and reduce unnecessary healthcare utilization. Yet, some are concerned that because providers do not receive marginal payments, they may cut back on necessary care or cherry-pick patients who have a lower cost of provision (Cutler and Ghosh, 2012; Fisher, 2016).

Most prior studies of bundled payments have been observational, focusing on the experience of a small number of hospitals that voluntarily participated. Many of these studies have found large savings associated with bundled paymentsfootnoteFor instance, there has been studies on bundled payments for coronary bypass (Cromwell, Dayhoff and Thoumaian, 1997), prenatal care (Carroll et al. 2017)., cancer (Newcomer et al., 2014), and hip and knee replacements (Doran and Zabinski, 2015; Dummit et al., 2016; Froemke et al., 2015; Navathe et al., 2017) However, voluntary participation makes separating treatment from selection difficult, and the small number of participating hospitals raises concerns about generalizability (Gronniger et al., 2017; King, 2019).

2.2 Comprehensive Care for Joint Replacement (CJR)

In this paper we focus on the Medicare bundled payment program for hip and knee replacement, known as Comprehensive Care for Joint Replacement (CJR). Hip and knee replacement (also re-

starting in 1982 and makes a fixed payment to the hospital for the hospital stay based on the patient's diagnosis related group (DRG) (Cutler, 1995).

ferred in the medical literature as lower extremity joint replacement, LEJR) is a large Medicare category; in 2014, the year before CJR was announced, Medicare had 486 thousand LEJR procedures, which accounted for \$6.2 billion in Medicare inpatient spending (about 4.5 percent of all Medicare inpatient spending in 2014) (Finkelstein et al., 2018).

Under CJR, an episode begins with a hospital stay in a qualifying diagnosis related group (DRG) and ends 90 days after hospital discharge. Hospitals are paid a fixed *target price*, and are responsible for medical spending over the entire episode (except for spending that is deemed as obviously unrelated). By contrast, under the status quo FFS regime, hospitals are paid a fixed amount for the hospital stay, while the surgical procedure and post-discharge care are reimbursed separately.

The level and targeting of the target price are key design elements in a bundled payment program. We denote by $t_{i,h}$ the target price for episode i at hospital h, let $y_{i,h}$ denote claims for this episode, and let t_h and y_h denote the average target price and average claims at the hospital. Under FFS, Medicare pays y_h on average. Under bundled payment, Medicare pays t_h on average.[4]

Under CJR, hospitals receive their annual hospital-specific target prices $t_{j,h}$ before each program year for each of four severity groups $j = 1, \ldots 4$, where we have suppressed the year subscript for notational simplicity. These four groups are determined by the 2-by-2 interaction of the patient's MS-DRG (469 or 470) and whether the patient had a hip fracture. Medicare has tried to set the target price for each severity group to be slightly lower than expected per-episode spending under FFS. Specifically, the target price included a small discount off a weighted average of historical hospital and regional (defined by the 9 census divisions) per-episode expenditures from three prior reference years, with the weight on regional expenditures increased over time from one third in the first two years of the program to 100% in the last two years;[5] The ''discount factor" was designed to reflect Medicare's portion of expected savings from CJR [6]. This design ensured that, if spending remained at past levels, Medicare spending would decrease under mandatory

[4]More specifically, under bundled payment, providers continue to be paid y_h on average over the year, as they would be under FFS. However, at the end of the year, hospitals receive a "reconciliation payment" of $t_h - y_h$ on average, so that the gross payment is t_h.

[5]The three reference years of historical spending are updated every other year. In 2016 and 2017 (the first two years of the program) historical spending from 2012 to 2014 was used, in 2018 and 2019 (the third and fourh program years), historical spending from 2014-2016 was used, and the final year of the program uses spending from 2016-2018 (Centers for Medicare & Medicaid Services, 2015*b*).

[6]The discount factor ranged from 1.5 to 3 percent depending on hospital quality (based on a composite quality score defined below), with smaller discounts for higher quality hospitals (Centers for Medicare & Medicaid Services, 2015*b*)

participation (although in practice there was a steady decline in per-episode cost, offsetting these built-in “discount”).

We abstract in our analyses from two other features of CJR. First, to mitigate concerns that bundled payments would create incentives to shirk on quality, hospitals were only eligible for positive reconciliation payments if they met a minimum quality standard.[7]. However, in practice the quality standard was not binding for the vast majority of the hospitals.[8]

Second, like most bundled payment programs, CJR is not a “pure" bundled payment model that exposed hospitals and Medicare to unbounded risk. Rather, to limit risk exposure, the reconciliation payment was subject to stop-loss and stop-gain provisions. In particular, if $t_h - y_h$ is less than the (weakly negative) stop loss, the hospital “only” has to pay Medicare the stop loss amount and similarly, if $t_h - y_h$ is greater than the stop gain, Medicare “only” has to pay the hospital the stop gain. The stop-loss and stop-gain amounts increased over time. In the first year, the stop-gain amount was set as 5 percent of t_h and the stop-loss was zero (meaning that hospitals would not need to make payments to Medicare). By the fifth year, the stop-gain and stop-loss amounts were each scheduled to be set at 20 percent of t_h. As we discuss more below, these do not affect the qualitative economic analysis and (we suspect) are unlikely to be quantitatively important, although we may try to incorporate them in future versions.

2.3 Experimental Design

CJR was initially designed by CMS as a 5-year, mandatory participation, randomized trial. Year 1 was defined as April 1 to December 31 of 2016, and years 2-5 were defined as the 2017-2020 calendar years. CMS randomized 196 eligible MSAs into treatment (bundled payment) or control (status quo FFS). Specifically, MSAs were divided into 8 strata based on the interaction of historical LEJR spending quartile and above- vs below- median MSA population, and MSA treatment probabilities varied by strata (ranging from 30% to 45%), with higher treatment probabilities for strata

[7]The quality standard is based on a composite quality score, which ranges from 0 to 20 points, and hospitals must score at least 5 points to be eligible for bonus payments. Up to 10 points are given based on a hospital’s quality performance percentile on a complication measure for total hip arthroplasty and total knee arthroplasty; up to 8 points are given based on a standardized national patient experience survey; up to 2 points are given for submitting the patient-reported outcomes and risk variable data. Finally, up to 1.8 points can be added to the final score for improvement in either of the first two measures relative to the previous performance year, as long as the final score does not exceed 20. See `https://innovation.cms.gov/Files/x/cjr-qualsup.pdf` for details

[8]For example, based on our calculation from the CJR reconciliation data, in the first year of the program fewer than 9% of treatment hospitals failed to meet the minimum quality standard for receiving a bonus.

with higher historical LEJR payments. CMS announced assignment to treatment and control in the July 2015 Federal Register (Centers for Medicare & Medicaid Services, 2015*b*). In Finkelstein et al. (2018), we show that treatment and control MSAs are balanced on outcome variables and MSA characteristics. After exclusions, the program covered 67 treatment MSAs and 104 control MSAs.[9] Within the 171 MSAs assigned to treatment or control, a small number of hospital types and episode types were further excluded from eligibility Finkelstein et al. (2018).

Participation was mandatory in treatment MSAs: eligible hospitals had no choice but to be reimbursed under the new bundled payment model. This mandatory participation feature was immediately controversial, with then-US representative Tom Price spearheading a letter in 2016, signed by 179 members of Congress, complaining that mandatory participation was unethical and unauthorized.[10] Subsequently, as the new Secretary of Health and Human Services – the federal agency charged with overseeing Medicare – Price spearheaded the effort to roll back mandatory participation bundled payment models. As a result, in a rule finalized in December 2017,Medicare unexpectedly decided to cancel two previously scheduled mandatory bundled payments models (Advancing Care Coordination Through Episode Payment and Cardiac Rehabilitation Incentive Payment Models) and modified CJR to be voluntary in half of treated MSAs for the remaining three program years (Centers for Medicare & Medicaid Services, 2017).

In the 33 MSAs with the lowest historical spending, each hospital had to make a one-time decision at the beginning of program-year 3 of whether to opt in, and continue to be paid under bundled payments. If they did not opt in, reimbursement would revert to FFS for the remaining three years. 73 hospitals – about one-quarter of the 279 hospitals in the voluntary bundled payment MSAs – chose to remain under bundled payments. In the 34 mandatory MSAs, hospitals did not face a choice and continued to be paid under bundled payments. Control group hospitals were unaffected by this change, and continued to be paid under FFS.

In the analysis that follows, we define three time periods. Period 1 is the period prior to bundled payments, when all providers were reimbursed under FFS. Period 2 covers the approxi-

[9] After the initial assignment, Medicare realized that they did not exclude some hospitals that were already (prior to assignment) signed up for BPCI (a different Medicare program), and subsequently excluded an additional 8 MSAs from the treatment group. Medicare later identified the 17 MSAs in the control group that would have been excluded based on this criteria. Since these exclusions were based on hospital decisions made prior to assignment we simply drop these 25 MSAs from the study.

[10] See `http://ascrs.org/legislative-and-regulatory/www/article/tom-price-along-other-ways-means-committee-members-spearhead-letter-cms-insisting-cmmi-stop`.

mately two years when the mandatory participation regime was in effect and expected to remain so. Period 3 is defined as the final 3 years of the program, when the program was voluntary for some hospitals.

Figure 1 provides a visual description of the experimental design. The top part of the figure shows the initial assignment to treatment and control for period 2, when the program was mandatory, and the bottom part shows period 3, where treatment MSAs where further divided into mandatory and voluntary treatment groups. Because this division was based on pre-determined historical MSA spending, we can analogously divide the control MSAs into mandatory and voluntary control MSAs based on this variable. For some of the subsequent analysis, we will compare mandatory treatment to mandatory control, and voluntary treatment to voluntary control.

3 Data and Descriptive Evidence

In this section, we describe the data and sample, and reproduce prior findings of the average impacts of bundled payments during the mandatory participation period. We then present evidence on selection on levels and slopes during the voluntary period. These patterns motivate our subsequent modeling decisions.

3.1 Data and Sample

The main data are the 100% Medicare enrollment and claims files. These contain basic demographic information on all enrollees (age, race, sex, and Medicaid enrollment) as well as claims for any inpatient, outpatient, and post-acute care provided.[11] The claims data include information on Medicare payments made and out-of-pocket payments owed, dates of admission and discharge, diagnoses, and discharge destinations.

We supplement these data with several ancillary data sources. First, we obtained data from the CJR website on the eligibility and treatment status of each hospital in each year, including its annual target price (for 2016 and 2017), annual reconciliation payments (for 2016 and 2017), and whether the hospital opted into bundled payment when it became voluntary in 2018.[12,13] Second, we use data from the 2016 American Hospital Association (AHA) annual survey to measure the

[11] Specifically, we use the following claims files: Inpatient, Outpatient, Carrier, Skilled Nursing Facility, Home Health Agency, Durable Medical Equipment, and Hospice.

[12] https://innovation.cms.gov/initiatives/CJR

[13] Target prices and reconciliation payments for 2018 will be added when the become available.

number of beds in each hospital, whether the hospital is for-profit, non-profit, or government owned, and the teaching status of the hospital. Third, we obtained data from Hospital Compare on each hospital's official quality measures (for 2016 and 2017), which is used to determine eligibility and the discount factor for reconciliation payments.[14]

We limit our analysis sample to the 171 eligible MSAs and, within these MSAs, to hospitals and episodes that were eligible for CJR. MSAs were excluded primarily due to low volume of hip and knee replacements. Within both treatment and control MSAs, hospitals were excluded from CJR if they were already participating in a pre-existing Medicare voluntary bundled payments model for LEJR.[15] Episodes were excluded if the patient did not have Medicare as the primary payer, was readmitted during the episode for LEJR, or died during the episode.[16]

We define period 2 (the period of mandatory participation bundled payment) to include all episodes admitted between April 1, 2016 and September 15, 2017. The start date corresponds to the program start date, and the end date was chosen so that nearly all 90-day episodes would finish by December 31, 2017, the close of the second year.[17] The end date also ensures that all admissions (and most discharges) occurred prior to the December 2017 announcement that participation would become voluntary for some MSAs starting on January 1, 2018 (Centers for Medicare & Medicaid Services, 2017). Following our prior work (Finkelstein et al., 2018), we define period 1 (pre period) to include all episodes admitted between April 1, 2013 and September 15, 2014. We omit 2015 from the analysis to avoid contamination from potential anticipatory effects; treatment and control MSAs were announced in July 2015 (Centers for Medicare & Medicaid Services, 2015*a*).

To construct our baseline sample, we start with the universe of 395,938 CJR episodes that occurred in period 2 in the 171 treatment and control MSAs; these episodes fall into 1,570 hospitals. In order to estimate our OLS specification, where we follow our prior work and control for the lags

[14] https://www.medicare.gov/hospitalcompare

[15] Model 1 or Phase 2 (Models 2 or 4) of the Bundled Payments for Care Improvement Initiative (BPCI).

[16] Finkelstein et al. (2018) provides more detail on these eligibility criteria and estimates that the eligible MSAs represented about 70 percent of all Medicare LEJR patients, and that within eligible MSAs, about 70 percent of LEJR patients were eligible for CJR. About 20 percent of LEJR patients within eligible MSAs were excluded because their hospital was not eligible, and another approximately 10 percent due to patient eligibility.

[17] Recall that the episode of care ends 90 days after hospital discharge. The mean length of stay for an LEJR admission is 3.1 days for DRG 470 and 7 days for DRG 469, so truncating the sample on Septmeber 15, 2017 guarantees that all claims associated with the episode would be included in the 2017 claim files (Centers for Medicare and Medicaid Services, 2016).

of the outcome variable from two prior years, we restrict the sample to the 1,455 hospitals that have at least one episode in both years of period 1 (Finkelstein et al., 2018). These 1,455 hospitals constitute our baseline sample, out of which 664 hospitals are located in a treatment MSA and 791 hospitals are in a control MSA. A total of 379,843 CJR episodes in period 2 fall into these 1,455 hospitals.

3.2 Average Treatment Effects

Average effects of CJR in the two-year mandatory participation period (period 2) have been well-studied (Finkelstein et al., 2018; The Lewin Group, 2018; Barnett et al., 2019; Haas et al., 2019); we reproduce some of the key results here. Since the program was mandatory and assignment was random at the MSA level, we simply estimate:

$$y_{j,2} = \beta_0 + \beta_1 BP_j + \beta_2 y_{j,2014} + \beta_3 y_{j,2013} + \delta_{s(j)} + \epsilon_j \tag{1}$$

where $y_{j,2}$ is the average per-episode outcome in MSA j and period 2, BP_j is an indicator for being randomly assigned to bundled payments and β_1 is the average treatment effect of bundled payment. As in Finkelstein et al. (2018), we include lagged outcomes from period 1 as controls to improve statistical power and, because randomization was conducted within strata, we include strata fixed effects, $\delta_{s(j)}$, to isolate the experimental variation. In all tables, we report heterogeneity robust standard errors.[18]

Table 1 shows the average treatment effects. To provide a baseline, the first two columns show the mean and the standard deviation of the outcome from the control group in period 2; the remaining columns show the average treatment effect, standard error, and p-value of the estimate.

Healthcare Use and Spending Panel A examines effects on health care spending and use. "Episode spending if paid under FFS" consists of eligible Medicare payments and patient cost sharing over the entire episode of care, but does not account for any reconciliation payment associated with bundled payments. Average per-episode spending in the control group was about $25,300, with roughly half of this spending on the index admission, which is already reimbursed as a fixed,

[18] Since MSAs are the unit of randomization, we follow (Finkelstein et al., 2018) and estimate the impact of bundled payment at the MSA level.

DRG-based prospective payment under the status quo. Of the remaining $11,800, about $4,000 comes from post-discharge spending at institutional Post Acute Care (PAC) – which are predominantly skilled nursing homes – $1,800 represents post-discharge spending on Home Health Care, and the remaining $5,800 includes categories such as payments to the suregon and other physicians (both inpatienta and outpatient), hospice, and durable medical equipment.

Bundled payments reduced total episode spending (as would be paid under FFS) by about $800, or about 3 percent, a statistically significant but economically modest result. [19] This reduction is primarily driven by a statistically significant $500 decline in spending on institutional PAC (12 percent of the control mean), with no statistically or economically significant effects on other categories of spending. Bundled payments did not impact average length of stay for the index admission, but decreased the unconditional average number of days spent in institutional PAC by about 0.6 days (7 percent of the control mean).

This decline in use of institutional PAC primarily reflects a large extensive margin response in where patients are discharged to following their index admission. In the control group, roughly one third each are discharged to institutional post acute care, home with home health care, and home without home health care. Bundled payments reduce discharged to institutional PAC by a statistically significant 3.4 percentage points (11%). The decline in discharges to institutional PAC is accompanied by similarly sized increase in discharges to home without home health.[20]

The experimental estimates are consistent with qualitative evidence. In a survey of hospital executives and administers, Zhu et al. (2018) find that hospitals respond to bundled payments by reducing SNF discharges using risk-stratification and home care support, and by forming networks of preferred SNFs to influence quality and costs, conditional on discharge.

[19] The estimated reduction is similar to prior findings of the impact of the first two years of CJR on total episode spending(Barnett et al., 2019). This is substantially smaller than several prior observational estimates from voluntary bundled payment programs for LEJR, which have found spending reductions of up to 21 percent (Navathe et al., 2017). Differences in the results may reflect selection into voluntary programs on both levels and slope. While these are typically viewed as econometric concerns for consistent estimates of average impacts (Gronniger et al., 2017), in our context they are the fundamental economic margin of interest.

[20] This can be either because the patients who would have been sent to institutional PAC are being sent home *without* home health, or because there is a cascading effect where the patients who would have been sent to institutional PAC are being sent home *with* home health, and patients who would have been sent home with home health are now being sent home without home health supports. A cascading effect seems (to us) more likely, but we cannnot differentiate between these two channels.

(Lack of) Cream Skimming and Quality Shirking A primary concern with bundled payment programs is that because providers are no longer paid on the margin, they will cut back on medically necessary care and/or cherry-pick patients who have lower costs of provision. Of course, the quality incentives provided by the program, as well as physician ethics, reputational concerns, and the threat of malpractice lawsuits may limit any quality response. Indeed, to the extent that low quality care creates downstream costs, providers may have incentives to improve care quality. Consistent with prior work on CJR (Finkelstein et al., 2018; The Lewin Group, 2018; Barnett et al., 2019; Haas et al., 2019), the bottom two panels of Table 1 show no evidence of an impact of CJR on quality of care or patient composition.

Panel B examines effects on measures of quality. Specifically, we estimate effects on a clinically-defined complication rate, whether the patient had an emergency room visit during the episode, and 90-day all-cause readmission. We estimate a fairly precise zero effect on all of these measures. Of course, the quality measures are limited. We cannot, for example, measure outcomes such as mobility or activities of daily living consistently across locations. That being said, the precise zero effects on our outcomes suggests limited quality response. Based on this, in our subsequent model and counterfactual exercises, we will assume that quality remains fixed and that patients' utility is unaffected by the hospital's response to incentives.

Panel C examines the impact of bundled payments on admissions and on patient composition. We estimate a precise zero effect on the number of LEJR admissions per 1,000 Medicare enrollees and the number of CJR-eligible admissions. [21] We examine patient composition by estimating effects on the Elixhauser Comorbidity Index of the patient pool, which is constructed as the sum of indicators for 31 different potential comorbidities (Elixhauser et al., 1998; Quan et al., 2005). We estimate a precise zero effect on this measure as well.

The lack of a patient volume response is consistent with LEJR being non-discretionary procedures, or at least procedures where the change in financial incentives from bundled payment is small relative to other determining factors. The lack of cream skimming may also reflect the fact that assignment to bundled payments in period 2 is determined at the MSA level, so that the closest substitutes for a given hospital are likely to be paid under the same regime. Cream skim-

[21] Both of these admissions are analyzed on the sample of 171 eligible MSAs. However, our analysis of CJR-eligible admissions - like most of our analyses - also excludes hospitals and episodes not eligible for CJR, while we include such admissions in analyzing LEJR admissions.

ming responses could potentially be different in period 3 once participation is voluntary, as there may now be hospitals paid under bundled payment and under FFS in the same MSA. Indeed, in a different voluntary payment model, Alexander (2017) documents that physicians strategically direct patients across hospitals within a local area to maximize hospital revenue.

3.3 Selection on levels and slopes

As we formalize in Section 4, hospitals have incentives to select into CJR on both levels and slopes. By selection on levels, we mean that hospitals have a larger incentive to select in if their average claims amount, holding behavior fixed, are below their target price. By selection on slopes, we mean that hospitals have a larger incentive to select in if they can more easily reduce their (per episode) claims below the target price. We present descriptive evidence on both here, looking at how the decision among voluntary treatment hospitals to select in or out of bundled payment in period 3 correlates with spending levels and behavioral responses to bundled payments. Table 2 presents the results.

Selection on Levels Selection on levels is motivated on what spending would be in the absence of bundled payment, relative to the target price. Because treated group hospitals change behavior in response to bundled payments (see Table 1), we do not observe what spending would have been for the treamtent group hospitals in period 3 in the absence of bundled payment. The model developed in the follow sections will allow us to formally recover these counterfactual levels. To provide model free evidence, for now we simply look at how selection correlates with hospital spending in period 1 - the period prior to when the bundled payment program was announced; all else equal, we expect hospitals with lower period 1 spending to be more likely to select in on levels. [22]

Panel A of Table 2 shows how the selection decision varies with period 1 levels. Specifically, we show mean outcomes and their standard deviation for three groups of hospitals: Column 1 shows hospitals in the voluntary control group, which we define as control group hospitals that would been assigned to voluntary based on their prior spending levels (see Figure 1). Columns 2

[22] We show in Appendix Table X that for control group hospitals, there is a strong auto-correlation between period 1 outcomes and period 2 outcomes, with a correlation coefficient that ranges from X to Y across. This suggests that period 1 outcomes are a reasonable proxy for what outcomes would have been for the treatment group hospitals in subsequent periods, had they not responded to the incentives of the bundled payment program.

and 3 shows period 1 outcomes for hospitals in the voluntary treatment group, split by those who in period 3 selected in to bundled payment (column 2) or out of bundled payment (column 3). The first three rows report results for the three outcomes where we observed a statistically significant impact of bundled payment in Table 1: total episode spending, spending in institutional PAC, and share of patients discharged to institutional PAC.

The results are consistent with selection on levels. Column 1 shows substantial heterogeneity across hospitals in the voluntary control group, indicating potential scope for selection. For example, total average Medicare episode spending is \$28,200, with a standard deviation of \$6,900. Columns 2 and 3 show that, as expected, hospitals who select into bundled payment have, on average \$3,000 lower average episode spending than those who select out, a statistically significant difference that is about 10% of the control mean. We see similar patterns for the other outcomes in Panel A.

Selection on Slopes We estimate hospital-specific slopes (i.e. behavioral changes in response to CJR in period 2) and then examine how selection into bundled payment in period 3 varies with this measure. Letting $y_{h,2}$ be the average outcome for hospital h in period 2, we estimate hospital-specific slopes with a modified version of our baseline specification (Equation 1) that allows the treatment effect coefficient to vary by hospital:

$$y_{h,2} = \beta_0 + \sum_h \beta_{1,h} BP_h + \beta_2 y_{h,2014} + \beta_3 y_{h,2013} + \delta_{(h)} + \epsilon_h \tag{2}$$

where BP_h is an indicator for being randomly assigned to bundled payments and $\beta_{1,h}$ is the hospital-specific treatment effect. As in our baseline specification, we include lagged outcomes as covariates to improve statistical power, although in this specification the lags are defined at the hospital level. As before, we include strata fixed effects because randomization was conducted within strata. We estimate this specification on the set of voluntary treatment and control hospitals (see Figure 1). The (admittedly strong) identifying assumption is that, conditional on the covariates, there are no hospital-specific trends, and thus any heterogeneity in the change in outcomes across hospitals reflects heterogeneous treatment effects.

Panel B shows the results. Specifically, we show the average etsimated hospital-specific treamtent effects ($\beta_{1,h}$) and their standard deviation separately across hosptials that select into bundled payment (column 2) and those that select out (column 3). The results once again show selection in the expected direction: on average, total Medicare episode spending under period 2 bundled payment declines by $772 for hospitals that opt to remain in bundled payment in period 3, and increases by $424 for hospitals that revert to FFS in period 3; these average slope differences are statistically distinguishable at the 10% level (column 4). For the other two outcomes in Panel B - changes in institutional PAC spending and share discharged to institutional PAC - the differences in avearge behavioral responses in period 2 *is* statistically distinguishable between those who select in and those who select out of bundled payment in period 3.

Selection on Hospital Characteristics Panel C briefly examines other characteristics of hospitals that opt in and opt out. Hospitals that select into bundled payments in period 3 have a higher volume of CJR episodes in period 1. This suggests there may be fixed costs to remaining in the program, a point we return to with our model specification in Section 5. There are no meaningful or statistically significant differences in hospital size (as measured by beds) or share of teaching hospitals across the select in or select out groups. Hospitals that select in are less likely to be government owned than those the select out (8.2% versus 18.4%) but there is no meaningful or statistically significant difference in selection patterns between for-profit and non-profit hospitals. Hospitals who select in have higher period 2 quality, and are more likely to be from the northweat or midwest than hospitals who select out.

4 Model of Voluntary Selection

4.1 Setting

We consider a pool of CJR episodes, each indexed by i, which are admitted to hospital h. We assume throughout that this pool is taken as given, and is known to the hospital.

Under fee-for-service (FFS), providers are reimbursed based on claims. Let λ_i denote the FFS claims generated by a given episode. The preceding sections' description of the institutional environment and of the average impacts of bundled payment suggest that it is useful to decompose $\lambda_i = f_i^{HOSP} + f_i^{OTH}$, where f_i^{HOSP} are the fixed, DRG-based claims submitted for the index hospitalization and f_i^{OTH} are the claims submitted by post-acute care and other downstream providers. Let c_i^{HOSP} denote the costs incurred by the hospital and c_i^{OTH} the costs incurred by the other providers . For tractability, we assume that other providers are reimbursed at cost, so that $f_i^{OTH} = c_i^{OTH}$.[23] In what follows, for each variable x_i, we focus on hospital-level averages, defined as $x_h = \frac{1}{n_h}\sum_{n=1}^{n_h} x_i$, where n_h is the number of episodes at the hospital.

Hospital Profits and the Participation Incentive Under FFS, average Medicare reimbursement is $\lambda_h = f_h^{HOSP} + f_h^{OTH}$. Hospitals are only exposed to reimbursement and costs within the hospital and earn profits $\pi_h^{FFS} = f_h^{HOSP} - c_h^{HOSP}$. Under bundled payments (BP), Medicare reimburses the admitting hospital the fixed target price t_h for the entire episode, and hospitals are, effectively, required to pay for not only hospital costs c_h^{HOSP} but also Medicare's reimbursement to downstream providers f_h^{OTH}. We assume that hospital can reduce average Medicare reimbursement outside of the hospital by e through the exertion of "effort" $\phi_h(e)$, where $\phi_h(0) = 0$, $\phi_h' > 0$, and $\phi_h'' > 0$.[24] Hospitals, thus, choose effort to maximize

$$\pi_h^{BP} = \max_e \left(t_h - \left[(c_h^{HOSP} + f_h^{OTH}) - e \right] - \phi_h(e) \right),$$

and optimal effort is pinned down by $\phi_h'(e_h^*) = 1$. Since these are also the social marginal cost and benefit of effort, participation in BP results in the first-best level of effort.

For tractability, we assume that the cost of effort is quadratic of the form $\phi_h(e) = \frac{e^2}{2\omega_h}$ where $\omega_h > 0$ is a hospital specific parameter. With this assumption, the hospital's optimal choice of effort is $e_h^* = \omega_h$, average claims incurred under BP are $f_h^{HOSP} + f_h^{OTH} - \omega_h = \lambda_h - \omega_h$, and hospital profits are $\pi_h^{BP} = t_h - (c_h^{HOSP} + f_h^{OTH} - \frac{\omega_h}{2})$.

Hospitals select into a voluntary BP program, denoted by the indicator $BP_h = 1$, if and only if $\pi_h^{BP} > \pi_h^{FFS}$. Substituting in yields the criteria

$$BP_h = 1 \Leftrightarrow (t_h - \lambda_h) + \frac{\omega_h}{2} > 0, \tag{3}$$

[23] This assumption is primarily made to simplify notation. It is straightforward in the context of the model to allow other providers to obtain a fixed markup, but reaonsable levels of such markups would only affect slightly the quantitative results and would have no impact on the qualitative conclusions.

[24] For simplicity, we assume that c_h^{HOSP} remains the same under the BP program and is not affected by e. This is not essential, and can be viewed as a normalization, although it is a natural assumption. If the effort to reduce hospital cost and the effort to reduce post-acute care cost are separable, the hospital cost level was already optimized under FFS given that hospitals were already paid (under FFS) a fixed amount for the hospital portion of the episode.

where the right hand side is the sum of a "level" effect ($t_h - \lambda_h$) and a "slope" effect ($\frac{\omega_h}{2}$). The level effect ($t_h - \lambda_h$) represents the transfer hospitals would receive under BP relative to FFS if they did not change their behavior from what it was under FFS. The slope effect ($\frac{\omega_h}{2}$) denotes the net savings that hospitals get from any change in behavior under BP, which is the reduced provider costs $e_h^* = \omega_h$ net of the effort cost that reduction entails ($\frac{\omega_h}{2}$).[25]

Social welfare. The distinction between selection on levels and slopes has important implications for the social welfare consequences of voluntary programs. To see this, define social welfare W as the sum of consumer surplus (S) and producer profits (π), minus Medicare spending (CMS) weighted by the marginal cost of public funds $\Lambda > 0$:

$$W = S + \pi - (1 + \Lambda)CMS.$$

The multiplier $\Lambda > 0$ captures the deadweight loss associated with raising government revenue through distortionary taxation; alternatively, it can be thought of as capturing a societal preference for money in the hands of Medicare rather than hospitals.[26] Consistent with the descriptive results in Section 3, we assume that the hospital's effort does not affect patient welfare (S).

Medicare expenditure is t_h under BP and λ_h under FFS. Plugging in these and the hospital profits implies that hospital participation in the bundled payments improves social welfare if and only if

$$W_{BP} > W_{FFS} = 1 \Leftrightarrow \frac{\omega_h}{2} - \Lambda(t_h - \lambda_h) > 0. \tag{4}$$

The key social welfare tradeoff is as follows: On one hand, BP incentivizes hospitals to exert the first best level of effort $e_h^* = \omega_h$, which increases social welfare by $\frac{\omega_h}{2}$. On the other hand, enticing hospitals to participate in bundled payments increases public spending by $t_h - \lambda_h$. Selection on levels $(t_h - \lambda_h)$ is thus socially costly: it creates a transfer from Medicare to hospitals that is

[25]These incentives are well understood by the hospital industry. For example, Arbormetrix, a healthcare consulting firm, advises their client hospitals to consider the following questions when deciding whether to participate in a bundled payment program: "One: 'How good is my target price?' Two 'What has changed [since the target prices were set]?' And three: What is my opportunity to improve?". See https://www.arbormetrix.com/press-releases/new-report-significant-variation-in-expected-savings-per-hospital-in-bpci-advanced

[26]Under this interpretation, it would be natural to multiply S by $(1 + \Lambda)$, so that Λ represents the wedge between hospitals and consumers and government. Since we net S out of the calculations below, this can be done without loss of generality.

associated with a social cost of Λ. By contrast, selection on slopes increases social welfare due to the reduction in real resource utilization net of effort costs ($\frac{\omega_h}{2}$). In other words, because of the cost of public funds Λ, and the need to ensure that hospitals are willing to participate in bundled payment, hospital participation is not always social-welfare enhancing.

4.2 Graphical intuition

We illustrate the setting graphically in Figure 2, which depicts the participation incentives for hospitals and the corresponding social welfare implications. Hospitals are represented by a $\{\lambda_h, \omega_h\}$ pair. If one could mandate participation without any additional Medicare costs, the social welfare maximizing outcome would be to mandate that all hospitals join the BP program (given that ω_h is positive, by design, for all hospitals). However, if participation is voluntary and Medicare's ability to encourage participation rests on the financial incentive, t_h, the tradeoff is represented in Figure 2.

To draw the figure, we hold the payment t fixed across hospitals. At this payment, the solid line represents the set of hospitals that are indifferent between participation in bundled payment and FFS. Hospitals to the left prefer bundled payments, because the sum of the transfer holding their behavior constant ($t - \lambda$) and the savings they get under bundled payment ($\frac{\omega}{2}$) is positive. Hospitals to the right of the solid line prefer to remain under FFS. Thus, hospitals have both a simple "level" incentive to participate (if $t - \lambda > 0$), as well as an additional "slope" incentive, which explains why the solid line slopes up: All else equal, a higher ω_h provides additional incentive for the hospital to join the BP program as it captures some of the savings it can generate.

The dashed line in Figure 2 represents the set of hospitals for which social welfare is the same whether they participate in bundled payment or FFS. While the "slope" effect enters similarly to the voluntary participation condition and social welfare condition (see Equations 3 and 4), the "level" effect $t - \lambda$ enters positively into the hospital's participation decision (Equation 3) but negatively, and multiplied by the cost of public funds $(1 + \Lambda)$, in the the social welfare calculus (Equation 4). This explains why the dashed line is downward sloping, and illustrates the central social welfare tension in designing a voluntary regime: Enticing providers to participate can be socially costly.

Taken together, Figure 2 partitions hospitals to three groups: those which would remain in the FFS regime, those that efficiently select into participating in the BP program, and those which select to participate in the BP program inefficiently because they get paid much more than they "should" but do not generate significant efficiency gains (due to low ω_h).

4.3 Targeting

The BP program aligns effort incentives, so if Medicare could generate BP participation without any additional public expenditure, it would be social welfare improving to do so (since we assume $\omega_h > 0$). However, in a voluntary regime Medicare must respect the hospitals' participation constraint. If Medicare has perfect information about $\{\lambda_h, \omega_h\}$, it could maximize social welfare by setting $t_h = \lambda_h - \frac{\omega_h}{2}$ for each hospital; under such payment, all hospitals would voluntarily participate in the BP program and Medicare spending would be lower.

Once information about the joint distribution of $\{\lambda_h, \omega_h\}$ is incomplete, setting the payment amount involves a tradeoff, similar to the one in the classic optimal regulation design problem of Laffont and Tirole (1993). Figure 3 illustrates this tradeoff in our setting. In Panel A we start with Figure 2 and super-impose on it the participation and social welfare indifference sets that are associated with a higher level of payment $t' > t$. The gray (solid and dashed) indifference lines that correspond to t' are analogous to the black lines, which correspond to t. Naturally, the higher payment amount increases the share of hospitals that select to participate in the BP program. For many of the marginal hospitals that opt into BP, participation is social welfare-increasing. At the same time, however, the greater payment increases the social welfare cost associated with infra-marginal participants, by a fixed amount of $\Lambda(t' - t)$, and in doing so makes participation social welfare reducing for some of these hospitals (those that lie in between the two dashed lines).

The ability to effectively target depends on the nature of the underlying joint distribution of $\{\lambda_h, \omega_h\}$ as well as the precision of the social planner's information is about this joint distribution. This is illustrated in the remaining panels of Figure 3, which shows three ovals illustrating three examples of possible underlying joint distributions.

Comparing Panels B and C shows the importance of the overall level of ω_h. When ω_h is high (i.e., the cost of effort associated with reducing claims under BP is lower and therefore optimal effort under BP is higher), as in Panel B, the participation incentives of the hospital and the social planner are more closely aligned. In this case, even if information about λ_h is imperfect, it is easier to generate social welfare enhancing participation in bundled payment. However, when ω_h is low (i.e., the cost of effort is higher and optimal effort is lower), as in Panel C, selection is primarily driven by levels, and it is difficult to generate social welfare enhancing participation by hospitals. In this case, it requires much more precise information on λ_h to be able to generate social welfare gains through the voluntary bundled payment program.

Comparing Panels C and D shows the importance of the relative uncertainty regarding λ_h and ω_h. Because the primary policy instrument is a fixed payment, high uncertainty about the level λ_h, as in Panel C, would lead to more inefficiency due to voluntary participation. In contrast, if the primary uncertainty is about the "slope" ω_h, as in Panel D, voluntary participation is more likely to lead to social welfare gains.

The joint distribution of λ_h and ω_h, perhaps conditional on (priced) observables, is therefore key in assessing the potential efficiency gains from the BP program. It is therefore the key object of interest in our econometric exercise in next section.

5 Specification and Results

5.1 Econometric specification

We now turn to econometrically estimating the economic model presented in the last section. Recall the way we defined the "periods" associated with the experimental setting: In period 1, prior to the introduction of the BP program, all hospitals in the sample are still under the FFS program. In period 2 a subset of the hospitals are randomly assigned to the BP program, and in period 3 a subset of those latter hospitals endogenously select to remain in the BP program, while the rest switch "back" to the FFS setting.

For each hospital in the sample, we observe two periods of spending data (y_{h1}, y_{h2}) and three periods of BP program participation indicators $(BP_{h1}, BP_{h2}, BP_{h3})$. For hospitals randomly assigned to BP in period 2, we also observe the hospital-specific BP payment rate, t_{ht}, in periods 2 and 3 (even if they select out of the program).

As discussed in the end of Section 4, our key object of interest is the joint distribution of $\{\lambda_h, \omega_h\}$, where λ_h is the average total expenditure per-episode under FFS and ω_h is the reduction in per-episode spending that is caused by the BP program (recall that $e_h^* = \omega_h$). Taken at face value, the model implies the following two relationships between the latent objects we are interested in and the observed variables: First, the spending equation (for $t = 1, 2$) is given by

$$y_{ht} = \lambda_h - BP_{ht}\,\omega_h \quad \text{for} \quad t = 1, 2,$$

and second, the BP participation equation (which is relevant only for period 3, when participation is voluntary) is given by

$$BP_{h3} = 1 \iff (t_h - \lambda_{h3}) + \frac{\omega_h}{2} > 0.$$

To help fit the model to the data, we introduce two additional econometric components. First, we allow λ_h to vary over time according to

$$\ln \lambda_{ht} = \ln \lambda_h + g(t) + \epsilon_{ht} \quad \text{for} \quad t = 1, 2, 3.$$

where $g(t) = \gamma\,(t-2)$ is a linear time trend, normalized to be zero in period 2, and ϵ_{ht} is drawn (iid) from $N(0, \sigma_\epsilon^2)$.

Second, we introduce a hospital-level choice shifter into the BP selection equation. We saw that hospitals with a higher volume of CJR episodes are more likely to select into bundled payments (see Table 2); survey evidence on how hospitals respond to bundled payments suggest both hospital and patient-level costs (Zhu et al., 2018). To allow for both of these forces, we add the term $\nu_h = \frac{\nu}{n_h^\kappa}$ to the hospital-level selection equation, where ν and $\kappa \in [0, 1]$ are parameters to be estimated. For $\kappa = 1$, the hospital-level choice shifter becomes $\frac{\nu}{n_h}$, which is equivalent to a choice shifter of ν in a episode-level specification. For $\kappa = 0$, the hospital-level choice shifter is ν and therefore does not include a episode-level component. Intermediate values of κ capture intermediate cases.[27]

[27] For now, we remain agnostic as to what these additional forces represent. A key assumption we need to make is whether this choice shifter only affect hospital choice, or also enter the social welfare calculus. We thus defer the discussion of what this may be to Section 6, where we entertain the two possibilities.

Putting everything together, the econometric model is given by the following relationships:

$$
\begin{aligned}
y_{h1} &= \lambda_{h1}, \\
y_{h2} &= \lambda_{h2} - BP_{h2}\,\omega_h, \\
BP_{h3} = 1 &\iff (t_{h3} - \lambda_{h3}) + \frac{\omega_h}{2} + \frac{\nu}{n_h^{\kappa}} > 0, \\
\ln \lambda_{ht} &= \ln \lambda_h + \gamma\,(t-2) + \epsilon_{ht} \quad \text{for} \quad t = 1, 2, 3.
\end{aligned}
$$

Finally, we assume that $\ln \omega_h$ and $\ln \lambda_h$ are joint normally distributed, so that

$$
\begin{pmatrix} \ln \lambda_h \\ \ln \omega_h \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{\lambda,s} \\ \mu_\omega \end{pmatrix}, \begin{pmatrix} \sigma_\lambda^2 & \rho_{\lambda\omega}\sigma_\lambda\sigma_\omega \\ \rho_{\lambda\omega}\sigma_\lambda\sigma_\omega & \sigma_\omega^2 \end{pmatrix} \right).
$$

where the s subscript on $\mu_{\lambda,s}$ indicates that we allow the mean of $\ln \lambda_h$ to vary across strata. [28]

5.2 Identification

The conceptual identification of the model is fairly straightforward given the random assignment associated with the BP program. The model has a set of parameters that correspond to the "level" of spending and its evolution over time $\mu_{\lambda,s}, \sigma_\lambda, \gamma, \sigma_\epsilon$; a set of parameters that correspond to the reduction in spending under the BP program $\mu_\omega, \sigma_\omega$; a correlation parameter that relates the levels and slopes $\rho_{\lambda\omega}$; and choice shifter parameters ν and κ. The intuition for the identification argument follows in three steps.

First, using data from the control group alone, which allows us to observe λ_{h1} and λ_{h2} for the same set of hospitals, we can identify $\mu_{\lambda,s}$, γ, σ_λ, and σ_ϵ. We can use the control group alone because random assignment guarantees that parameters estimated from the control group are valid for the entire sample.

Second, using data from the treatment group as well as the control group, we can identify μ_ω, σ_ω, and $\rho_{\lambda\omega}$. We observe λ_{h1} and $\lambda_{h2} - \omega_h$ for all hospitals in the treatment group. Since the average change between λ_{h1} and λ_{h2} is γ, the remaining difference identifies μ_ω. The dispersion in the change in spending between period 1 and period 2 is driven by a combination of the stochastic

[28] Since assignment to BP in period 2 was random conditional on strata, allowing the mean to vary by strata isolates the experimental variation and is the analogue to controlling for strata fixed effects in Equation 1.

evolution of λ_{ht} and the dispersion in ω_h. Since the stochastic evolution of λ_{ht} is already identified from the control group, we can (loosely) net it out, and the residual dispersion identifies σ_ω. The intuition for identifying $\rho_{\lambda\omega}$ is similar: We observe the reduction in spending for each hospital in the treatment group, and can correlate it with the hospital's period-1 spending, and adjust it appropriately for the additional independent noise that is driven by the stochastic evolution of λ_{ht}, which is already identified by the control group.

Finally, the BP selection equation identifies the distribution of the remaining choice shifter parameters ν and κ. This equation resembles a probit equation, but the error term has an economic interpretation as reflecting hospitals' profit maximizing choices. The joint distribution of λ_{h3} and ω_h, which is identified from the previous two steps, together with our model, generates predictions for the overall take-up rate of the BP program (among the hospitals participating in the treatment group). Any deviation from this "predicted" take-up rate identifies ν, with κ identified by the extent to which hospitals with greater numbers of episodes are, all else equal, more likely to select into the BP program.

5.3 Estimation

We estimate the model in two steps using maximum likelihood, following the identification argument quite closely. We summarize the estimation procedure here, and provide more details in Appendix A.

In the first step we follow the first part of the identification argument, by estimating $\mu_{\lambda,s}$, γ, σ_λ, and σ_ϵ using data from the control group alone. An observation is a pair of spending for a control group hospital in period 1 and period 2 $\{y_{h1}, y_{h2}\}$.

We could continue along the identification argument, and estimate the remaining parameters in two additional steps, but in order gain efficiency we combine them into one. Specifically, we now use observations on the treatment group hospitals, where each independent observation is given by $\{y_{h1}, y_{h2}, BP_{h3}\}$. The likelihood is given by

$$L_h = Pr(BP_{h3} = 1) \cdot f(y_{h1}, y_{h2}|BP_{h3} = 1) + Pr(BP_{h3} = 0) \cdot f(y_{h1}, y_{h2}|BP_{h3} = 0),$$

and can be evaluated numerically. As a way to speed up computation (with some small loss in efficiency), we can use the parameters estimated in the first step and the observed value of y_{h1} to generate a "posterior distribution" for $\ln \lambda_{h2}$ and (identically) $\ln \lambda_{h3}$ for each hospital. This simplifies the second step of the estimation. When we estimate the model, we weight each hospital by the number of episodes in period 2, so that the resulting parameters are representative at the episode level.

5.4 Results

Table 3 presents the estimation results. Panel A reports the parameter estimates and Panel B presents some of the key summary statistics that are implied by these estimates, first for all hospitals (Panel B.1) and then limited to the sample of voluntary treatment hospitals (Panel B.2) – i.e. those subject to the voluntary bundled payment regime (see Figure 1).

Panel A indicates a slightly negative trend in episode spending ($\gamma = -0.07$) – which is consistent with the time series pattern in the control group – and a relatively small standard deviation for the idiosyncratic disturbances in λ_{ht} ($\sigma_\epsilon = 0.07$ versus $\sigma_\lambda = 0.17$), which yields a λ_{h3} with an expected value of \$24,200. We estimate that ω_h has an expected value of \$450, which is slightly smaller than the average effect estimated in Table 1. The distributions of λ_h and ω_h have a modest positive correlation ($\rho_{\lambda,\omega} = 0.27$). The estimate of κ is 0.58, implying that the choice shifter scales up with the number of episodes but less than proportionally, suggesting that it can be thought of as some combination of hospital-level and episode-level costs.

The model emphasizes the importance of heterogeneity in levels and heterogeneity in slopes in determining the nature of selection, Medicare spending and social welfare under a voluntary bundled payment regime. The results in Panel B.1 indicate substantially more heterogeneity in levels than in slopes – the standard deviation of $\lambda_h 3$ is \$4,400 compared to a standard deviation of ω_h of \$1,300. This raises concerns that the voluntary system may primarily produce inefficient transfers to hospitals through substantial selection on levels. However, the potential for selection on levels may be limited by the design of target prices. We explore this in Panel B.2 which is restricted to the subsample of voluntary treated hospitals, where we can also observe target prices and which will be the focus of our counterfactuals in the next section. Panel B.2 indicates that netting out target prices does not noticeably reduce heterogeneity in levels: the \$4,900 standard deviation of $\lambda_{h3} - t_h$ for voluntary treatment hospitals is only slightly less than the \$5,500 standard deviation of λ_{h3} for these hospitals.[29]

To further examine the role of target prices, Figure 4 produces empirical analogues of the selection figures we used to illustrate the model in Section 4, incrementally accounting for target prices and the choice-shifter ν_h. Once again we restrict analysis to the subsample of voluntary treatment hospitals. To provide a baseline, Panel A of Figure 4 plots simulated hospitals from the joint distribution of ω_h (vertical axis) and λ_{h3} (horizontal axis), without netting out the target price. As would be expected from the results in Table 3 Panel B.2, the plot suggests that selection on levels is a primary concern, with a large mass of hospitals selecting BP inefficiently or selecting FFS. In Panel B of Figure 4 we examine the role of targeting by plotting $\lambda_{h3} - t_h$ on the horizontal axis, instead of λ_{h3}. Netting out target prices does not noticeably shrink the heterogeneity along the horizontal axis, with large masses of hospitals continuing to select BP inefficiently. In Panel C, we further add the choice-shifter by plotting $\lambda_{h3} - t_h + \nu_h$ on the horizontal axis, thus capturing all the components of the selection decision.

6 Counterfactuals

We use the estimated model to perform a set of counterfactual exercises. Throughout, we focus on the sample of voluntary treated hospitals, defined as treatment group hospitals that were given a choice in period 3 of whether to remain in BP or revert back to FFS (see Figure 1).

6.1 Voluntary vs. Mandatory

We first compare outcomes under the observed voluntary bundled payment program to two counterfactuals: all hospitals mandated to be under the status quo FFS regime, or all hospitals mandated to participate in the bundled payment program. These counterfactuals can be thought about as measuring the impact of the Trump Administration's decision to make the BP program voluntary for these hospitals, relative to cancelling the BP program entirely or keeping it mandatory. Once again, we focus our analysis on hospitals subject to the voluntary bundled payment program (see Figure 1).

To operationalize this counterfactual exercise, we use our model and parameter estimates

[29]From a mathematical perspective, this should not be surprising. Since $Var(\lambda_{h3} - t_h) = Var(\lambda_{h3}) + Var(t_h) - 2Cov(\lambda_{h3}, t_h)$, the distributions of λ_{h3} and t_h can have a modest positive covariance and still yield a case where $Var(\lambda_{h3} - t_h) < Var(\lambda_{h3})$.

reported in Table 3 to simulate hospital-specific values for $\{\lambda_{h3}, \omega_h\}$ in period 3 conditional on the hospital's period 3 selection decision BP_{h3}, target price t_{h2}, and number of CJR episodes n_{h2} from period 2.[30] We assume a social cost of funds of $\Lambda = 0.15$.

Panel A of Table 4 shows the results. The first row reports results if there were no bundled payment program and all hospitals are paid under FFS. Medicare spending (i.e. "CMS" in the defintion of social welfare from Equation 4.1) averages \$24,306 per episode in this counterfactual (which corresponds to the average λ_{h3} reported in Panel C of Table 3). The remainder of the entries are normalized to zero; the mandatory FFS counterfactual will serve as a benchmark from which to compare other regimes.

The second row of Panel A considers the counterfactual where hospitals are mandated to enroll in the BP program in period 3, as was intended under the *initial* design of the BP program. Under mandatory BP, hospitals receive a transfer from Medicare target prices $(t_h - \lambda_h)$ and are also residual claimants on the ω-related savings they generate (i.e. $\frac{\omega_h}{2}$.) If target prices had been calibrated to equal counterfactual FFS costs on average ($\mathbb{E}_h t_h = \mathbb{E}_h \lambda_{h3}$), Medicare spending would have been unaffected relative to the baseline. However, as seen in Panel B.2 of Table 3, target prices t_h ended up being on average slightly higher (\$52) than counterfactual FFS spending, so Medicare spending is \$52 higher and (multiplied by $1 + \Lambda = 1.15$) social costs thus increase by \$60 (column 3).

In columns (4) through (7) we consider two different versions of the welfare analysis, depending on whether we treat the choice shifter ν_h term as non-welfare relevant or welfare-relevant. In columns (4) and (5) we assume ν_h is not welfare relevant (e.g., because it represents a choice friction such as status quo bias rather than a real fixed costs). As a result, hospital profits relative to FFS (i.e. $(t_h - \lambda_h) + \frac{\omega_h}{2}$) rise by \$302 (column 4), [31] and social surplus rises by \$242 (column 5). In other words, the incentive effects of BP (which generate $\frac{\omega_h}{2}$ in social savings) are larger than the social cost of the transfer from Medicare to hospitals $((\Lambda)((t_h - \lambda_h))$, leading to a modest increase in social surplus per episode. Naturally, if ν_h is taken into account (as in columns 6 and 7), both hospital profits and relative social surplus are lowered by its average of \$1,668 (see Table 3 panel

[30] CMS has not yet made target prices for period 3 available, and thus we rely on period 2 target prices in this draft of the paper. We expect CMS to make these data available soon and will update our analysis when they are available.

[31] Panel B.2 of Table 3 indicates that ω_h has an expected value of of \$500, so that hospital profits from changes in behavior (i.e. $\frac{\omega_h}{2}$) are \$250. The additional \$52 in hospital profits represent a transfer from CMS from target prices being on average \$52 higher than counterfactual FFS spending.

B.2).

The third row of Table 4 Panel A considers the voluntary selection scenario that actually took place. Outcomes for voluntary selection are simulated based on the model parameters. In particular, for each hospital we simulate a binary participation decision and the resulting Medicare costs, hospital profits, and social surplus values. We find that 39% of episode-weighted hospitals select into BP, which is almost identical to the actual selection percentage (37% in Table 2), providing assurances about the in-sample fit of our model. As we discussed in Section 5, the much greater heterogeneity in $t_h - \lambda_h$, relative to $\omega_h/2$, suggests that selection into BP is primarily on "levels." Consequently, voluntary participation raises Medicare costs by $1,701 relative to the baseline and by $1,649 relative to the mandatory scenario.

Since we are giving hospitals a choice, hospital profits must be weakly higher under voluntary relative to the mandatory regime. When we treat the ν_h term as non-welfare relevant, hospital profits rise to $1,807 above the mandatory FFS benchmark, which is $1,505 higher than under the mandatory bundled payment program. Ignoring ν_h, social surplus under voluntary is less than either the mandatory or baseline scenarios (-$149, column 5), due to both the larger transfer and smaller share of hospitals generating efficiency gains of $\omega_h/2$. When we treat ν_h as welfare relevant, hospital profits rise by a smaller amount, of $1,241 above the mandatory FFS benchmark, and social surplus is correspondingly lower. It is worth noting however that, while negative, social surplus is less negative than under the mandatory counterfactual (see column 7). Intuitively, if we think that these costs are real, it is important to let hospitals avoid them if they do not expect the benefits from reducing spending to offset the fixed costs of BP.

While these estimates indicate that a voluntary BP program reduces welfare, they do not imply that such a program is necessarily ineffective. The BP program generates positive ω-related efficiency gains. The reason the overall effect is negative is because these ω-related gains are small relative to the social welfare losses associated with "over paying" participating hospitals relative to their counterfactual FFS cost. It is therefore useful to explore how the voluntary program performance could improve if Medicare were able to set target prices to better reflect underlying hospital-specific costs, which is what we turn to next.

6.2 Targeting

In order to explore price targeting under voluntary BP in a systematic fashion, we approximate the observed target prices using a parametric distribution, and then we examine the impacts of shifting its parameters. Specifically, we assume that hospital-level target prices t_h are log-normally distributed, and are correlated with hospital costs. We then explore voluntary participation under different values for the level of target prices μ_t, their variance σ_t^2, and their correlation with hospitals costs $\rho_{t,\lambda}$. Appendix B provides more details.

Figure 5 summarizes the outcomes from this exercise, plotting social surplus relative to the mandatory FFS benchmark (y-axis) against Medicare costs (x-axis) for different values of $\{\mu_t, \sigma_t^2, \rho_{t,\lambda}\}$. In the plot, we focus on the social surplus values that do not consider ν_h (column 5 of Table 4). Panel B of Table 4 reports additional outcomes associated with each exercise. The black dot in Figure 5 corresponds to the the observed distribution of t_h, which is similar to the third row of Panel A of Table 4 (it is not exactly the same because of the log-normal parameterization of the target price distribution), and serves as a benchmark.

We consider three other possible targeting policies. The first, indicated by the point labeled as "perfect targeting", sets target prices exactly equal to counterfactual fee-for-service costs ($t_h = \lambda_{h3}$). Under this contract, there is no transfer to hospitals (no selection on levels) and hospitals are the full residual claimants on ω-related savings they generate. Because there are no transfers to offset the ν_h, only 19.5% of hospitals select into BP under this contract. These 19.5% generate ω-related surplus gains of \$XX, but because there are few of them, average total surplus only increases by \$73 (ignoring the choice shifter).

"Perfect targeting" is a useful benchmark, but not feasible in our model: Medicare setting target prices of $\lambda_{h3} = \lambda_h + g(t) + \epsilon_{h3}$ is infeasible, since ϵ_{h3} is only known to the hospital. Therefore, to gauge the benefits of a more feasible contract, we consider a second scenario where Medicare sets a target price $t_h = \lambda_h + g(t)$, which is based on the hospital's underlying type and the time trend. Given our stochastic assumptions, the mean of this contract is unchanged, the standard deviation is reduced to $\sigma_\lambda = \sqrt{\sigma_{\lambda,3}^2 - \sigma_\epsilon^2}$, and the correlation with λ_{h3} reduced to $\frac{\sigma_\lambda}{\sigma_\lambda + \sigma_\epsilon} = 0.72$. This contract, which is labeled as "feasible targeting" generates slightly lower social surplus (of \$31) and higher Medicare spending.

A third exercise (labeled as "no targeting") considers a case where target prices are uniform across hospitals at a value equal to the average of λ_{h3}. Relative to the observed targeting, the no targeting case leads to even greater (inefficient) selection on levels, greater spending and social cost, and greater participation. Relative to the potential benefits from feasible targeting (of $192 = 31 - (-161) per episode), the observed targeting seems to generate approximately two thirds of the feasible gains ($133 = -28 - (-161) out of $192).

While we mostly view this exercise as illustrative, we should point out that improved targeting – that is, higher values of $\rho_{t,\lambda}$ in the context of this exercise – does not have to rely on better information. It could also rely on a narrower definition of the bundle. To illustrate, consider the case, which is broadly consistent with the evidence shown in Section 3, where hospital episode spending consists of two additive separable components: hospital spending and other spending, which includes post-acute care. As discussed, given that we expect hospitals to have had sufficient incentive to make efficient choices regarding their own spending (see even under "FFS" hospitals are paid a capitated amount per admission), it is natural to attribute all the potential ω-related saving to all other spending, and post-acute spending in particular. However, if much of the heterogeneity across hospitals in their episode-level spending – which underlies the difficulty in targeting – is associated with hospital spending, it might be better to only target the post-acute spending; this can be thought of as approximately increasing the correlation coefficient. In particular, we consider two such cases. In the first ("narrow bundling, no targeting") we assume that we only apply such narrow bundling, but cannot target price further. Within our framework, this is equivalent to setting a target price of the sum of realized in-hospital spending and mean other spending, which results in correlation of 0.76.[32] In the second ("narrow bundling, observed targeting"), we assume that we can target prices even more efficiently, by narrowing the bundle and in addition targeting prices such that the we can obtain the same observed correlation between hospital spending and target prices also when we consider non-hospital spending only. This exercise is equivalent to raiding the correlation coefficient further, to 0.87.[33].This last exercise

[32]This target price is set to be $f_h^{HOSP} + \mathbb{E}_h[f_h^{OTH}]$, and we calculate its mean, standard deviation, and correlation using the empirical distributions of λ and this target price.

[33]This target price is set to $f_h^{HOSP} + t_h^{OTH}$, where t_h^{OTH} is the other component of the target prices. We set the mean of t_h^{OTH} equal to the mean of f_h^{OTH}, the standard deviation to have the same ratio with f_h^{OTH} as the observed target price does with λ_h, and correlation of t_h^{OTH} and f_h^{OTH} equal the correlation between the observed target price and λ_h

generates considerable social gains relative to the observed coluntary bundled payment program and brings overall social welfare to the levels under the status quo mandatory FFS regime.

Overall, our finding suggest that while the observed voluntary program is socially costly, there are feasible improvements in targeting – which could arise by a cobination of tailoring target prices better to reflect the cost structure of each hospital, and by more narrowly focusing the budle on the subset of services in which cost-saving are more likely to occur – that could make a voluntary program that would eliminate social losses or generate small social gains. A feature of all the exercises considered so far is that they attempt to stay within the nature of the observed program, where there is only a single target price that is used to incentivize each hospital. In the next set of exercises we go further out of sample, and explore a richer set of contract designs.

6.3 Screening Contracts

TBA (To be added in future version).

7 Conclusion

TBD

References

Alexander, Diane. 2017. "How do doctors respond to incentives? unintended consequences of paying doctors to reduce costs."

Barnett, Michael L, Andrew Wilcock, J Michael McWilliams, Arnold M Epstein, Karen E Joynt Maddox, E John Orav, David C Grabowski, and Ateev Mehrotra. 2019. "Two-year evaluation of mandatory bundled payments for joint replacement." *New England Journal of Medicine*, 380(3): 252–262.

Carroll, Caitlin, Michael Chernew, A Mark Fendrick, Joe Thompson, and Sherri Rose. 2018. "Effects of episode-based payment on health care spending and utilization: Evidence from perinatal care in Arkansas." *Journal of health economics*, 61: 47–62.

Centers for Medicare and Medicaid Services. 2016. "Medicare program; hospital inpatient prospective payment systems for acute care hospitals and the long-term care hospital prospective payment system and policy changes and fiscal year 2017 rates; quality reporting requirements for specific providers; graduate medical education; hospital notification procedures applicable to beneficiaries receiving observation services; technical changes relating to costs to organizations and Medicare cost reports; finalization of interim final rules with comment period on LTCH PPS payments for severe wounds, modifications of limitations on redesignation by the Medicare Geographic Classification Review Board, and extensions of payments to MDHs and low-volume hospitals. Final rule."

Centers for Medicare & Medicaid Services. 2015*a*. "Medicare Program; Comprehensive Care for Joint Replacement Payment Model for Acute Care Hospitals Furnishing Lower Extremity Joint Replacement Services. Final rule." *Federal register*, 80(226): 73273.

Centers for Medicare & Medicaid Services. 2015*b*. "Medicare program; hospital inpatient prospective payment systems for acute care hospitals and the long-term care hospital prospective payment system policy changes and fiscal year 2016 rates; revisions of quality reporting requirements for specific providers, including changes related to the Electronic Health Record Incentive Program; extensions of the Medicare-dependent, Small Rural Hospital Program and the low-volume payment adjustment for hospitals. Final rule; interim final rule with comment period." *Federal Register*, 80(158): 49325.

Centers for Medicare & Medicaid Services. 2017. "Medicare Program; Cancellation of Advancing Care Coordination Through Episode Payment and Cardiac Rehabilitation Incentive Payment Models; Changes to Comprehensive Care for Joint Replacement Payment Model: Extreme and Uncontrollable Circumstances Policy for the Comprehensive Care for Joint Replacement Payment Model." *Federal Register*, 82: 57066–57104.

Clemens, Jeffrey, and Joshua D Gottlieb. 2014. "Do physicians' financial incentives affect medical treatment and patient health?" *American Economic Review*, 104(4): 1320–49.

CMS. 2019. "CMS Fast Facts." Centers for Medicare & Medicaid Services.

Cromwell, Jerry, Debra A Dayhoff, and Armen H Thoumaian. 1997. "Cost savings and physician responses to global bundled payments for Medicare heart bypass surgery." *Health Care Financing Review*, 19(1): 41.

Cutler, David M. 1995. "The Incidence of Adverse Medical Outcomes Under Prospective Payment." *Econometrica: Journal of the Econometric Society*, 29–50.

Cutler, David M, and Kaushik Ghosh. 2012. "The potential for cost savings through bundled episode payments." *New England Journal of Medicine*, 366(12): 1075–1077.

DeAngelis, Corey, Lindsey Burke, and Patrick Wolf. 2018. "The Effects of Regulations on Private School Choice Program Participation: Experimental Evidence from Florida." Working Paper.

Doran, James P, and Stephen J Zabinski. 2015. "Bundled payment initiatives for Medicare and non-Medicare total joint arthroplasty patients at a community hospital: bundles in the real world." *The Journal of arthroplasty*, 30(3): 353–355.

Dummit, Laura A, Daver Kahvecioglu, Grecia Marrufo, Rahul Rajkumar, Jaclyn Marshall, Eleonora Tan, Matthew J Press, Shannon Flood, L Daniel Muldoon, Qian Gu, et al. 2016. "Association between hospital participation in a Medicare bundled payment initiative and payments and quality outcomes for lower extremity joint replacement episodes." *Jama*, 316(12): 1267–1278.

Einav, Liran, Amy Finkelstein, and Neale Mahoney. 2018. "Provider incentives and healthcare costs: Evidence from long-term care hospitals." *Econometrica*, 86(6): 2161–2219.

Einav, Liran, Amy Finkelstein, Raymond Kluender, and Paul Schrimpf. 2016. "Beyond statistics: the economic content of risk scores." *American Economic Journal: Applied Economics*, 8(2): 195–224.

Einav, Liran, Amy Finkelstein, Stephen P Ryan, Paul Schrimpf, and Mark R Cullen. 2013. "Selection on moral hazard in health insurance." *American Economic Review*, 103(1): 178–219.

Eliason, Paul J, Paul LE Grieco, Ryan C McDevitt, and James W Roberts. 2018. "Strategic Patient Discharge: The Case of Long-Term Care Hospitals." *American Economic Review*, 108(11): 3232–65.

Elixhauser, Anne, Claudia Steiner, D Robert Harris, and Rosanna M Coffey. 1998. "Comorbidity measures for use with administrative data." *Medical care*, 8–27.

Finkelstein, Amy, Yunan Ji, Neale Mahoney, and Jonathan Skinner. 2018. "Mandatory Medicare bundled payment program for lower extremity joint replacement and discharge to institutional postacute care: interim analysis of the first year of a 5-year randomized trial." *Jama*, 320(9): 892–900.

Fisher, Elliott S. 2016. "Medicare's bundled payment program for joint replacement: promise and peril?" *Jama*, 316(12): 1262–1264.

Frakt, Austin. 2019. "Which Health Policies Actually Work? We Rarely Find Out."

Froemke, Cecily C, Lian Wang, Matthew L DeHart, Ronda K Williamson, Laura Matsen Ko, and Paul J Duwelius. 2015. "Standardizing care and improving quality under a bundled payment initiative for total joint arthroplasty." *The Journal of arthroplasty*, 30(10): 1676–1682.

GAO. 2018. "Voluntary and Mandatory EpisodeBased Payment Models and Their Participants." United States Government Accountability Office GAO-19-156.

Gronniger, T, M Fiedler, K Patel, L Adler, and P Ginsberg. 2017. "How should the Trump Administration handle Medicare's new bundled payment programs." *Health Affairs blog. April*.

Haas, Derek A, Xiaoran Zhang, Robert S Kaplan, and Zirui Song. 2019. "Evaluation of Economic and Clinical Outcomes Under Centers for Medicare & Medicaid Services Mandatory Bundled Payments for Joint Replacements." *JAMA internal medicine*.

Heckman, James J, and Bo E Honore. 1990. "The empirical content of the Roy model." *Econometrica*, 58(5): 1121–1149.

Ho, Kate, and Ariel Pakes. 2014. "Hospital choices, hospital prices, and financial incentives to physicians." *American Economic Review*, 104(12): 3841–84.

Ida, Takanori, Koichiro Ito, and Makoto Tanaka. in preparation. "Policy Design with Advantageous Selection: Experimental Evidence from Electricity Plan Choice."

Jack, B Kelsey, and Seema Jayachandran. 2019. "Self-selection into payments for ecosystem services programs." *Proceedings of the National Academy of Sciences*, 116(12): 5326–5333.

King, Robert. 2019. "CMS to use mandatory models 'very judiciously,' official says."

Laffont, Jean-Jacques, and Jean Tirole. 1993. *A theory of incentives in procurement and regulation.* MIT press.

Levy, S, N Bagley, and R Rajkumar. 2018. "Reform at Risk-Mandating Participation in Alternative Payment Plans." *The New England journal of medicine*, 378(18): 1663–1665.

Navathe, Amol S, Andrea B Troxel, Joshua M Liao, Nan Nan, Jingsan Zhu, Wenjun Zhong, and Ezekiel J Emanuel. 2017. "Cost of joint replacement using bundled payment models." *JAMA internal medicine*, 177(2): 214–222.

Newcomer, Lee N, Bruce Gould, Ray D Page, Sheila A Donelan, and Monica Perkins. 2014. "Changing physician incentives for affordable, quality cancer care: results of an episode payment model." *Journal of oncology practice*, 10(5): 322–326.

Quan, Hude, Vijaya Sundararajan, Patricia Halfon, Andrew Fong, Bernard Burnand, Jean-Christophe Luthi, L Duncan Saunders, Cynthia A Beck, Thomas E Feasby, and William A Ghali. 2005. "Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data." *Medical care*, 1130–1139.

Shatto, John D. 2016. "Center for Medicare and Medicaid Innovation's Methodology and Calculations for the 2016 Estimate of Fee-for-Service Payments to Alternative Payment Models." Centers for Medicare and Medicaid Services.

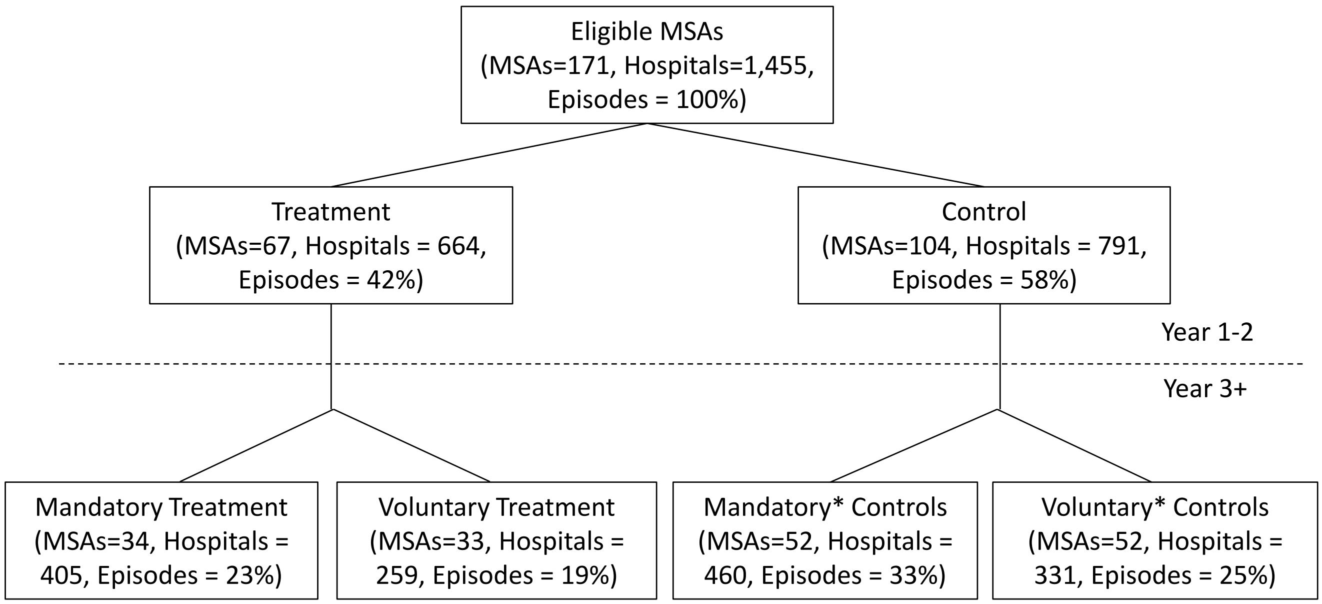
Shepard, Mark. 2016. "Hospital network competition and adverse selection: Evidence from the Massachusetts health insurance exchange."

Thaler, Richard H, and Cass R Sunstein. 2003. "Libertarian paternalism." *American economic review*, 93(2): 175–179.

The Lewin Group. 2018. "CMS Comprehensive Care for Joint Replacement Model: Performance Year 1 Evaluation Report."

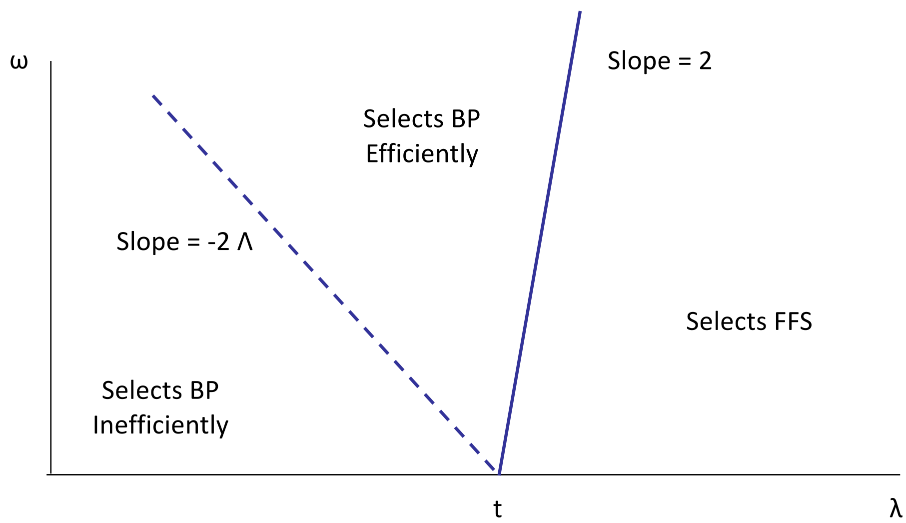
Zhu, Jane M, Viren Patel, Judy A Shea, Mark D Neuman, and Rachel M Werner. 2018. "Hospitals using bundled payment report reducing skilled nursing facility use and improving care integration." *Health Affairs*, 37(8): 1282–1289.

Figure 1: Experimental Design



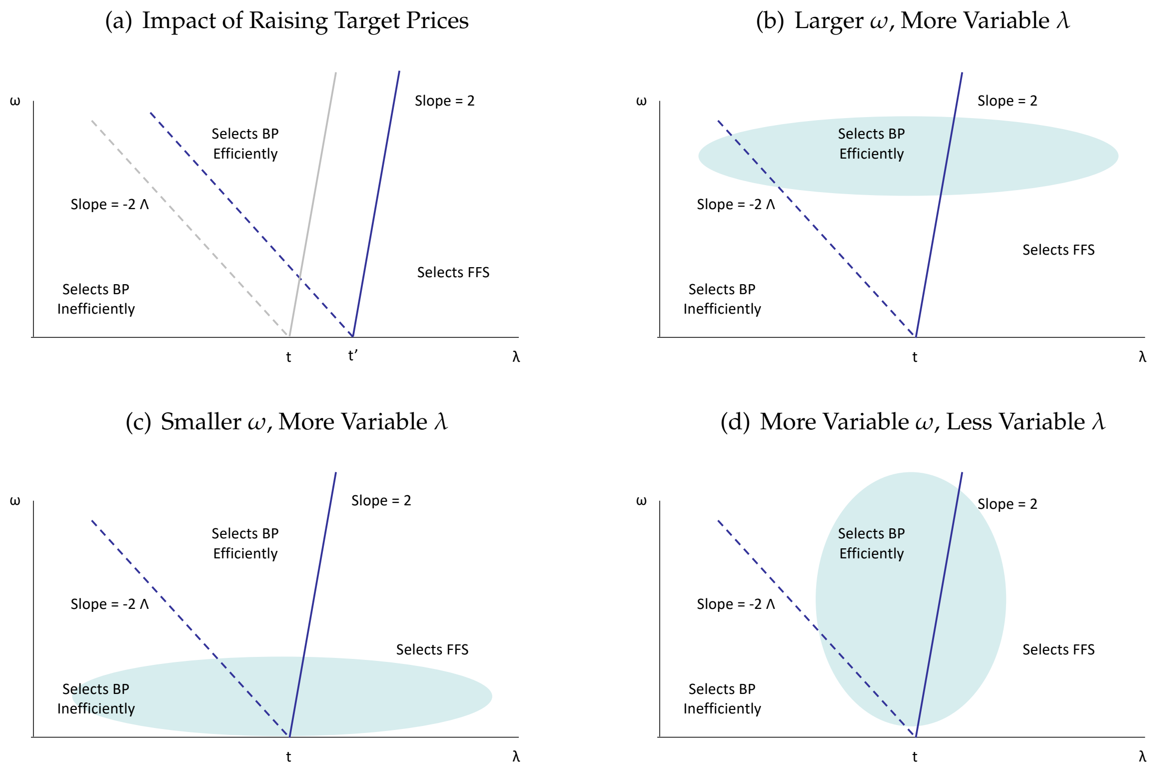
Note: Figure shows experimental design of bundled payments experiment. The top half shows the initial mandatory design in program years 1 and 2, the bottom half shows the partially voluntary design in program years 3-5. Episode shares are based on data from program years 1 and 2. *Control group MSAs are assigned to mandatory vs. voluntary by authors using historical spending

Figure 2: Hospital Selection Into Bundled Payment and Social Welfare Implications



Note: Figure shows - for a given target price t, the hospital participation decision and social welfare implications as a function of the hospital's λ and ω.

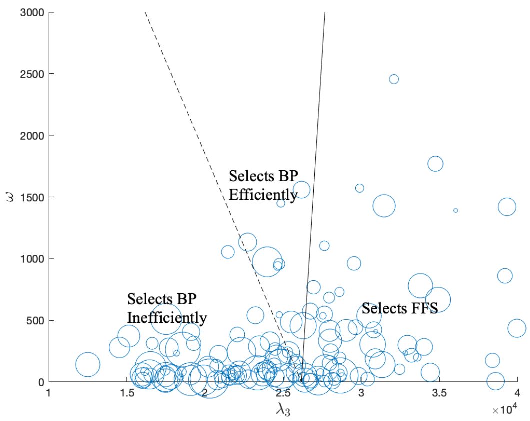
Figure 3: Model Illustration



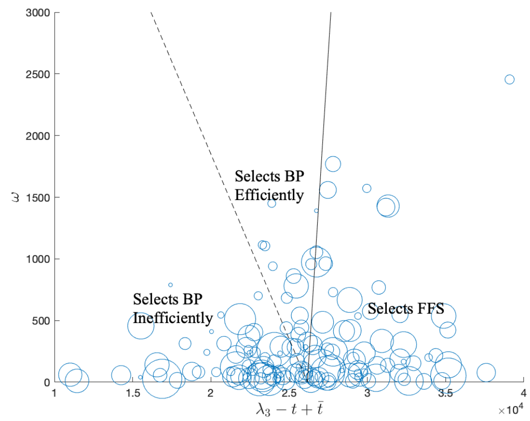
Note: Figure ilustrates some of the key analytics in voluntary bundled payment design. Panel A illustrates the tradeoffs involved in setting higher target payments $t' > t$; Panels B through D consider the impact of different primitives and targeting, with Panel B vs C comparing outcomes with higher vs. lower ω and Panel C vs D comparing outcomes with more vs less unobserved heterogeneity in λ.

Figure 4: Model Estimates

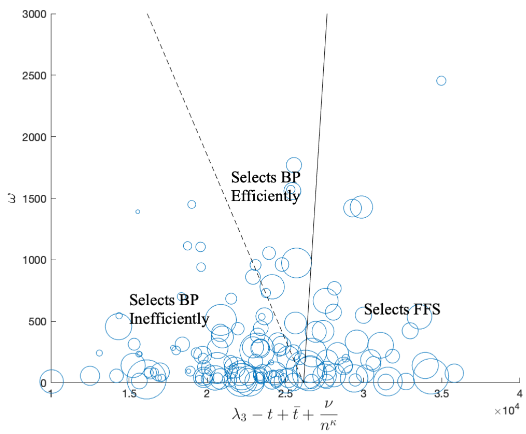
(a) No Targeting or Choice Shifter



(b) Target Price, Ignoring Choice Shifter

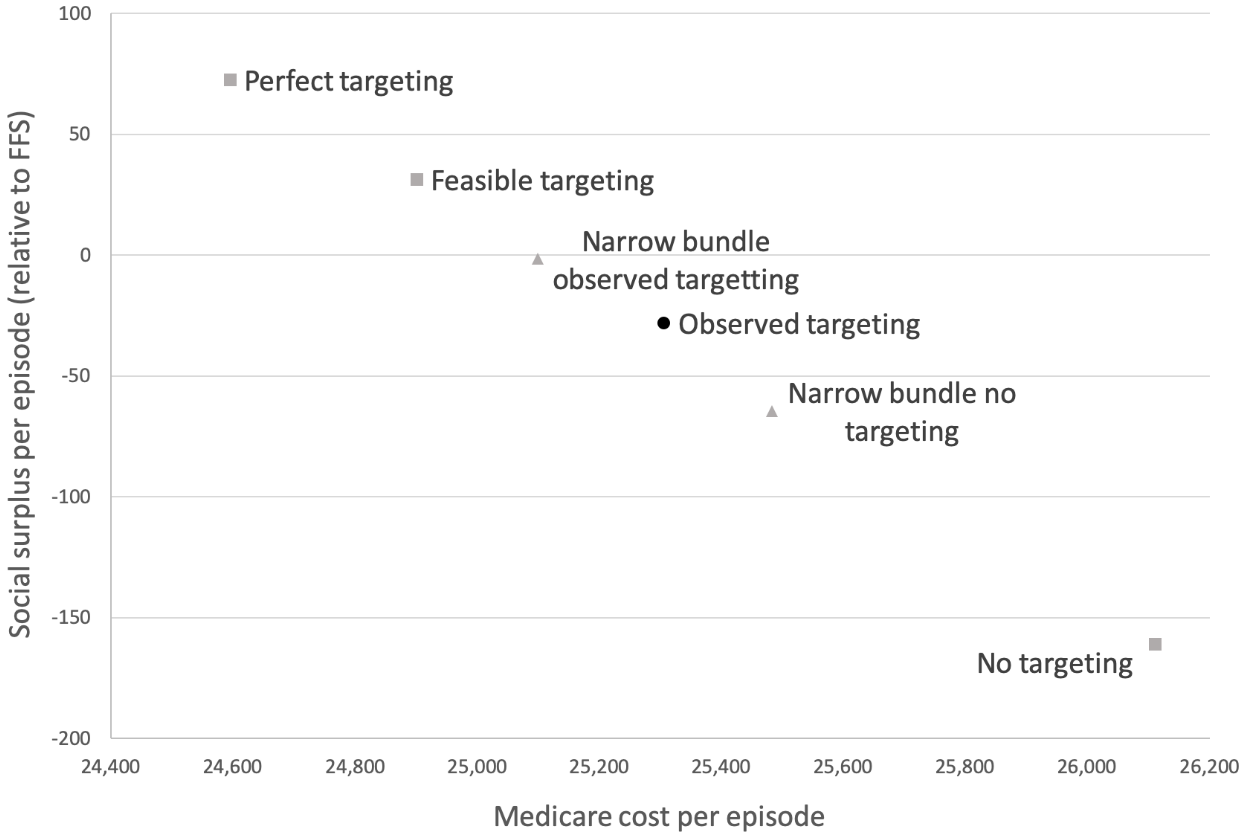


(c) Target Price and Choice Shifter



Note: Figure reports the empirical analog of the selection Figures 2 and 3. Specifically, it reports simulated hospitals (weighted by the number of episodes) based on our estimates for the 259 hospitals in the voluntary treatment sample (see Figure 1). Panel (a) reports results assuming there are no target prices and the choice shifter ν_h in the hospital selection equation is not welfare relevant. Panel (b) adds considers the role of target prices by plotting $\lambda_{h3} - t_h - \bar{t}$ on the horizontal axis, instead of λ_{h3} – we subtract $\bar{t}$ (the average target price) so that the axis remains on the same scale as in Panel (a). finally in Panel (c) we not only net out target prices but also allow the choice shifter ν_h to be welfare relevant (and add it onto the x axis again to keep the scale the same).

Figure 5: Medicare Costs and Social Surplus Under Alternative Target Prices



Note: TBD. mention weighting by episodes.

Table 1: Experimental Estimates

	Control Mean	Control SD	Average Treatment Effect	Standard Error	P-value
Panel A: Healthcare Use and Spending					
Total Episode Spending if paid under FFS	25,304	3,601	-797	204	0.001
Episode Spending in Index Admission	13,546	2,387	-175	89	0.06
Episode Spending in Institutional PAC Admission	4,122	1,380	-498	128	0.001
Episode Spending in Home Health	1,801	918	-89	59	0.14
Other Episode Spending	5,835	531	26	55	0.64
Total Episode Spending Net of Reconciliation	25,304	3,601	33	208	0.88
Number of Days in Index Admission	2.6	0.4	-0.1	0.0	0.21
Number of Days in Institutional PAC	7.7	2.3	-0.6	0.2	0.02
Discharge Destination					
Institutional Post Acute Care	0.313	0.104	-0.034	0.009	0.001
Home Health Agency	0.339	0.196	0.004	0.018	0.81
Home (w/o Home Health Agency)	0.329	0.232	0.042	0.018	0.02
Other	0.020	0.032	-0.004	0.002	0.05
Panel B: Quality Measures					
Complication Rate	0.011	0.005	0.001	0.001	0.26
ER Visit During Episode	0.199	0.027	0.003	0.003	0.41
90-day All Cause Readmission Rate	0.102	0.015	-0.001	0.002	0.69
Panel C: Admissions and Patient Composition					
LEJR Admissions (per 1,000 enrollees)	29.9	15.8	-0.8	0.5	0.10
CJR-eligible LEJR Admissions (per 1,000 enrollees)	23.6	11.3	0.1	0.5	0.89
Elixhauser Comorbidity Score	2.4	0.3	0.0	0.0	0.98

Note: Table shows results from estimating equation (1) by OLS on period 2 data; the regression includes strata fixed effects and lagged outcomes from period 1. Standard errors are heteroskedasticity robust. Control means and standard deviations are from period 2.“ Total episode spending if paid under FFS” consists of Medicare payments and patient cost-sharing over the entire episode. Complication rate is defined, as in Finkelstein et al. (2018), as the share of CJR-eligible patients who have at least one of eight underlying complications that go into the total hip arthroplasty / total knee arthroplasty 90-day complication measure used in the targeted quality score.

Table 2: Selection

	Voluntary Control	Voluntary Select-In	Voluntary Select-Out	P-Value of Select-In vs. Select-Out Difference
Panel A: Selection on Levels				
Total Episode Spending if paid under FFS	28,177	27,124	30,181	0.003
	(6,919)	(4,369)	(7,957)	
Episode Spending in Institutional Post Acute Care	5,783	5,522	6,696	0.02
	(3,464)	(2,238)	(3,607)	
Share Discharged to Institutional Post Acute Care	42.2%	43.2%	44.6%	0.59
	(20.9%)	(19.1%)	(18.3%)	
Panel B: Selection on Slopes				
Change in Total Episode Spending		-772	424	0.07
		(2,180)	(5,274)	
Change in Institutional PAC Spending		-510	667	0.003
		(1,219)	(3140)	
Change in Share Discharged to Institutional PAC		-3.2%	3.8%	0.001
		(9.3%)	(14.6%)	
Panel C: Selection on Hospital Characteristics				
Above Median CJR Episodes	53.8%	68.5%	51.1%	0.02
Above Median Beds	47.1%	47.9%	46.2%	0.81
Teaching	11.2%	5.5%	11.4%	0.16
For-profit	18.8%	23.3%	17.3%	0.28
Non-profit	73.6%	68.5%	64.3%	0.53
Government-owned	7.6%	8.2%	18.4%	0.05
Share with Above Median Quality	55.6%	69.9%	47.8%	0.002
Northeast	8.5%	8.2%	1.6%	0.009
Midwest	30.5%	31.5%	50.0%	0.007
South	20.2%	12.3%	15.6%	0.51
West	35.0%	47.9%	32.8%	0.03
Sample Sizes				
Number of Hospitals	331	73	186	
Number of Episodes in Period 1	51,599	14,664	24,789	
Percent of Episodes in Period 1		37.2%	62.8%	

Note: Table reports means (and standard deviations in parentheses). In Panel A, all outcomes are defined in period 1 (although the target price is defined in period 2). Panel B reports results the average (and standard deviation) over different hospitals of $\beta_{1,h}$ from estimating equation (2) by OLS. In Panel C, the ''above median CJR episodes" comes from period 1 data, and "Above Median Beds", "Teaching", "For-profit", "Non-profit", and "Government" variables are all based on data from the 2016 American Hospital Association annual survey; We are unable to match 3 hospitals to these survey. For these outcomes, the number of hospitals in control, select-in, select-out are 329, 73, and 185, respectively. Finally, "share with above median quality" is based on a modified version of the hospitals' composite quality scores from period 2, which is based on the first 18 points of the score (see footnote 7). p-values of differences are computed based on a simple t-test of equality of the means.

Table 3: Parameter Estimates

Panel A: Parameter Estimates							
	Estimate	Std. Err.					
μ_{λ} *	10.14	TBA					
σ_{λ}	0.17	TBA					
γ	-0.07	TBA					
σ_{ε}	0.07	TBA					
μ_{ω}	5.00	TBA					
σ_{ω}	1.50	TBA					
$\rho_{\lambda,\omega}$	0.27	TBA					
ν	-50,000	TBA					
κ	0.58	TBA					
Panel B: Implied Distributions							
	E(x)	SD(x)	P5	P25	P50	P75	P95
B.1. All hospitals							
λ_h	25,802	4,333	19,341	22,738	25,445	28,475	33,477
λ_{h3}	24,189	4,374	17,721	21,090	23,803	26,865	31,973
ω_h	457	1,332	13	54	148	408	1,750
B.2. Only Voluntary Treatment Hospitals							
λ_h	25,852	5,329	18,385	21,910	25,244	28,974	35,541
λ_{h3}	24,306	5,546	16,738	20,148	23,670	27,445	34,420
$t_h - \lambda_{h3}$	52	4,924	-8,207	-2,859	346	2,918	7,660
ω_h	500	1,724	13	55	150	420	1,788
$(t_h - \lambda_{h3}) + \omega/2$	302	4,903	-7,879	-2,579	593	3,115	7,905
ν/n_h^K	-1,668	981	-3,387	-1,865	-1,408	-1,143	-754

Note: Panel A reports parameter estimates from the model and Panel B reports some key summary statistics implied by these estimates for all 1,455 hospitals. Panel C reports these implied statistics separately for the 259 voluntary treatment hospitals (see Figure 1). The model is estimated weighting each hospital by the number of episodes, and the summary statistics in panels B and C are computed from hospital-level simulated data, weighted by the number of episodes per hospital.

* Episode-weighted average of strata-specific estimates.

Table 4: Counterfactuals

	Share selecting in	Medicare Spending	Relative Social Costs	Ignoring Choice Shifter		Incorporating Choice Shifter	
				Relative Hospital Profit	Relative Social Surplus	Relative Hospital Profit	Relative Social Surplus
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Mandatory vs. Voluntary							
Mandatory FFS (Benchmark)	0.0%	24,306	0	0	0	0	0
Mandatory Bundled Payment	100.0%	24,358	60	302	242	-1,366	-1,426
Voluntary Bundled Payment	38.8%	26,007	1,956	1,807	-149	1,241	-715
Panel B: Alternative Voluntary Regimes with Different Target Prices							
Perfect targeting	19.5%	24,594	332	404	73	174	-157
Feasible targeting	28.0%	24,900	684	715	31	363	-320
Observed targeting	34.3%	25,305	1,149	1,121	-28	661	-488
No targeting	40.9%	26,110	2,075	1,914	-161	1,343	-732
Narrow bundle, no targeting	37.6%	25,483	1,354	1,290	-65	791	-564
Narrow bundle, observed targeting	32.2%	25,100	913	912	-2	498	-415

Note: All counterfactuals are done on 259 the voluntary treated hospitals (See Figure 1). We weight the hospital-level simulated data by the number of episodes per hospital, so that the statistics are representative of the average episode. In Panel A, row 1 reports results from the counterfactual in which BP does not exist and all hospitals are paid FFS, row 2 reports results from a mandatory participation bundled payment counterfactual, and row 3 reports results from the observed voluntary participation bundled payment regime. Panel B reports results from counterfactual voluntary participation regimes that vary in their target prices. Column 2 reports Medicare spending ("CMS" from Equation 4.1). All other columns report results relative to the FFS counterfactual. Column 3 reports relative social costs (i.e. $(1+\Lambda)((t_h-\lambda_h)$. Columns 4 and 5 report hospital profits and social surplus relative to the FFS counterfactual, under the assumption that ν_h is not welfare relevant; therefore hospital profits relative to FFS are given by $(t_h-\lambda_h)+\frac{\omega_h}{2}$, and social surplus relative to FFS is given by $\frac{\omega_h}{2}-\Lambda)(t_h-\lambda_h)$. Columns 6 and 7 report relative hospital profits and social surplus when ν_h is welfare relevant and therefore subtracted from both.

Voluntary Regulation: Evidence from Medicare Bundled Payments

Online Appendix

Liran Einav Amy Finkelstein Yunan Ji Neale Mahoney

A Maximum Likelihood Estimation

We estimate the parameters of the model using two-step simulated maximum likelihood. For notational convenience, let $H_C = \{h : BP_{h2} = 0\}$ be the set of control group hospitals, and let $H_{C,s}$ be the subset of control group hospitals in strata s. Similarly, let $H_T = \{h : BP_{h2} = 1\}$ be the set of treatment group hospitals. Finally, let H_V be the set of treatment group hospitals who were given the decision whether to voluntarily select into the BP program in period 3 and $H_M \equiv H_T \setminus H_V$ be the set of treatment group hospitals who were mandated to remain in the program.

We weight hospitals according to the average number of CJR episodes at a given hospital so that our estimates are representative of the average episode in our sample. Let w_h denote these hospital importance weights and let $W_C = \sum_{h \in H_C} w_h$ denote the sum of control hospital weights and $W_{C,s} = \sum_{h \in H_{C,s}} w_h$ be the sum of hospital weights in strata s.

In the first step, we estimate the parameters $\theta_1 = \{\mu_\lambda, \sigma_\lambda, \gamma, \sigma_\epsilon\}$ that determine the evolution of λ using data from the control group alone. The first step log likelihood is

$$\ln L = \sum_{h \in H_c} w_h \ln f(y_{h1}, y_{h2} | \theta_1) \tag{5}$$

where $\{y_{h1}, y_{h2}\}$ are data and f is their joint density function.

Our distributional assumptions imply that $\ln y_{h1}$ and $\ln y_{h2}$ are jointly normally distributed, allowing us to derive the maximum likelihood estimators in closed form. The first step maximum likelihood estimators are:

$$\widehat{\gamma} = \frac{1}{W_C} \sum_{h \in H_C} w_h \left[\ln y_{h2} - \ln y_{h1}\right] \tag{6}$$

$$\widehat{\mu_{\lambda,s}} = \frac{1}{2W_{C,s}} \sum_{h \in H_{C,s}} w_h \left[\ln y_{h1} + \widehat{\gamma} + \ln y_{h2}\right] \quad \text{for} \quad s = 1 \ldots 8 \tag{7}$$

$$\widehat{\sigma_\lambda} = \sqrt{\frac{1}{W_C} \sum_{h \in H_C} w_h (\ln y_{h1} - \widehat{\mu_{\lambda,s}} + \widehat{\gamma})(\ln y_{h2} - \widehat{\mu_{\lambda,s}})} \tag{8}$$

$$\widehat{\sigma_\epsilon} = \sqrt{\frac{1}{2W_C} \left[\sum_{h \in H_C} w_h \left[(\ln y_{h1} - \widehat{\mu_{\lambda,s}} + \widehat{\gamma})^2 + (\ln y_{h2} - \widehat{\mu_{\lambda,s}})^2\right]\right] - \widehat{\sigma_\lambda}^2} \tag{9}$$

In the second step, we estimate the remaining parameters $\theta_2 = \{\mu_\omega, \sigma_\omega, \rho_{\lambda\omega}, \nu\}$ of the joint distribution of $\{\lambda, \omega\}$ and choice shifter ν, conditional on the first step estimates $\widehat{\theta}_1$. Because the control-group hospitals provide no information about these second-step parameters, the entire second step relies only on the hospitals in the treatment group. The second-step log likelihood is thus given by

$$\ln L = \sum_{h \in H_V} w_h \ln g_1(BP_{h3}, y_{h2} | \widehat{\theta}_1, \theta_2, y_{h1}) + \sum_{h \in H_M} w_h \ln g_2(y_{h2} | \widehat{\theta}_1, \theta_2, y_{h1}) \tag{10}$$

where g_1 is the joint distribution of $\{BP_{h3}, y_{h2}\}$ and g_2 is the marginal distribution of y_{h2}.

Operationally, we estimate the joint distribution by decomposing it into the conditional choice probability and marginal distribution

$$\ln g_1(BP_{h3}, y_{h2} | \widehat{\theta}_1, \theta_2, y_{h1}) = \ln \Pr(BP = BP_{h3} | \widehat{\theta}_1, \theta_2, y_{h1}, y_{h2}) + \ln g_2(y_{h2} | \widehat{\theta}_1, \theta_2, y_{h1}), \tag{11}$$

so that

$$\ln L = \sum_{h \in H_V} w_h \ln \Pr(BP = BP_{h3} | \widehat{\theta}_1, \theta_2, y_{h1}, y_{h2}) + \sum_{h \in H_T} w_h \ln g_2(y_{h2} | \widehat{\theta}_1, \theta_2, y_{h1}). \tag{12}$$

Since neither the choice probability nor the marginal distribution of y_{h2} has a closed-form solution (because they depend on $y_{h2} = \lambda_{h2} - \omega_h$, which is the difference of two log normally distributed variables), we need to use simulation to construct the likelihood function. For a given set of parameters θ_2, we simulate many values for λ_{h2} and ω_h. We then estimate the marginal density using a kernel estimator and multiple it by the choice probability to construct the simulated likelihood.

We tested our estimator by simulating data based on known parameters and then estimating the model on these simulated data. We found that the estimator performed well when we draw X values for each set of candidate parameter values, and estimate the marginal density function g_2 using a Epanechnikov kernel with a bandwidth of X. We maximize the likelihood by conducting a grid search over θ_2, which in testing we found worked more reliably than other methods.

B Target Price Distribution

For the estimation of the model we use the observed target prices. However, in order to explore the impact of better targeting in a systematic fashion in our counterfactual exercises, we approximate the observed target prices using a parametric distribution, and then change these parameters. In this appendix we provide more details about this exercise.

We model target prices as log normally distributed, such that they are correlated with hospitals costs λ_h (but only correlated with ω_h via the correlation between λ_h and ω_h). Since target prices are partially based on lagged hospital spending, and we allow mean hospital spending to vary by strata, we also allow the mean of the target price distribution to vary by strata.

Following the notation in Appendix A, let $H_T = \{h : BP_{h2} = 1\}$ be the set of treatment group hospitals. We weight hospitals according to the average number of CJR episodes at a given hospital. Let w_h denote these hospital importance weights and let $W_T = \sum_{h \in H_T} w_h$ denote the sum of treatment hospital weights and $W_{T,s} = \sum_{h \in H_{T,s}} w_h$ be the sum of hospital weights in strata s.

We only observe target prices for treatment group hospitals in period 2. Using these data, the maximum likelihood estimators for the mean and standard deviation of the log target price distribution are given by:

$$\widehat{\mu_{t,s}} = \frac{1}{W_{T,s}} \sum_{h \in H_{T,s}} w_h \ln t_{h2} \quad \text{for} \quad s = 1 \ldots 8 \tag{13}$$

$$\widehat{\sigma_t} = \sqrt{\frac{1}{W_T} \sum_{h \in H_T} w_h \left(\ln t_{h2} - \widehat{\mu_{t,s}}\right)^2}. \tag{14}$$

We estimate $\rho_{\lambda t}$ using the covariance between log spending in period 1 ($\ln y_{h1}$) and the period 2 log target price ($\ln t_{h2}$). The covariance is given by

$$Cov(\ln y_{h1}, \ln t_{h2}) = Cov(\ln \lambda_h - \gamma + \epsilon_{h1}, \ln t_{h2}) = Cov(\ln \lambda_h, \ln t_{h2}) = \rho_{\lambda t} \sigma_\lambda \sigma_t \tag{15}$$

where the ϵ_{h1} drops out because it is assumed to be independently drawn and thus $Cov(\epsilon_{h1}, \ln t_{h2}) = 0$.

It follows that the maximum likelihood estimator of the correlation is

$$\widehat{\rho_{\lambda t}} = \frac{1}{\widehat{\sigma_\lambda} \widehat{\sigma_t}} \frac{1}{W_T} \sum_{h \in H_T} w_h (\ln y_{h1} - \widehat{\mu_{\lambda,s}} + \widehat{\gamma})(\ln t_{h2} - \widehat{\mu_{t,s}}). \tag{16}$$

where $\widehat{\sigma_\lambda}$, $\widehat{\mu_{\lambda,s}}$, and $\widehat{\gamma}$ are the estimates described in Appendix A. For our counterfactuals, it will be more natural to adjust the correlation between λ_{h3} and $\ln t_h$ (than the correlation between λ_h and $\ln t_h$). This object is given by $\widehat{\rho_{\lambda_3 t}} = \frac{\widehat{\sigma_\lambda}}{\widehat{\sigma_\lambda} + \widehat{\sigma_\epsilon}} \widehat{\rho_{\lambda,t}}$.

To examine the impact of better targeting, we simulate data using different parameters for the target price distribution, and then examine how these alternative target prices impact selection, Medicare costs, hospital profits, and social surplus. We simulate data in two steps. First we draw values for $\{\lambda_{h3}, \omega_h\}$ conditional on the observed t_h and n_h. To do so, we draw from the unconditional distribution of $\{\lambda_{h3}, \omega_h\}$ and then keep the draws that imply an optimal selection decision such that $BP = BP_{h3}$ at these values.

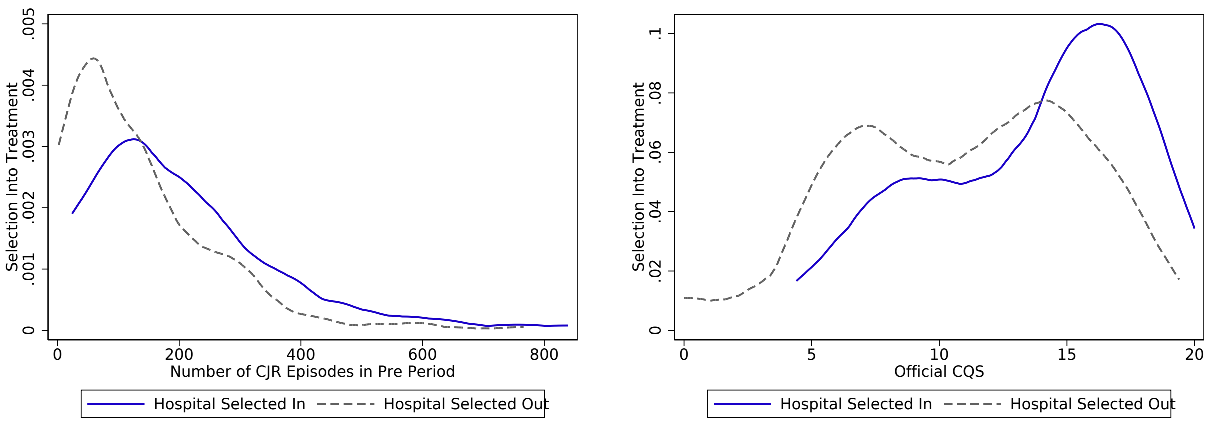
Second, we draw target prices for a given set of parameters, conditional on the simulated values of λ_{h3} from step 1. Let j be a counterfactual associated with the triplet $\{\mu_t^j, \sigma^j, \rho^j_{\lambda_3,t}\}$ for the distribution of target prices t_h^j. Since t_h^j and λ_{h3} are jointly log normal, the conditional values of t_h^j can be simulated as

$$t_h^j = \mu_t^j + \frac{\rho^j_{\lambda_3,t} \sigma_t^j}{\sigma_{\lambda,3}} (\lambda_{h3} - \mu_\lambda) + \sqrt{1 - (\rho^j_{\lambda_3,t})^2}\, \sigma_t^j\, \epsilon_h \tag{17}$$

where $\epsilon_h \sim N(0,1)$ is an independent, normally distributed random variable. Importantly, we keep the draws of ϵ_h for each hospital fixed throughout the set of counterfactual exercises, so that differences in outcomes across exercises are not driven by simulation variation.

Figure A1: Selection on Hospital Characteristics

(a) Pre-Period CJR Episodes (b) Composite Quality Score



Note: Figure shows kernel densities of hospital characteristics by whether the hospital selected in or out of the CJR program in period 3.

Table A1: Autocorrelation Between Period 1 and Period 2

	(
Total Episode Spending if paid under FFS	
Episode Spending in Institutional PAC	
Share Discharged to Institutional PAC	

Note: Estimated on sample of control hospitals (N=791)