Voluntary Regulation:
Evidence from Medicare Payment Reform[*]

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

May 13, 2020

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, predetermined payment for a sequence of related healthcare services, instead of separate service-specific payments. This "bundled payment" 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 substantially reduce these inefficient transfers. Our analysis highlights key design elements to consider under voluntary regulation.

[†]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 some government programs allow these actors to voluntarily decide whether they would like to participate in the program. 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 need only 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 actors 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, can simply receive a higher government transfer, the resulting "selection on levels" could lead to higher government spending without the desired behavior change. Thus, the extent to which a voluntary program is more or less socially desirable depends critically on the nature and extent of selection into the program.

We explore these tradeoffs in the context of the U.S. Medicare program, the public health insurance program for the elderly and the disabled. Over the last decade, Medicare has rapidly expanded the use of alternative models for reimbursing healthcare providers, such as Accountable Care Organizations, bundled payments, and primary care coordination models. By 2016, over 30% of Traditional Medicare spending was based on alternative payments models (Shatto, 2016). With the partial exception of the payment model we study, provider participation in all of these payment models has been voluntary (GAO, 2018). However, there is an ongoing and active debate over whether these programs should be made mandatory (e.g., Gronniger et al., 2017; King, 2019; Frakt, 2019*b*; Levy, Bagley and Rajkumar, 2018; Liao, Pauly and Navathe, 2020).

We analyze the Medicare bundled payments 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 and $10.7 billion in Medicare spending in 2014. Under

bundled payments, Medicare makes a single payment (known as the target price) to the hospital for all services related to the episode of care, including the initial hospital stay and physician fees, as well as any subsequent care by other medical providers during the recovery period. By contrast, under the status quo Fee-for-Service (FFS) system, Medicare makes separate payments to different providers based on the care provided. The idea behind the bundled payments reform is that by making the hospital the residual claimant on the costs related to the entire episode of care, it will internalize the incentives to provide care efficiently, including coordination with downstream providers. In practice, this was implemented by providing financial bonuses or penalties – known as "reconciliation payments" – to hospitals when the submitted Medicare FFS claims for the episode deviated from the target price.

CJR was initially designed by Medicare administrators as a mandatory participation, 5-year randomized trial. Randomization was conducted at the Metropolitan Statistical Area (MSA) level. In the 67 treatment MSAs, hospitals were paid under the bundled payments program. In the 104 control MSAs, hospitals were paid under the status quo FFS system. The program was implemented as designed in April 2016. 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 (CMS, 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 submitted Medicare claims, driven predominantly by reduced discharges to post-acute care (PAC) facilities (Finkelstein et al., 2018; Dummit et al., 2018; Barnett et al., 2019; Haas et al., 2019). Once reconciliation payments were taken into account, however, the mandatory bundled payment regime had no effective impact on government expenditures. We then examine the nature of selection into the program once it became voluntary. Consistent with selection on levels, we find that hospitals with lower claims under FFS – who, holding behavior constant, would benefit more from bundled payments – are more likely to opt in. Consistent with selection on slopes, we also find that hospitals that achieved larger reductions in claims under

mandatory bundled payments are more likely to opt in when it becomes voluntary.

Motivated by these patterns, we specify and estimate a stylized model of responsiveness to and selection into the bundled payment program. In the model, hospitals are characterized by a hospital-specific "level" (average claims per episode under FFS incentives) and a hospital-specific "slope" (the reduction in claims under bundled payments). Under a voluntary regime, the selection decision depends on the hospital-specific "target price" – the bundled payment the hospital receives from the government 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 that average episode claims under the status quo FFS incentives would be about $24,000 and that bundled payments reduced claims, on average, by about $400 per episode. These averages, however, mask substantial heterogeneity across hospitals in both levels and slopes. Heterogeneity is particularly large in levels, where the standard deviation across hospitals in claims under FFS incentives 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, once participation becomes voluntary.

We use the estimated model 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 payments designs. We define social welfare as the sum of consumer surplus and producer (hospital) profits, minus the social cost of public spending. We assume that consumer surplus is not affected by the payment regime; this is consistent with evidence from the randomized trial that healthcare quality, patient mix, and patient volume did not change with bundled payments (Finkelstein et al., 2018; Dummit et al., 2018; Barnett et al., 2019; Haas et al., 2019). We define the social cost of public spending as government (i.e., Medicare) spending multiplied by the shadow cost of public funds, which we assume (conservatively) to be 0.15. This "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.

The model estimates, like the reduced form descriptive work, indicate substantial selection on levels: hospitals that opt into the voluntary bundled payment regime would have had (counterfactual) average claims per episode under the FFS status quo that were substantially lower than hospitals which do not opt in. Moreover, the magnitude of the favorable selection (on "slopes") is too small to offset it. As a result, we estimate that the voluntary bundled payments regime lowers social surplus relative to the FFS status quo, due to inefficient transfers from the government to hospitals.

We also show that better targeting could improve the net social welfare impact of the voluntary regime, by reducing the transfers and aligning better participation incentives. For example, we estimate that a perfectly targeted voluntary bundled payments regime would raise social surplus relative to the status quo. While perfect targeting may be difficult to achieve in practice, we show that compared to the observed targeting, other feasible designs – based on better exploiting existing information or defining the bundle more narrowly – can come close to generating the social gains from perfect targeting, generating up to three-fifths of the social welfare 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 program we study, when participation was randomly assigned and mandatory (Finkelstein et al., 2018; Dummit et al., 2018; Barnett et al., 2019; Haas et al., 2019). It also includes evaluations of the much larger number of voluntary participation bundled payment programs for a host of conditions, including coronary bypass, prenatal care, cancer, and hip and knee replacement.[1] It is well understood that non-random selection into voluntary models can bias the evaluation of these program (e.g., Gronniger et al., 2017; Levy, Bagley and Rajkumar, 2018). Our focus here is on how voluntary participation affects the actual impact of the program, rather than the estimated impact.

Within health economics more generally, our work contributes to the growing literature on the impact and optimal design of financial incentives for healthcare providers (e.g., Cutler, 1995; Clemens and Gottlieb, 2014; Ho and Pakes, 2014; Einav, Finkelstein and Mahoney, 2018; Eliason

---

[1]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).

et al., 2018). It also relates 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; Marone and Sabety, 2019); here, we examine selection on moral hazard from the provider side rather than the consumer side.

More broadly, our emphasis of the potential for selection not only on levels but also on slopes relates to work in labor economics on selection on gains (Heckman and Honore, 1990), as well as to work on selection into voluntary regulation in other sectors. For example, recent work has analyzed selection by landowners into voluntary incentive programs for providing environmental services (Jack and Jayachandran, 2019), by polluting firms into whether to pay taxes based on their disclosed and verifiable emissions or the average emission rate among non-disclosers (Cicala, Hémous and Olsen, 2020), by private schools into whether or not to accept public vouchers (DeAngelis, Burke and Wolf, 2019), and by residential electricity consumers into whether to face a constant or time-varying regulated electricity price schedule (Ida, Ito and Tanaka, 2020).

The rest of the paper proceeds as follows. Section 2 provides background on our setting. Section 3 describes the data and presents reduced form, descriptive evidence of the impact of bundled payments under mandatory participation 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 the public health insurance program for the elderly and the disabled in the United States. We focus on the Traditional Medicare program, which provides coverage to about two-thirds of enrollees. In 2017, Traditional Medicare (hereafter "Medicare") had 38.7 million enrollees and annual expenditures of $377 billion (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 for each post-operative visit by the surgeon, or for renting a wheelchair during the recovery period. Moreover, within most of these categories, the payment would depend on the specific services provided.[2]

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, 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 alternative models (Shatto, 2016).

Our focus is on bundled payments, which represent a middle ground between FFS and fully capitated models, such as ACOs, in which providers are paid a fixed per capita amount per annum. Under bundled payments, Medicare makes a predetermined, single payment to one provider 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 government savings associated with bundled payments (e.g., 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). However, voluntary participation

---

[2]One exception to this system is hospital reimbursements. Starting in 1982, Medicare adopted the Prospective Payment System (PPS) in which it makes a fixed payment for the hospital stay based on the patient's diagnosis-related group (DRG) (Cutler, 1995).

makes separating treatment from selection difficult, and the small number of participating hospitals raises concerns about generalizability (e.g., Gronniger et al., 2017; King, 2019).

2.2 Comprehensive Care for Joint Replacement (CJR)

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 referred in the medical literature as lower extremity joint replacement, or LEJR) is a large Medicare category; in 2014, the year before CJR was announced, Medicare had almost half a million LEJR procedures, accounting for about 5% of Medicare admissions and inpatient spending (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. Medicare pays the hospital a predetermined *target price* for the episode. The hospital is then financially responsible for medical claims over the entire episode (except for care that is deemed as obviously unrelated). By contrast, under the status quo FFS regime, Medicare pays the hospital a fixed amount for the hospital stay, and reimburses the surgical procedure and post-discharge care separately based on those providers' submitted claims.

The level and targeting of the target price are key design elements in a bundled payment program. Let t_h denote the average per episode target price at hospital h in a given year, and let y_h denote average per-episode claims submitted that year. Under FFS, Medicare pays y_h on average. Under bundled payments, Medicare pays t_h on average. More specifically, under bundled payments, providers continue to submit claims and receive reimbursement of y_h on average as if they were under FFS, allowing us to observe y_h even under bundled payments. At the end of the year, hospitals under bundled payments receive a "reconciliation payment" of $t_h - y_h$ per episode, so that the gross Medicare payment is t_h.

Under CJR, Medicare tried to set the target price before each program year to be slightly lower than expected per-episode claims under FFS.[3] To do so, the target price included a small discount off a weighted average of historical hospital and regional (defined by the 9 census divisions) per-episode claims from three prior reference years, with the weight on the regional component increased over time from one-third in the first two years of the program to 100% in the last two

---

[3]In particular, Medicare set hospital-specific target prices for four severity groups determined by the 2-by-2 interaction of the patient's DRG (469 or 470) and whether the patient had a hip fracture.

years.[4] The "discount factor" was designed to reflect Medicare's portion of expected savings from CJR.[5]

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.[6] However, in practice the quality standard was not binding for the vast majority of the hospitals.[7] In addition, prior research has not detected effects on either incentivized or non-incentivized measures of quality (Finkelstein et al., 2018).

Second, like most bundled payment programs, CJR is not a "pure" bundled payment model that exposes hospitals and Medicare to unbounded risk. Rather, to limit risk exposure by both hospitals and Medicare, the reconciliation payment was subject to stop-loss and stop-gain provisions. In particular, if $t_h - y_h$ was less than (the negative of) the stop-loss amount, the hospital "only" has to pay Medicare the stop-loss amount. Similarly, if $t_h - y_h$ is greater than the stop-gain amount, Medicare "only" has to pay the hospital the stop-gain amount. The stop-loss and stop-gain amounts increased over time.[8] These provisions complicate the model and its estimation presented later, but do not affect the qualitative economic analysis and do not seem to be quantitatively important (see Appendix C and Appendix Table A3), so we abstract from them throughout the main text.

---

[4]The three reference years of historical claims are updated every other year. In 2016 and 2017 (the first two years of the program) historical claims from 2012 to 2014 were used, in 2018 and 2019 (the third and fourth program years), historical claims from 2014-2016 were used, and the final year of the program uses claims from 2016-2018 (CMS, 2015*b*).

[5]The discount factor ranged from 1.5% to 3% depending on hospital quality (based on a composite quality score defined below), with smaller discounts for higher quality hospitals (CMS, 2015*b*). This "discount factor" was intended to ensure that if claims remained at past levels, Medicare expenditures would decrease under mandatory participation. In practice, however, there was a steady secular decline in per-episode claims, which offset the built-in "discount."

[6]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

[7]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.

[8]In the first year, the stop-gain amount was set as 5% of t_h and the stop-loss was zero (meaning that hospitals would never need to make payments to Medicare). By years 4 and 5, the stop-gain and stop-loss amounts were each scheduled to be set at 20% of t_h.

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 payments) or control (status quo FFS). Specifically, MSAs were divided into 8 strata based on the interaction of historical LEJR spending quartile and above- versus below-median MSA population. MSA treatment probabilities varied by strata (ranging from 30% to 45%), with higher treatment probabilities for strata with higher historical LEJR payments. CMS announced assignment to treatment and control in the July 2015 Federal Register (CMS, 2015*b*). Treatment and control MSAs are balanced on outcome variables and MSA characteristics (Finkelstein et al., 2018). 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 (see Section 3 for more details).

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 lead 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 payment models (Advancing Care Coordination Through Episode Payment and Cardiac Rehabilitation Incentive Payment Models) and modified CJR to be voluntary in half of the treated MSAs for the remaining three program years (CMS, 2017).

Specifically, hospital participation in CJR was made voluntary in 33 of the 67 treatment MSAs

---

[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 these criteria. Since these exclusions were based on hospital decisions made prior to assignment we simply drop these 25 MSAs from the study.

[10]See https://thehill.com/policy/healthcare/298769-lawmakers-call-for-end-to-medicare-experiments.

with the lowest historical episode claims under FFS.[11] In these "voluntary treatment" MSAs, hospitals 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 for the remaining 3 program years. If they did not opt in, reimbursement would revert to FFS for the remaining 3 years. Slightly more than one-quarter of the hospitals in the voluntary bundled payment MSAs (73 out of 279) chose to remain under bundled payments. In the 34 mandatory bundled payment 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 approximately 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 shows a flow chart 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 were further divided into mandatory and voluntary treatment groups. Because this division was based on predetermined 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.

---

[11]Specifically, this determination was made using spending over July 1, 2011 through June 30, 2014, which was the same period used to determine MSA eligibility in CJR, and occurred entirely before random assignment was announced (CMS, 2017).

3.1 Data and Sample

Our main data are the 100% Medicare enrollment and claims files. These contain basic demographic information (age, race, sex, and Medicaid enrollment) and claims for inpatient, outpatient, and post-acute care.[12] The claims data include information on Medicare payments made to providers and out-of-pocket payments owed, dates of admission and discharge, diagnoses, and discharge destinations.

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

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 a 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 payment model for LEJR.[14] 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.[15,16]

We define period 2 (the period of mandatory participation bundled payments) to include all episodes with an index admission 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

---

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

[13]Source: https://innovation.cms.gov/initiatives/CJR. Target prices and reconciliation payments for 2018 were not available at the time of writing.

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

[15]In Table 1, we conduct analysis that expands the sample to include readmitted patients and show there is no effect on readmissions.

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

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 (CMS, 2017). Following prior work (Finkelstein et al., 2018), we define period 1 (the period when all hospitals are under FFS) to include all episodes admitted between April 1, 2013 and September 15, 2014, and omit 2015 from the analysis to avoid contamination from potential anticipatory effects; treatment and control MSAs were announced in July 2015 (CMS, 2015*a*).

To construct our baseline sample, we start with the universe of 1,570 hospitals in the 171 treatment and control MSAs that had a CJR episode during period 2; these hospitals had a total of 395,938 CJR episodes during period 2. So that we can observe outcomes in the years prior to the intervention, we restrict the sample to the 1,455 hospitals that have at least one episode in both years of period 1. 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; Dummit et al., 2018; Barnett et al., 2019; Haas et al., 2019); we reproduce some key results here. Since the program was mandatory and assignment was random at the MSA level, we follow Finkelstein et al. (2018) and estimate:

$$outcome_{j2} = \beta_0 + \beta_1 BP_j + \beta_2 outcome_{j,2014} + \beta_3 outcome_{j,2013} + \delta_{s(j)} + \epsilon_j \tag{1}$$

where $outcome_{j2}$ 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 payments. We include lagged outcomes from 2013 and 2014 (i.e., period 1) as controls to improve statistical power. Because the probability of random assignment to treatment varied across strata, we include strata fixed effects, $\delta_{s(j)}$, to isolate the experimental variation. In all tables, we report heteroskedasticity robust standard errors.

---

[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 September 15, 2017 guarantees that virtually all claims associated with the episode would be included in the 2017 claim files (CMS, 2016).

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 Claims, Use and Government Spending. Panel A of Table 1 examines effects on healthcare claims, healthcare use, and government spending per episode. "Claims" consist of Medicare claims paid and patient cost sharing owed over the entire episode of care, but do not account for any reconciliation payment associated with a bundled payment; they thus correspond to y_h in the notation from Section 2.2, and are measured as average per episode claims. Average per-episode claims in the control group were about $25,300, with roughly half of this spending on the index admission, which is already reimbursed with a DRG-based, prospective payment under the status quo. Of the remaining $11,800, about $4,100 comes from post-discharge claims for institutional Post Acute Care (PAC) – predominantly skilled nursing facilities – $1,800 represents post-discharge claims for Home Health Care, and the remaining $5,800 includes categories such as claims for the surgeon and other physicians (both inpatient and outpatient), hospice, and durable medical equipment, such as wheelchair rental.

We estimate that bundled payments reduced average episode claims by about $800, or about 3 %, a statistically significant but economically modest result. This reduction is primarily driven by a statistically significant $500 decline in claims for institutional PAC (12 % of the control mean), with no statistically or economically significant effects on other categories of claims. The impacts on claims are reflected in the impacts on utilization. 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 (8 % of the control mean).

The decline in use of institutional PAC in turn reflects at least in part an extensive margin response in where patients are discharged to following their index admission. In the control group, patients are discharged to institutional PACs, home with home health care, and home without home health care in equal proportion. Bundled payments reduced discharges to institutional PAC by a statistically significant 3.4 percentage points (11 %). The decline in discharges to institutional PAC is accompanied by a similarly sized increase in discharges to home without home health.[18]

---

[18]This can be either because the patients who would have been sent to institutional PAC are being sent home *without*

These experimental estimates are consistent with qualitative evidence on how hospitals respond to CJR. In a survey of hospital executives and administers, Zhu et al. (2018) find that hospitals report responding 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.

Although bundled payments reduced episode claims as they would be paid under FFS, actual government spending (claims, y_h, under FFS and the target price, t_h, under bundled payments) does not change. The point estimate indicates a statistically insignificant increase in average government spending per episode of $33 (standard error of $208). The lack of a reduction in government spending – despite modest reductions in submitted claims – reflects the design feature of bundled payments, according to which target prices were set to approximate counterfactual claims under FFS.

(Lack of) Quality Shirking and Cream Skimming. The primary concern with bundled payments is that because providers are no longer paid on the margin, they will cut back on medically necessary care, cherry-pick patients who have lower costs of provision, or both. 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 increases downstream costs, hospitals may have incentives to improve quality.

Consistent with prior work on CJR, 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 three measures of quality: 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. Panel C examines patient volume and 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.[19] 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

---

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 more likely (to us), but we cannot differentiate between these two mechanisms.

[19]Specifically, while our analysis of LEJR admissions includes non-CJR eligible admissions in the sample of 171 eligibles MSAs, our analysis of CJR-eligible admissions excludes hospitals and episodes in these MSAs that are not eligible for CJR.

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 as a non-discretionary procedure, or at least a procedure where the change in financial incentives from bundled payments 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 skimming responses could potentially be different in period 3 once participation is voluntary, as there may now be hospitals paid under bundled payments 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 revenue.

Since our paper focuses on selection – which is a *hospital-level* choice – all of our remaining analyses are conducted at the hospital level, with hospitals weighted by the number of episodes in period 2 so that the results are representative for the average episode. Appendix Table A1 shows that the main results from Table 1 are largely similar when estimated at the MSA level weighted by the number of episodes, or at the hospital level weighted by the number of episodes, although some of the point estimates shrink in magnitude. In addition, one of our three quality measures (the complication rate) shows a statistically significant increase of 0.2 percentage points (off a base of 1.1 %) in one of the three specifications in Appendix Table A1.

In interpreting this evidence, it is important to point out that the quality measures are limited. We cannot, for example, measure outcomes such as morbidity, mobility, or activities of daily living consistently across locations. We therefore cannot definitively rule out any impact of bundled payments on quality of care. However, given the lack of compelling evidence of a quality response on the margins we can observe, 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.

3.3 Selection on Levels and Slopes

As we formalize in the next section, 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, holding behavior fixed at what it would be under FFS, would be 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 average claims below the target price. We present descriptive evidence on both margins, examining how the decision among voluntary treatment hospitals to select in or out of bundled payments in period 3 correlates with episode claim levels in period 1 and behavioral responses to bundled payments in period 2. Table 2 presents the results.

Selection on Levels.   In principle, we would want to examine selection on slopes by estimating the correlation between the selection decision and period 3 spending in the absence of the incentives from bundled payments. Because treated hospitals change behavior in response to bundled payments (see Table 1), we do not observe unaffected spending levels. The model developed in the following sections will allow us to formally recover these counterfactual spending levels. To provide model free evidence, for now we simply look at how selection correlates with the hospital's average episode claims in period 1, when all hospitals were paid under FFS. Since spending levels are strongly auto-correlated within hospitals over time, we expect hospitals with lower average episode claims in period 1 to be more likely to select in (on levels) in period 3.[20]

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 have been assigned to voluntary based on their prior spending levels (see Figure 1). Columns 2 and 3 show period 1 outcomes for hospitals in the voluntary treatment group, split by those who in period 3 selected into bundled payments (column 2) or out of bundled payments (column 3). The first three rows report results for the three outcomes where we observed a statistically significant impact of bundled payments in Table 1: episode claims, claims for 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, average episode claims are $26,523, with a standard deviation of $5,368. Columns 2 and 3 show that, as expected, hospitals who select into bundled payments have, on average, about

---

[20]Specifically, for control group hospitals, the correlation coefficient between period 1 and period 2 is 0.77 for total episode claims, 0.65 for institutional PAC claims, and 0.71 for the share discharged to institutional PACs.

$1,600 lower average episode claims than those who select out, a statistically significant difference that is about 6 % of the control mean. The patterns are similar for claims at and the share discharged to institutional PAC.

Selection on Slopes. To assess selection on slopes, we estimate hospital-specific slopes (i.e., behavioral changes in response to bundled payments in period 2) and then examine how selection into bundled payments in period 3 varies with this measure. Letting $outcome_{h2}$ denote the average episode outcome for hospital h in period 2, we estimate a modified version of Equation 1 that allows the treatment effect of bundled payments to vary by hospital. Specifically, we estimate:

$$outcome_{h2} = \beta_0 + \beta_{1h}BP_h + \beta_2 outcome_{h,2014} + \beta_3 outcome_{h,2013} + \delta_{s(h)} + \epsilon_h, \qquad (2)$$

where BP_h is an indicator for being randomly assigned to bundled payments and β_{1h} is the hospital-specific treatment effect. As in Equation 1, 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 also 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).

Panel B of Table 2 shows the results. Specifically, we show the average estimated hospital-specific treatment effects (β_{1h}) and their standard deviation separately across hospitals that select into bundled payments (column 2) and those that select out (column 3). The results once again show selection in the expected direction: for hospitals that selected into bundled payments in period 3, bundled payments in period 2 reduces average claims per episode by $796, compared to a lower reduction ($669) for hospitals that revert to FFS in period 3. However, these differences in average slopes are not statistically distinguishable (column 4). Selection is also in the expected direction for the other two outcomes in Panel B: hospitals that experienced greater declines in institutional PAC claims and in the share of patients discharged to institutional PAC due to bundled payments are more likely to remain under bundled payments. The differences in the impact on institutional PAC claims in period 2 between those who remain in bundled payments ($519) and those who select out ($177) in period 3 is statistically distinguishable (p=0.05).

The random assignment of payment regime in period 2 is not particularly important when

17

estimating heterogeneity in treatment effects across hospitals (i.e., the β_{1h}'s from Equation 2); the control hospitals simply identify the secular time trend β_0. Rather, estimation requires the (admittedly strong) identifying assumption that, conditional on the covariates, there are no hospital-specific time trends. Any heterogeneity in the change in outcomes across hospitals is interpreted as reflecting heterogeneous treatment effects. To probe the sensitivity to this assumption, we estimate an alternative specification in which we include hospital-specific linear time trends as controls. Theses time trends are identified off the hospital specific outcomes in 2013 and 2014. In this specification, the (weaker) identifying assumption is that, conditional on covariates, there are no hospital specific deviations from the time trend. Appendix Table A2 shows that the qualitative pattern is robust to this alternative specification: hospitals that select into bundled payments in period 3 experienced larger changes in outcomes in period 2 than those who opt out, although none of these differences are statistically distinguishable.

Selection on Hospital Characteristics.   Panel C of Table 2 briefly examines other characteristics of hospitals that select in and select out of bundled payments. Hospitals that select into bundled payments in period 3 have a somewhat higher volume of CJR episodes in period 1, suggesting there may be fixed costs to remaining in the program, a point we return to with our model specification in Section 5.[21] Hospitals that select in are less likely to be teaching hospitals, more likely to be for-profit, and less likely to be government-owned; they are also associated with higher measured quality.

4   Model of Voluntary Selection

4.1   Setting

We consider a pool of CJR episodes, 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 FFS, providers are reimbursed based on claims. Let λ_i denote the claims generated under FFS incentives by a given episode. The preceding sections' description of the institutional environment and the estimates of the average effects of bundled payments 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

---

[21] We use period 1 volume since it is necessarily unaffected by period 2 treatment assignment.

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 $f_i^{OTH} = c_i^{OTH}$.[22] 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 Participation Incentive. Under FFS, average government spending (i.e., Medicare reimbursement) per episode is $\lambda_h = f_h^{HOSP} + f_h^{OTH}$. Hospitals only incur costs and receive Medicare payment for costs within the hospital, so they earn profits $\pi_h^{FFS} = f_h^{HOSP} - c_h^{HOSP}$. Under bundled payment, Medicare reimburses the admitting hospital the fixed target price t_h for the entire episode, so average government spending per episode is t_h. Hospitals are, effectively, required to incur not only hospital costs c_h^{HOSP} but also downstream providers' claims f_h^{OTH}, which would have been reimbursed by Medicare under FFS. We assume that the hospital can reduce claims outside of the hospital by e through the exertion of "effort" $\phi_h(e)$, where $\phi_h(0) = 0$, $\phi_h' > 0$, and $\phi_h'' > 0$.[23] Hospitals, thus, choose effort to maximize

$$\pi_h^{BP} = \max_e \left(t_h - \left[(c_h^{HOSP} + f_h^{OTH}) - e \right] - \phi_h(e) \right), \tag{3}$$

and optimal effort is pinned down by $\phi_h'(e_h^*) = 1$. Since the hospital internalizes both the social marginal cost and benefit of effort, bundled payments 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, under bundled payments the hospital's optimal choice of effort is $e_h^* = \omega_h$, average claims 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 bundled payment program, denoted by the indicator $BP_h =$

---

[22]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 reasonable levels of such markups would only slightly affect the quantitative results and would have no impact on the qualitative conclusions.

[23]For simplicity, we assume that c_h^{HOSP} remains the same under bundled payments 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.

1, if and only if $\pi_h^{BP} > \pi_h^{FFS}$. Substituting in yields the criteria

$$BP_h = 1 \Leftrightarrow (t_h - \lambda_h) + \frac{\omega_h}{2} > 0, \tag{4}$$

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 payment hospitals would receive from the government under bundled payments relative to under 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 bundled payments, which is the reduced provider costs $e_h^* = \omega_h$ net of the effort cost that reduction entails $(\frac{\omega_h}{2})$. 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 payments 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?".[24]

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

$$W = S + \pi - (1 + \Lambda)G. \tag{5}$$

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 government (or consumers) rather than in the hands of hospitals.[25] Consistent with the descriptive results in Section 3, we assume that the hospital's effort does not affect patient welfare (S).

---

[24]See https://www.arbormetrix.com/press-releases/new-report-significant-variation-in-expected-savings-per-hospital-in-bpci-advanced.

[25]Recall that y_h in practice measures claims paid by Medicare or owed out of pocket by consumers. Under this interpretation, it would be natural to multiply S by $(1 + \Lambda)$, so that Λ represents the wedge between hospitals on the one hand and consumers and government on the other. Since we net S out of the calculations below, this can be done without loss of generality.

Government spending per episode (G) is t_h under bundled payments and λ_h under FFS. Plugging in for these and for hospital profits implies that hospital participation in the bundled payments improves social welfare if and only if

$$W_{BP} > W_{FFS} \Leftrightarrow -\Lambda(t_h - \lambda_h) + \frac{\omega_h}{2} > 0. \tag{6}$$

Equation 6 illustrates the key social welfare tradeoff. On one hand, bundled payments 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 government spending by $t_h - \lambda_h$. Selection on levels ($t_h - \lambda_h$) is thus socially costly: it creates a transfer from the government to hospitals that is 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 payments, 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 government 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 target price t fixed across hospitals. At this payment, the solid line represents the set of hospitals that are indifferent between participation in bundled payments 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 payments ($\frac{\omega}{2}$) 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 $t_h - \lambda_h > 0$), as well as an additional "slope" incentive,

which explains why the solid line slopes up. All else equal, a higher ω_h provides an additional incentive for the hospital to join the bundled payment 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 payments or FFS. While the "slope" effect ($\frac{\omega}{2}$) enters identically (and positively) into the private participation condition (Equation 4) and the social welfare condition (Equation 6), the "level" effect $t - \lambda$ enters positively into the hospital's participation decision (Equation 4) but negatively in the social welfare calculus (Equation 6). 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: hospitals that choose the FFS regime, hospitals that efficiently select into bundled payments, and hospitals that select into bundled payments 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 bundled payment program aligns effort incentives. If Medicare could generate 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 these target prices, all hospitals would voluntarily participate and government 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 target price $t' > t$. The black (solid and dashed) indifference lines that correspond to t' are analogous to the gray lines, which correspond to t. Naturally, the higher payment amount increases the share of hospitals that select into bundled payments. For many of the marginal hospitals that opt in, participation increases social welfare. 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 underlying joint distribution of $\{\lambda_h, \omega_h\}$ as well as any information about this joint distribution that the social planner can condition on in setting target prices. This is shown in the remaining panels of Figure 3, which use ovals to illustrate three examples of underlying joint distributions, conditional on (priced) observables.

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 is lower and therefore optimal effort under bundled payments 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 λ_h is heterogenous after conditioning on available information, it is easier to generate social welfare enhancing participation in bundled payments. 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 less heterogeneity in λ_h, or more precise information, 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 heterogeneity in λ_h and ω_h. Because the primary policy instrument is a fixed payment, large heterogeneity in λ_h, as in Panel C, leads to more inefficient selection into bundled payments. In contrast, if the primary source of heterogeneity is the "slope" ω_h, as in Panel D, voluntary participation is more likely to generate social welfare gains.

The joint distribution of λ_h and ω_h, conditional on (priced) observables, is therefore key in assessing the potential social welfare gains from alternative payment models. It is therefore the key object of interest in our econometric exercise in the 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, all hospitals are still under FFS, and assignment to bundled payments has not been announced. In period 2, a subset of the hospitals are randomly assigned to bundled payments. In period 3, a subset of these latter hospitals endogenously choose to remain under bundled payments while the rest switch back to FFS.

For each hospital in the sample, we observe two periods of claims data (y_{h1}, y_{h2}) and three periods of bundled payments participation indicators $(BP_{h1}, BP_{h2}, BP_{h3})$. For hospitals randomly assigned to bundled payments in period 2, we also observe the hospital-specific target price, t_h, in period 2 (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 per-episode claims under FFS and ω_h is the reduction in claims that is caused by bundled payments. The model implies the following relationships between the latent objects of interest and the variables we observe. First, the claims 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 hospital's participation equation (which is relevant only for period 3, when participation in bundled payments 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

24

(iid) from $N(0, \sigma_\epsilon^2)$.[26]

Second, we introduce a hospital-level choice shifter into the participation equation. Survey evidence on how hospitals respond to bundled payments suggest that participation may be influenced by both hospital-level fixed costs and marginal per-episode 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 n_h is the number of episodes at the hospital, and 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 an episode-level specification. For $\kappa = 0$, the hospital-level choice shifter becomes ν, implying no episode-level component. Intermediate values of κ capture intermediate cases. For now, we remain agnostic as to what these additional forces represent and, in particular, whether they only affect hospital choice or are also relevant for social welfare. In our counterfactual analyses in Section 6 we will consider both possibilities.

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

$$y_{h1} = \lambda_{h1},$$

$$y_{h2} = \lambda_{h2} - BP_{h2}\, \omega_h,$$

$$BP_{h3} = 1 \iff (t_h - \lambda_{h3}) + \frac{\omega_h}{2} + \frac{\nu}{n_h^\kappa} > 0,$$

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

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.[27] Because the DRG portion of spending is fixed, we restrict ω_h from being infeasibly large by truncating it from above at $0.71\,\lambda_h$.[28]

---

[26]Our use of the term "linear" is slightly imprecise, because the trend is in periods, which do not perfectly correspond to calendar time (e.g., we exclude 2015 due to potential anticipatory effects).

[27]Since assignment to bundled payments 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.

[28]The institutions and empirical evidence indicate that virtually all of the savings from bundled payments arise from

5.2 Identification

The conceptual identification of the model is fairly straightforward given random assignment to bundled payments in period 2. The model has a set of parameters that correspond to the "level" of claims under FFS incentives and its evolution over time: $\mu_{\lambda s}, \sigma_{\lambda}, \gamma, \sigma_{\epsilon}$; a set of parameters that correspond to the reduction in claims under bundled payments: $\mu_{\omega}, \sigma_{\omega}$; a correlation parameter that relates the levels and slopes: $\rho_{\lambda\omega}$; and choice shifter parameters: ν, κ. The intuition for the identification argument follows in three steps.

First, using data from the control group alone, in which we observe λ_{h1} and λ_{h2} for the same set of hospitals, we can identify $\mu_{\lambda s}$, γ, σ_{λ}, and σ_{ϵ}. Intuitively, these would be identified from a random effects model estimated on the control group. We can use the control group alone to estimate these parameters 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 λ_{h2} for all hospitals in the control group, and λ_{h1} and $\lambda_{h2} - \omega_h$ for all hospitals in the treatment group. The average change between λ_{h1} and λ_{h2} for control hospitals identifies γ, the time trend in λ_{ht}. Again, because of random assignment we know that γ estimated from the control group is valid for the whole sample. For treatment hospitals, the average difference in λ_{h1} and $\lambda_{h2} - \omega_h$ in excess of γ identifies μ_{ω}.

The dispersion within the treatment hospitals in the change in episode claims between period 1 and period 2 is driven by a combination of the stochastic 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 for treatment hospitals identifies σ_{ω}. The intuition for identifying $\rho_{\lambda\omega}$ is similar: we observe the reduction in claims for each hospital in the treatment group, and can correlate it with the hospital's period-1 claims, 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 bundled payment participation equation identifies the distribution of the remain-

---

"other" claims that are not associated with fixed, DRG-based payment for the index admission. The value of 0.71 represents the 99th percentile in the data for the ratio of these other claims to total episode claims. By truncating the distribution of ω_h at the 99th percentile of other claims, we are essentially making the assumption that savings cannot exceed the 99th percentile of "other" spending.

ing 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 participation rate in a voluntary bundled payment program. Any deviation from this "predicted" participation 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 bundled payments.

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 average episode claims 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 to increase 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.

Standard errors are calculated using the bootstrap method. Specifically, we construct bootstrap samples by drawing the observed number of hospitals with replacement from cells defined

27

by the full interaction of treatment assignment and strata. When estimate the parameters of the model on each bootstrap sample and calculate the standard errors as the standard deviation of these bootstrap parameter estimates.

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, in Panel B.2, limited to the hospitals in the voluntary treatment group only (see Figure 1).

Panel A indicates a negative time 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 $386, which is smaller than the average effect estimated in Table 1 (and closer to the hospital-level estimates in Appendix Table A1). The distributions of λ_h and ω_h have a modest positive correlation ($\rho_{\lambda,\omega} = 0.30$). The estimate of κ is 0.40, implying that the choice shifter scales up with the number of episodes but less than proportionally, and can thought of as representing some combination of hospital-level and episode-level costs. With the exception of the choice-shifter parameter ν, the parameters are precisely estimated.

The model emphasizes the importance of heterogeneity in levels and heterogeneity in slopes in determining the nature of selection, government spending, and social welfare under a voluntary bundled payments regime. The results in Panel B.1 indicate that the heterogeneity in levels is substantially greater than either the level or the heterogeneity in the slopes – the standard deviation of λ_{h3} is $4,400 compared to a standard deviation of ω_h of $1,100, and an even lower average of ω_h. This raises concerns that the voluntary system may primarily produce inefficient transfers to hospitals through 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 hospitals in the voluntary treatment regime, 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 $5,700 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 hospitals in the voluntary treatment group. 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 Panel B.2 of Table 3, the plot suggests that selection on levels is a primary concern, with a large mass of hospitals selecting bundled payments 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 bundled payments 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 hospitals in the voluntary treatment group – i.e., treatment group hospitals that were given a choice in period 3 of whether to remain under bundled payments or revert back to FFS (see Figure 1). We perform two main sets of exercises. First, we compare social welfare under mandatory FFS, mandatory bundled payment, and voluntary bundled payment. Second, within the voluntary bundled payment regime, we consider how alternative target prices affect social welfare.[30]

---

[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 $Var(\lambda_{h3} - t_h) \approx Var(\lambda_{h3})$.

[30]The current bundled payment option offers the choice of only one target price, paired with incentives that make the hospital the full residual claimant on effort. It would be natural to explore social welfare under menus of contracts which trade off higher target prices in return for shallower incentives, as in the classic optimal regulation problem in Laffont and Tirole (1993). However, given that, as shown in Figure 4, selection on levels rather than slopes is the primary concern, and that the average slope is relatively modest, better targeting of the price level rather than better design of the screening contracts seems a more fruitful avenue to explore.

6.1 Voluntary vs. Mandatory

We first compare outcomes under the observed period 3 voluntary bundled payment program to two period 3 counterfactuals: all hospitals mandated to be under the status quo FFS regime (as they were in period 1), or all hospitals mandated to participate in the bundled payment program (as they were in period 2). These counterfactuals can be thought of as measuring the impact of the government's decision to make the bundled payment program voluntary, relative to cancelling the program entirely or keeping it mandatory.

To operationalize this counterfactual exercise, we use our model, together with the parameter estimates 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}, their target price t_h, and number of CJR episodes n_{h2} from period 2. We assume a social cost of funds of $\Lambda = 0.15$.

Panel A of Table 4 shows the results. The first row reports the results that would occur if there were no bundled payment program and all hospitals were paid under FFS. Government spending (i.e., G in the definition of social welfare from Equation 5) averages \$24,443 per episode in this counterfactual (which corresponds to the average λ_{h3} reported in Panel B.2 of Table 3). The remainder of the entries are normalized to zero; the mandatory FFS counterfactual will serve as a benchmark to which we compare other regimes.

The second row of Panel A considers a counterfactual in which all hospitals are mandated to enroll in bundled payments in period 3, as was intended under the initial design. Under mandatory bundled payment, hospitals receive a transfer of $(t_h - \lambda_h)$ and are residual claimants on the ω-related savings they generate ($\frac{\omega_h}{2}$). If target prices had been calibrated to equal counterfactual claims under FFS incentives ($\mathbb{E}_h t_h = \mathbb{E}_h \lambda_{h3}$), government spending would have been unaffected relative to the baseline. However, as seen in Panel B.2 of Table 3, target prices t_h were on average \$85 lower than FFS claims, so government spending is \$85 lower and (multiplied by $1 + \Lambda = 1.15$) social costs decrease by \$98 (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. The choice shifter would be welfare-relevant if it represents a real hospital cost - such as fixed or variable costs of changing behavior, or perhaps uncertainty about λ_h, which might be greater

for smaller hospitals. The choice shifter might not be welfare relevant, however, if it represents a choice friction such as status quo bias. In columns (4) and (5), where we assume ν_h is not welfare relevant, hospital profits relative to FFS rise by \$116 (column 4). This rise reflects the difference between the ω-related savings of \$201 (half of the expected value of ω_h from Panel B.2 of Table 3) and the \$85 reduction in government payments. Social surplus rises by \$214 (column 5). In other words, the incentive effects of bundled payment (which generate $\frac{\omega_h}{2} = \$201$ in social savings on average) represent most of the social gain, with the remainder coming from the reduction in government spending. 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 \$1,419 (see Panel B.2 of Table 3).

The third row of Panel A considers the voluntary selection scenario that actually took place. To be consistent with the other counterfactuals, 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 government spending, hospital profits, and social surplus. We find that 38.8% of episode-weighted hospitals select into bundled payments, which is almost identical to the actual selection percentage (37.2% 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 $\frac{\omega_h}{2}$, suggests that selection into bundled payment is primarily on "levels." Consequently, voluntary participation raises government spending relative to the other counterfactuals; specifically, we estimate that voluntary participation raises government spending per episode by \$1,682 relative to the mandatory FFS benchmark, and by \$1,767 relative to the mandatory bundled payment regime.

Since hospitals are given a choice, their profits must be weakly higher under the voluntary regime relative to a mandatory regime. When we treat the ν_h term as non-welfare relevant, hospital profits rise to \$1,754 above the mandatory FFS benchmark, which is \$1,638 higher than under mandatory bundled payments. Ignoring ν_h, social surplus under voluntary bundled payments is lower than under either the mandatory FFS benchmark (-\$179, column 5) or the mandatory bundled payment regime (-\$393); lower social surplus reflects both the larger transfers to hospitals and the smaller share of hospitals generating ω-related efficiency gains. When we treat ν_h as welfare relevant, hospital profits rise by a smaller amount of \$1,248 (instead of \$1,754) above the mandatory FFS benchmark, and social surplus is correspondingly lower. It is worth noting

31

however that, while negative, social surplus in the voluntary bundled payment counterfactual is less negative than under the mandatory bundled payment counterfactual (see column 7). Intuitively, if we think that the ν_h-related costs are real, it is important to let hospitals avoid them if the offsetting ω-related benefits are not large enough.

Whether or not we treat the ν_h term as welfare relevant, the results indicate that the voluntary bundled payment model reduces social surplus relative to the mandatory FFS status quo. However, the results do not imply that a voluntary program is necessarily ineffective. Voluntary bundled payments generates positive ω-related efficiency gains. The reason overall social welfare is reduced is because these ω-related gains are small relative to the social welfare losses associated with "overpaying" participating hospitals relative to their counterfactual FFS claims. We therefore turn now to exploring whether and how the voluntary program performance could improve if Medicare were able to set target prices to better reflect underlying hospital-specific costs and thus capture more of the potential \$100 million dollars of potential annual ω-related gains in social surplus (from approximately 500,000 CJR episodes per year, with an average $\frac{\omega_h}{2}$ of about 200).

6.2 Targeting

In order to explore price targeting under voluntary participation in a systematic fashion, we approximate the observed target prices using a parametric distribution, and then 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 parameter values. 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 government spending (x-axis) for different target price counterfactuals. 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, including analyses where ν_h is considered welfare-relevant. The black dot in Figure 5 corresponds to the observed distribution of t_h, and serves as a baseline. The observed target prices are parameterized by means μ_{ts} that vary by strata, a standard deviation of $\sigma_t = 0.17$, and a correlation of $\rho_{\lambda_3, t} = 0.81$

32

with claims under FFS (λ_{h3}).[31] Because of the log-normal parameterization, the outcomes for observed targeting in Panel B of Table 4 are slightly different to those in the voluntary bundled payments row of Panel A.

We consider three counterfactual targeting policies. The first, indicated by the point labeled "perfect targeting," sets target prices exactly equal to realized claims under FFS ($t_h = \lambda_{h3}$). Under this contract, there is no transfer to hospitals (no selection on levels) and hospitals are the full residual claimants on the ω-related savings they generate. Because there are no transfers to offset the ν_h, only 19.9% of hospitals select into bundled payments. These 19.9% generate non-trivial ω-related surplus gains, but because there are few of them, average total surplus only increases by $39 per episode (i.e. about one-fifth of the increase that would be achieved if all hospitals selected in with no transfers).

"Perfect targeting" is a useful benchmark, but not feasible in the context of our model. To see this, recall that we model $\lambda_{h3} = \lambda_h + g(t) + \epsilon_{h3}$, where ϵ_{h3} is an iid random variable that is not predictable in advance. To gauge the benefits of a more feasible contract, we consider a second scenario (labelled "feasible targeting") in which Medicare sets a target price of $t_h = \lambda_h + g(t)$, which is based on the hospital's underlying type and the time trend.[32] The correlation with realized claims under FFS is reduced to $\frac{\sigma_\lambda}{\sqrt{\sigma_\lambda^2 + \sigma_\epsilon^2}} = 0.93$, which is two-thirds of the way between the correlation for observed target prices ($\rho_{\lambda_3,t} = 0.81$) and perfect targeting ($\rho_{\lambda_3,t} = 1$). We estimate that relative to the observed targeting, feasible targeting generates approximately three-fifths of the social welfare gains that could be achieved with perfect targeting; as a result, under feasible targeting, social surplus is virtually identical to the FFS benchmark (-$5).

The above analysis asks how well observed targeting does compared to its potential; the mirror-image of this is to ask how well it does relative to no targeting. We therefore undertake a third exercise, labeled "no targeting," which considers a case in which 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 higher levels of participation in bundled payment, greater (inefficient) selection on levels, greater government spending and higher social costs. Relative to no targeting,

---

[31]The values of $\mu_{t,s}$ for the 8 strata are $[10.00, 10.04, 10.04, 10.11, 10.23, 10.05, 10.14, 10.26]$

[32]Given our stochastic assumptions, the mean of this contract is unchanged, and the standard deviation is reduced to $\sigma_t = \sqrt{\sigma_{\lambda,3}^2 - \sigma_\epsilon^2} = 0.17$, from $\sigma_t = \sigma_{h3} = 0.18$ under perfect targeting.

feasible targeting has the potential to generate $189 per episode in social surplus; the observed targeting in turn generates approximately two-thirds of these feasible gains ($128 = (-66)- (-194) out of $189).

While improved information is a natural way for Medicare to achieve better targeting, it is not the only way to do so. Medicare could also achieve better targeting through a narrower definition of the bundle. To illustrate, suppose that we can decompose average per-episode claims into two additive separable components: hospital claims and other claims, which includes post-acute care. Because hospitals were already paid a predetermined, DRG-based amount under FFS, we do not expect hospital claims to respond to the incentives from bundled payments. Indeed, the empirical evidence shows that most of the ω-related savings come from other claims, and post-acute claims in particular. However, while hospital claims are not a source of ω-related savings, they are heterogenous across hospitals, and thus a source of selection on levels. Eliminating hospital claims from the bundle thus effectively increases the degree of targeting.

We simulate the effects of two such narrower bundles. In the first, labelled "narrow bundling, no targeting", we remove hospital claims from the bundle but assume that we cannot target prices for other claims. Within our framework, this is equivalent to setting the target price as the sum of ex post, realized hospital claims and a predetermined, fixed payment for other costs.[33] This target price has correlation coefficient of $\rho_{\lambda_3,t} = 0.76$ and reduces social welfare relative to the observed target prices.

In the second, called "narrow bundling, observed targeting", we continue to assume that hospitals are paid their ex post realized claims, however, we now assume that Medicare can obtain the same degree of targeting for other claims as the actual target price achieved for total claims.[34] Doing so increases the correlation coefficient to $\rho_{\lambda_3,t} = 0.87$. This contract generates social gains relative to the observed voluntary bundled payment program and brings overall social welfare (-$39) closer to the level from feasible targeting.

Overall, our findings suggest that while the observed voluntary program is socially costly,

---

[33]That is, the target price is set to be $f_h^{HOSP} + \mathbb{E}_h[f_h^{OTH}]$. We calculate $\{\mu_t, \sigma_t^2, \rho_{\lambda_3,t}\}$ for this target price using the empirical distributions of $t_h = f_h^{HOSP} + \mathbb{E}_h[f_h^{OTH}]$ and λ_h.

[34]Specifically, the target price is set to $t_h = f_h^{HOSP} + t_h^{OTH}$, 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 to the correlation between the observed target price and λ_h.

there are feasible improvements in targeting which could create a voluntary program that would mostly eliminate social losses (or perhaps even generate small social gains). These could arise through tailoring target prices better to reflect claims under FFS of each hospital, or focusing the bundle on a narrower set of services in which cost-savings can be realized.

7 Conclusion

Government regulations are sometimes based on voluntary participation, allowing market actors to "choose their own incentives." These voluntary regimes may be socially beneficial if they induce selection of actors with private information about the net benefits from changing their behavior, but they may be socially costly if they primarily attract actors who can receive higher government payments with no behavior change. We explored this tradeoff between "selection on slopes" and "selection on levels" in the context of Medicare payment reform.

Our analysis takes advantage of a unique setting in which Medicare introduced an alternative payment model – called bundled payments – as a randomized trial and then modified the experimental design in midstream. Bundled payments was originally imposed as a mandatory participation model for hospitals in 67 randomly selected markets, with hospitals in 104 other randomly selected markets paid under the status quo. Two years into this five-year experiment, however, hospitals in half of the 67 randomly selected markets were unexpectedly allowed to choose whether to remain under bundled payments or revert back to the status quo payment model. This provided a rare opportunity to observe all hospitals' behavior under the alternative model and then to observe which hospitals voluntarily choose to continue under it.

Both the descriptive evidence and the model estimates indicate that selection was primarily based on levels rather than on slopes. The main driver of participation in the voluntary bundled payment model was whether the hospital would benefit financially from the alternative regime without any change in behavior. We also found selection on slopes – hospitals who changed their behavior more in response to bundled payments were also more likely to opt in – but selection on this margin was much less quantitatively important.

As a result, we estimated that the voluntary bundled payment model generated inefficient transfers to hospitals and reduced social welfare relative to imposing the status quo fee-for-service (FFS) payment regime on all hospitals. However, we also estimated that alternative (feasible) vol-

untary designs that targeted reimbursement more closely to hospitals' claims level under FFS could reduce these inefficient transfers substantially. Of course, any design with less generous reimbursements to "better-performing" actors may raise concerns about fairness, as well as concerns about ratchet effects (Freixas, Guesnerie and Tirole, 1985).

Our quantitative results are, of course, specific to our setting. The impact of alternative payment models in Medicare have tended to be modest – "singles not home runs" in the words of Frakt (2019*a*); in our setting, this limits the scope for social-welfare improving selection on slopes. In other contexts, this scope could be greater. In addition, the design feature that is so helpful for our empirical work – that we observe hospital behavior under the alternative regime and then their choices of whether to continue in it – may also affect the selection dynamics relative to a more typical setting in which actors are given a choice of opting into a new regime without any prior experience with that regime. For example, hospitals' prior experience under bundled payments may have increased hospitals' information about their levels and/or slopes, thus creating more selection on one or both of these dimensions. In addition, the expectation that they would have to participate in bundled payments for five years may have induced hospitals to undertake sunk investments, which may in turn have affected their participation decisions when it unexpectedly became optional.

Beyond the specific quantitative results, our analysis suggests the importance of considering – and ideally estimating – selection on both slopes and levels in settings in which the regulator is considering a voluntary regime. While there is an active and ongoing debate over the merits of voluntary versus mandatory payment reforms, voluntary participation is currently the norm across Medicare's alternative payment models. With the partial exception of the program we study, all other bundled payment models, Accountable Care Organizations, and primary care coordination models have been implemented in a voluntary manner. Moreover, as we noted in the introduction, voluntary regulation is widespread outside of healthcare, in sectors such as education, environmental regulation, and electricity regulation. Exploring the impact and optimal design of such voluntary programs in these other settings is an important and fruitful direction for further work.

References

Alexander, Diane. 2017. "How Do Doctors Respond to Incentives? Unintended Consequences of Paying Doctors to Reduce Costs." *Working Paper*. `https://www.chicagofed.org/publications/working-papers/2017/wp2017-09`.

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.

Cicala, Steve, David Hémous, and Morten Olsen. 2020. "Adverse Selection as a Policy Instrument: Unraveling Climate Change." *Working Paper*. `https://home.uchicago.edu/~scicala/papers/unraveling/unraveling_climate_change_draft.pdf`.

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. 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): 73273–73554.

CMS. 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): 49325–49886.

CMS. 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." *Federal Register*, 81(162): 56761–57345.

CMS. 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.

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*, 63(1): 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 A, Lindsey M Burke, and Patrick J Wolf. 2019. "The Effects of Regulations on Private School Choice Program Participation: Experimental Evidence from Florida." *Social Science Quarterly*, 100(6): 2316–2336.

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.

Dummit, Laura A, Kimberly Smathers, OJ Bright, Rebecca Cherry, McCulloch Cline, Amy Cowell, et al. 2018. "CMS Comprehensive Care for Joint Replacement Model: Performance Year 1 Evaluation Report." The Lewin Group.

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*a*. "'Value' of Care Was a Big Goal. How Did It Work Out?" *The New York Times*, September 23.

Frakt, Austin. 2019*b*. "Which Health Policies Actually Work? We Rarely Find Out." *The New York Times*, September 9.

Freixas, Xavier, Roger Guesnerie, and Jean Tirole. 1985. "Planning Under Incomplete Information and the Ratchet Effect." *The Review of Economic Studies*, 52(2): 173–191.

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*, 179(7): 924–931.

Heckman, James J, and Bo E Honore. 1990. "The Empirical Content of the Roy Mode." *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. 2020. "Selection on Welfare Gains and Policy Design: Experimental Evidence from Electricity Plan Choice." *Working Paper*.

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." *Modern Healthcare*, April 26.

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." *New England Journal of Medicine*, 378(18): 1663–1665.

Liao, Joshua M, Mark V Pauly, and Amol S Navathe. 2020. "When Should Medicare Mandate Participation In Alternative Payment Models?" *Health Affairs*, 39(2): 305–309.

Marone, Victoria R, and Adrienne Sabety. 2019. "Should There be Vertical Choice in Health Insurance Markets?" *Working Paper*. https://victoriamarone.github.io/jobmarket/Marone_JMP_Vertical_Choice.pdf.

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." *Working Paper*. `https://www.nber.org/papers/w22600`.

Thaler, Richard H, and Cass R Sunstein. 2003. "Libertarian Paternalism." *American Economic Review*, 93(2): 175–179.

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
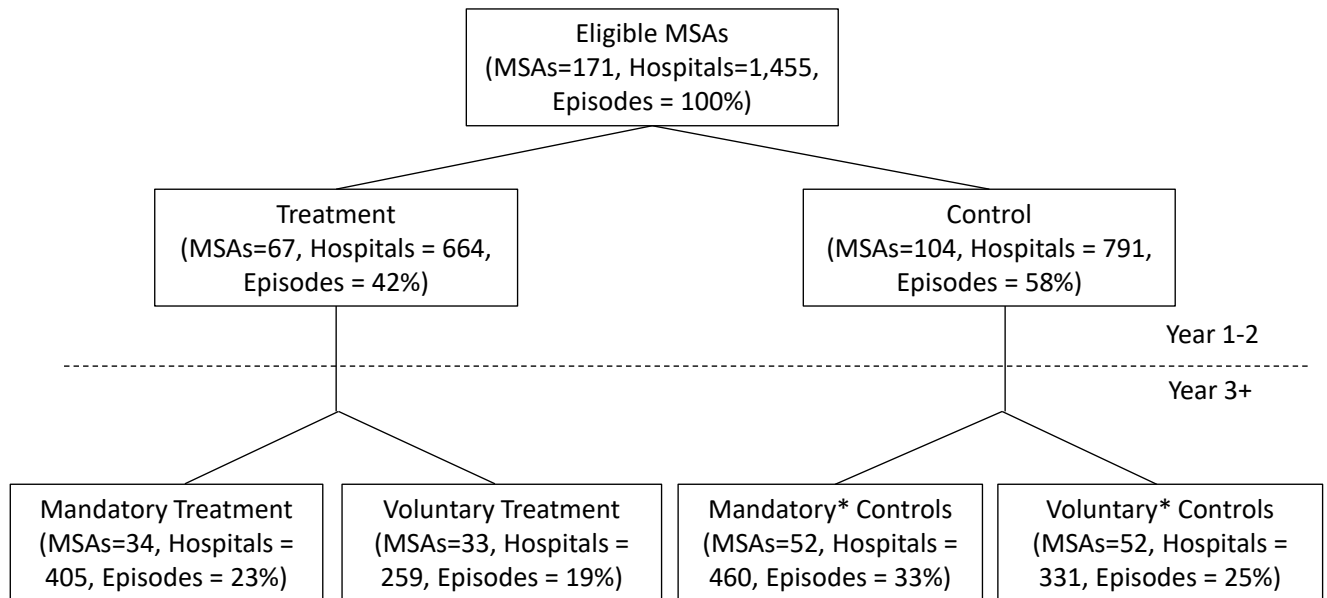
Note: Figure shows the design of the bundled payments experiment. The top half shows the initial mandatory design in program years 1-2, and the bottom half shows the partially voluntary design in program years 3-5. Episode shares are based on data from program years 1-2.

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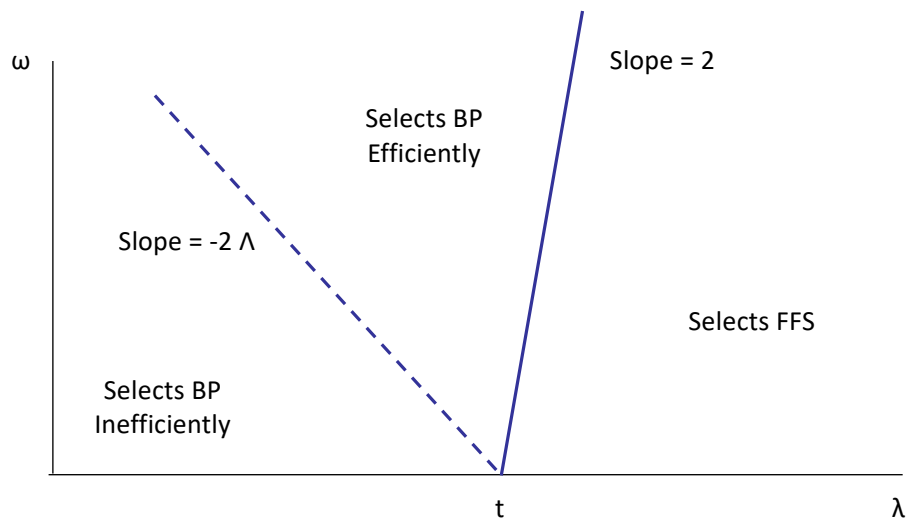
Note: Figure shows, for a given target price t, the hospital participation decision and social welfare implications as a function of the hospital's λ and ω.

Figure 3: Model Illustration

(a) Impact of Raising Target Prices

(b) Larger ω, More Variable λ

(c) Smaller ω, More Variable λ

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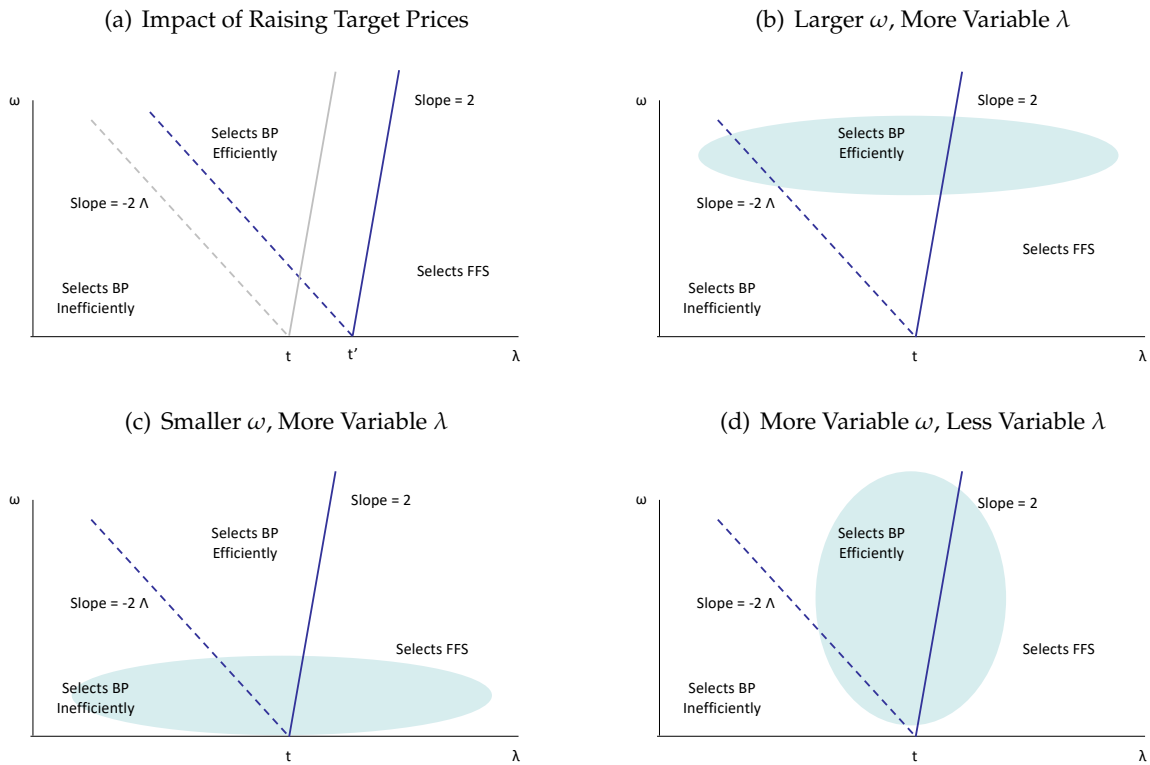
Note: Figure illustrates some of the key analytics in voluntary bundled payment design. Panel A illustrates the tradeoffs involved in setting higher target payments $t' > t$; Panels B through D consider the impact of different primitives and targeting, with Panel B vs. C comparing outcomes with higher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

Figure 4: Model Estimates

(a) No Target Price or Choice Shifter

(b) Target Price, Ignoring Choice Shifter

(c) Target Price and Choice Shifter

Note: Figure reports the empirical analogues of Figures 2 and 3. Specifically, it reports simulated hospitals based on our estimates for the 259 hospitals in the voluntary treatment group (see Figure 1), with circles proportional to the number of episodes. Panel (a) reports results assuming there are no target prices and the choice shifter ν_h in the hospital selection equation is not decision-relevant. Panel (b) considers the role of target prices by plotting $\lambda_{h3} - t_h - \bar{t}$ on the horizontal axis. In addition to adding t_h, we subtract the average target price $\bar{t}$, so that the axis remains on the same scale as in Panel (a). In Panel (c), we not only net out target prices but also allow the choice shifter ν_h to be decision-relevant.

Figure 5: Medicare Costs and Social Surplus Under Alternative Target Prices
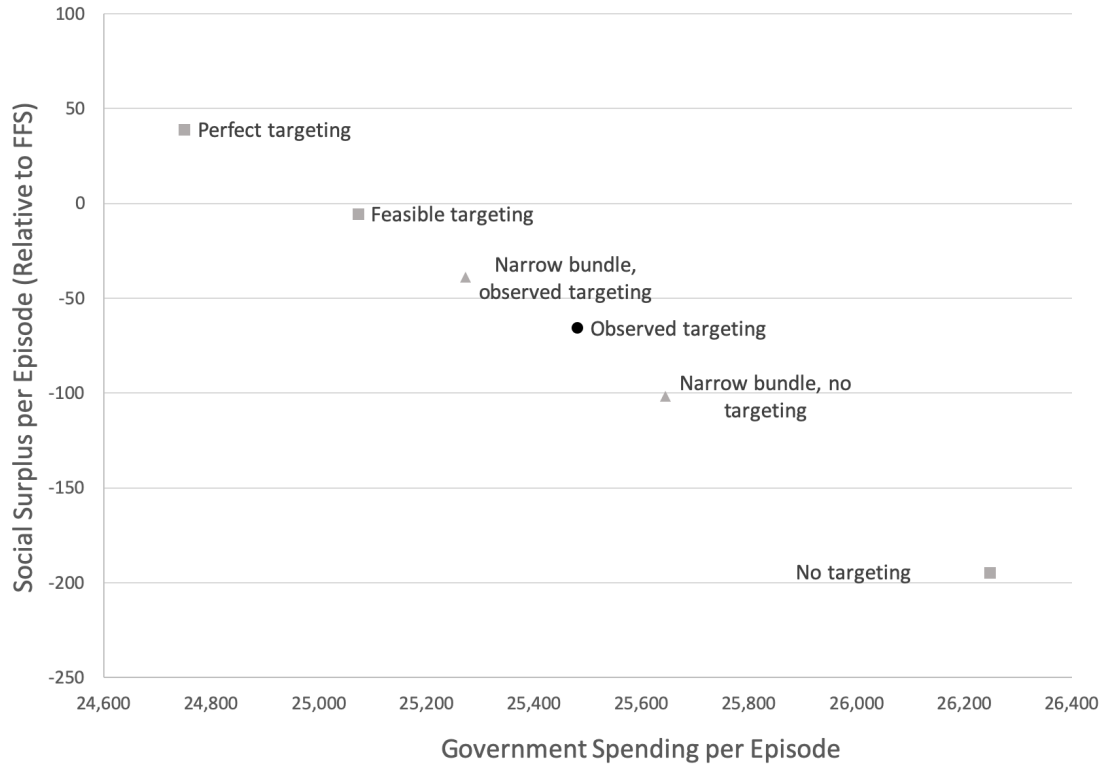
Note: Figure plots social surplus per episode (relative to the FFS counterfactual) against government spending, and correspond to the values in columns (5) and (2) of Table 4. See that table note for more details.

Table 1: Experimental Estimates

	Control Mean	Control SD	Average Treatment Effect	Standard Error	P-value
Panel A: Claims, Utilization, and Gov't Spending (per Episode)					
Claims	25,304	3,601	-797	204	0.001
Claims for Index Admission	13,546	2,387	-175	89	0.06
Claims for Institutional PAC	4,122	1,380	-498	128	0.001
Claims for Home Health	1,801	918	-89	59	0.14
Other Claims	5,835	531	26	55	0.64
Utilization					
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
Government Spending	25,304	3,601	33	208	0.88
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. "Claims" include patient cost-sharing. 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
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%	
Panel A: Selection on Levels (Period 1 Outcomes per Episode)				
Claims	26,523	26,146	27,777	0.03
	(5,368)	(4,176)	(5,498)	
Claims for Institutional Post Acute Care	4,812	4,681	5,552	0.02
	(2,403)	(2,054)	(2,448)	
Share Discharged to Institutional Post Acute Care	36.9%	37.3%	41.6%	0.10
	(16.0%)	(16.2%)	(14.5%)	
Panel B: Selection on Slopes				
Impact on episode claims		-796	-669	0.73
		(1,932)	(2,831)	
Impact on Institutional PAC claims		-519	-177	0.05
		(973)	(1,474)	
Impact on Share Discharged to Institutional PAC		-3.3%	-1.2%	0.13
		(7.8%)	(9.2%)	
Panel C: Selection on Hospital Characteristics				
Mean Number of CJR Episodes	306	320	251	0.14
Above Median Beds	60.6%	62.0%	60.1%	0.82
Teaching	18.7%	4.3%	16.1%	0.02
For-profit	17.5%	26.6%	12.5%	0.06
Non-profit	78.5%	63.4%	68.6%	0.57
Government-owned	4.0%	10.0%	18.9%	0.20
Share with Above Median Quality	60.0%	79.2%	59.3%	0.01
Northeast	8.4%	5.3%	0.6%	0.08
Midwest	33.4%	33.1%	52.2%	0.03
South	22.4%	14.5%	22.4%	0.31
West	35.3%	47.1%	24.8%	0.02

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 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-owned" 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 6). 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	0.018					
σ_λ	0.167	0.004					
γ	-0.067	0.004					
σ_ε	0.066	0.005					
μ_ω	4.833	0.708					
σ_ω	1.500	0.253					
$\rho_{\lambda,\omega}$	0.300	0.072					
v	-15,000	13,667					
κ	0.400	0.101					
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	386	1,124	11	46	125	344	1,476
B.2. Hospitals in the Voluntary Treatment Group Only							
λ_h	25,978	5,412	18,425	21,987	25,376	29,079	35,756
λ_{h3}	24,443	5,652	16,805	20,200	23,789	27,605	34,670
$t_h - \lambda_{h3}$	-85	4,895	-8,238	-2,952	185	2,711	7,562
ω_h	403	1,088	11	47	128	356	1,482
$(t_h - \lambda_{h3}) + \omega/2$	116	4,847	-7,946	-2,735	388	2,871	7,672
v/n_h^K	-1,419	513	-2,368	-1,573	-1,297	-1,124	-845

Note: Panel A reports parameter estimates from the model and Panel B.1 reports some key summary statistics implied by these estimates for all 1,455 hospitals. Standard errors are calculated using the bootstrap method, based on 20 samples of hospitals drawn with replacement from cells defined by the full interaction of treatment assignment and strata. Panel B.2 reports these implied statistics separately for the 259 hospitals in the voluntary treatment group (see Figure 1). The model is estimated weighting each hospital by the number of episodes. The summary statistics in panels B.1 and B.2 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	Government 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%	24,443	0	0	0	0	0
Mandatory Bundled Payment	100.0%	24,358	-98	116	214	-1,302	-1,204
Voluntary Bundled Payment	38.8%	26,125	1,933	1,754	-179	1,248	-685
Panel B: Alternative Voluntary Regimes with Different Target Prices							
Perfect targeting	19.9%	24,749	352	391	39	156	-196
Feasible targeting	29.4%	25,073	723	718	-5	357	-366
Observed targeting	35.8%	25,480	1,191	1,126	-66	670	-521
No targeting	41.4%	26,247	2,074	1,880	-194	1,351	-723
Narrow bundle, no targeting	38.6%	25,644	1,381	1,279	-102	796	-585
Narrow bundle, observed targeting	33.6%	25,273	953	914	-39	497	-456

Note: All counterfactuals are conducted on the 259 hospitals in the voluntary treatment group (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 (G from Equation 5). 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_{Cs} be the subset of control group hospitals in stratum 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 $H_M := H_T \setminus H_V$ 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 that 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 $W_{Cs} = \sum_{h \in H_{Cs}} w_h$ be the sum of hospital weights in stratum 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

$$\ln L = \sum_{h \in H_c} w_h \ln f(y_{h1}, y_{h2} | \theta_1) \tag{7}$$

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:

$$\widehat{\gamma} = \frac{1}{W_C} \sum_{h \in H_C} w_h \left[\ln y_{h2} - \ln y_{h1} \right] \tag{8}$$

$$\widehat{\mu_{\lambda,s}} = \frac{1}{2W_{Cs}} \sum_{h \in H_{Cs}} w_h \left[\ln y_{h1} + \widehat{\gamma} + \ln y_{h2} \right] \quad \text{for} \quad s = 1 \dots 8 \tag{9}$$

$$\widehat{\sigma_\lambda} = \sqrt{\frac{1}{W_C} \sum_{h \in H_C} w_h (\ln y_{h1} - \widehat{\mu_{\lambda,s}} + \widehat{\gamma})(\ln y_{h2} - \widehat{\mu_{\lambda,s}})} \tag{10}$$

$$\widehat{\sigma_\epsilon} = \sqrt{\frac{1}{2W_C} \left[\sum_{h \in H_C} w_h \left[(\ln y_{h1} - \widehat{\mu_{\lambda,s}} + \widehat{\gamma})^2 + (\ln y_{h2} - \widehat{\mu_{\lambda,s}})^2 \right] \right] - \widehat{\sigma_\lambda}^2} \tag{11}$$

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

$$\ln L = \sum_{h \in H_V} w_h \ln g_1(BP_{h3}, y_{h2} | \widehat{\theta}_1, \theta_2, y_{h1}) + \sum_{h \in H_M} w_h \ln g_2(y_{h2} | \widehat{\theta}_1, \theta_2, y_{h1}) \tag{12}$$

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

$$\ln g_1(BP_{h3}, y_{h2} | \widehat{\theta}_1, \theta_2, y_{h1}) = \ln \Pr(BP = BP_{h3} | \widehat{\theta}_1, \theta_2, y_{h1}, y_{h2}) + \ln g_2(y_{h2} | \widehat{\theta}_1, \theta_2, y_{h1}), \tag{13}$$

so that

$$\ln L = \sum_{h \in H_V} w_h \ln \Pr(BP = BP_{h3} | \widehat{\theta}_1, \theta_2, y_{h1}, y_{h2}) + \sum_{h \in H_T} w_h \ln g_2(y_{h2} | \widehat{\theta}_1, \theta_2, y_{h1}). \tag{14}$$

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 50,000 values for each set of candidate parameter values, and estimate the marginal density function g_2 using a Epanechnikov kernel with a bandwidth of 400. 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 $W_{Ts} = \sum_{h \in H_{Ts}} w_h$ 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:

$$\widehat{\mu_{ts}} = \frac{1}{W_{Ts}} \sum_{h \in H_{Ts}} w_h \ln t_{h2} \quad \text{for} \quad s = 1 \dots 8 \tag{15}$$

$$\widehat{\sigma_t} = \sqrt{\frac{1}{W_T} \sum_{h \in H_T} w_h \left(\ln t_{h2} - \widehat{\mu_{ts}}\right)^2}. \tag{16}$$

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

$$Cov(\ln y_{h1}, \ln t_{h2}) = Cov(\ln \lambda_h - \gamma + \epsilon_{h1}, \ln t_{h2}) = Cov(\ln \lambda_h, \ln t_{h2}) = \rho_{\lambda t} \sigma_\lambda \sigma_t \tag{17}$$

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

$$\widehat{\rho_{\lambda t}} = \frac{1}{\widehat{\sigma_\lambda} \widehat{\sigma_t}} \frac{1}{W_T} \sum_{h \in H_T} w_h (\ln y_{h1} - \widehat{\mu_{\lambda s}} + \widehat{\gamma})(\ln t_{h2} - \widehat{\mu_{ts}}). \tag{18}$$

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 $\widehat{\rho_{\lambda_3 t}} = \dfrac{\widehat{\sigma_\lambda}}{\sqrt{\widehat{\sigma_\lambda}^2 + \widehat{\sigma_\epsilon}^2}} \widehat{\rho_{\lambda t}}$.

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 $\{\mu_t^j, \sigma^j, \rho_{\lambda_3 t}^j\}$ 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

$$t_h^j = \mu_t^j + \frac{\rho_{\lambda_3 t}^j \sigma_t^j}{\sigma_{\lambda 3}} (\lambda_{h3} - \mu_\lambda) + \sqrt{1 - (\rho_{\lambda_3 t}^j)^2} \, \sigma_t^j \, \epsilon_h \tag{19}$$

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.

C Stop-Loss and Stop-Gain

As we discussed in Section 2, CJR was not implemented as a "pure" bundled payment program. Instead, like most bundled payment programs, it limited risk exposure to both hospitals and Medicare. In particular, the reconciliation payments for the difference between target prices and realized FFS claims were capped on either side by stop-gain and stop-loss amounts. These amounts increased from a stop-gain of 5 percent of t_h and a stop-loss of zero (meaning that hospitals would never need to make payments to Medicare) in year 1, to stop-gain and stop-loss amounts of ± 20 percent of t_h in years 4 and 5 of the program.

Qualitatively, we do not see these provisions as affecting the basic economics of the environment. For instance, while some hospitals receive smaller reconciliation payments when the stop-gain binds, because this provision targets the tail of the distribution, the affected hospitals are unlikely to be marginal in their selection decision. The same logic holds for the cap on losses.

To quantitively assess the importance of the stop-gain and stop-loss provisions, we estimate an alternative version of the model that caps reconciliation payments at ± 20 percent and we redo our counterfactuals with these estimates. In testing, we found that we required a larger number of draws for reliable estimation of the model, and this was not feasible on the computer server with the research identifiable claims data. To estimate this alternative model, we therefore removed hospital-level data for the 1,326 of 1,455 hospitals where we had adequate sample size to preserve anonymity (11 or more episodes in periods 1 and 2) and estimated the model on a more powerful computer server. The hospitals we dropped are those with the smallest number of episodes, and we confirmed that the estimates of the baseline model (without stop-gain or stop-loss) were very similar when they were excluded. Appendix Table A3 reproduces Table 4 using the stop-loss and stop-gain version of the model. While the exact numbers are slightly different, the qualitative patterns in the tables are identical and the quantitive estimates are very similar.

Figure A1: Selection on Hospital Characteristics

(a) Pre-Period CJR Episodes

(b) Composite Quality Score
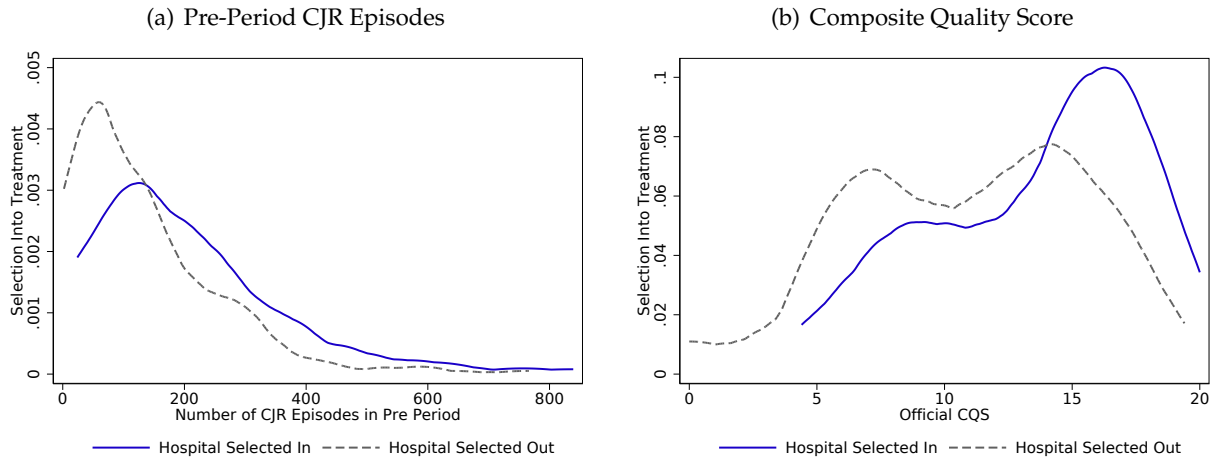
Note: Figure shows kernel densities of hospital characteristics by whether the hospital selected in or out of the CJR program in period 3.

Table A1: Experimental Estimates: Alternative Specifications

	MSA-Level, Unweighted		MSA-Level, Weighted		Hospital-Level, Weighted	
	Control Mean	Average Treatment Effect	Control Mean	Average Treatment Effect	Control Mean	Average Treatment Effect
	(SD)	(SE) *P-Value*	(SD)	(SE) *P-Value*	(SD)	(SE) *P-Value*
Panel A: Healthcare Use and Spending						
Claims	25,304	-797	25,899	-593	25,899	-407
	(3,601)	(204)	(3,344)	(213)	(5,142)	(286)
		0.001		*0.007*		*0.158*
Claims for Index Admission	13,546	-175	13,754	-117	13,754	-47
	(2,387)	(89)	(2,079)	(91)	(2,670)	(133)
		0.06		*0.20*		*0.73*
Claims for Institutional PAC	4,122	-498	4,117	-333	4,117	-323
	(1,380)	(128)	(1,220)	(121)	(2,411)	(123)
		0.001		*0.007*		*0.010*
Claims for Home Health	1,801	-89	1,997	17	1,997	26
	(918)	(59)	(872)	(87)	(996)	(89)
		0.14		*0.85*		*0.78*
Other Claims	5,835	26	6,032	10	6,032	89
	(531)	(55)	(538)	(44)	(1049)	(86)
		0.64		*0.82*		*0.30*
Utilization Measures						
Number of Days in Index Admission	2.6	-0.1	2.6	-0.01	2.6	0.0
	(0.4)	(0.04)	(0.3)	(0.04)	(0.6)	(0.04)
		0.21		*0.73*		*0.87*
Number of Days in Institutional PAC	7.7	-0.6	7.6	-0.3	7.6	-0.4
	(2.3)	(0.23)	(1.9)	(0.21)	(4.2)	(0.21)
		0.02		*0.14*		*0.08*
Discharge Destination						
Institutional Post Acute Care	0.313	-0.034	0.320	-0.021	0.320	-0.021
	(0.104)	(0.009)	(0.093)	(0.009)	(0.151)	(0.009)
		0.001		*0.020*		*0.020*
Home Health Agency	0.339	0.004	0.378	0.025	0.378	0.026
	(0.196)	(0.018)	(0.173)	(0.019)	(0.222)	(0.018)
		0.81		*0.19*		*0.16*
Home (w/o Home Health Agency)	0.329	0.042	0.289	0.013	0.289	0.005
	(0.232)	(0.018)	(0.212)	(0.018)	(0.265)	(0.020)
		0.02		*0.45*		*0.82*
Other	0.020	-0.004	0.013	-0.004	0.013	-0.004
	(0.032)	(0.018)	(0.023)	(0.002)	(0.030)	(0.002)
		0.05		*0.08*		*0.05*
Panel B: Quality Measures						
Complication Rate	0.011	0.001	0.011	0.001	0.011	0.002
	(0.005)	(0.001)	(0.004)	(0.0005)	(0.010)	(0.0005)
		0.26		*0.02*		*0.01*
ER Visit During Episode	0.199	0.003	0.194	0.002	0.194	0.002
	(0.027)	(0.003)	(0.021)	(0.002)	(0.047)	(0.003)
		0.41		*0.42*		*0.51*
90-day All Cause Readmission Rate	0.102	-0.001	0.103	-0.002	0.103	-0.001
	(0.015)	(0.002)	(0.004)	(0.002)	(0.031)	(0.002)
		0.69		*0.22*		*0.40*
Panel C: Admissions and Patient Composition						
LEJR Admissions (per 1,000 enrollees)	29.9	-0.8	27.3	-0.5	7.9	-0.5
	(15.8)	(0.5)	(13.6)	(0.3)	(12.4)	(0.2)
		0.10		*0.16*		*0.02*
CJR-eligible Admissions (per 1,000 enrollees)	23.6	0.1	21.1	0.1	6.7	-0.4
	(11.3)	(0.5)	(9.7)	(0.5)	(9.5)	(0.2)
		0.89		*0.87*		*0.02*
Elixhauser Comorbidity Score	2.4	-0.001	2.3	-0.006	2.3	-0.0001
	(0.3)	(0.029)	(0.2)	(0.028)	(0.4)	(0.027)
		0.98		*0.85*		*1*

Note: Table compares regression estimates in Table 1 with estimates from two alternative specifications: an MSA-level specification, where observations are weighted based on number of episodes in the MSA in period 2, and a hospital-level specification where observations are weighted based on the number of episodes at the hospital in period 2. See Table 1 note for more details.

Table A2: Selection on Slopes with Hospital-Specific Trends

Specification:	Baseline			With Hospital-Specific Trend		
	Voluntary Select-In	Voluntary Select-Out	P-Value of Select-In vs. Select-Out Difference	Voluntary Select-In	Voluntary Select-Out	P-Value of Select-In vs. Select-Out Difference
Impact on episode claims	-796	-669	0.73	-362	385	0.42
	(1,932)	(2,831)		(4,335)	(10,018)	
Impact on Institutional PAC claims	-519	-177	0.05	-305	46	0.49
	(973)	(1,474)		(2,896)	(4,876)	
Impact on Share Discharged to Institutional PAC	-3.3%	-1.2%	0.13	-3.0%	-0.7%	0.46
	(7.8%)	(9.2%)		(16.0%)	(22.5%)	

Note: The panel titled "Baseline" replicates Table 2 Panel B from the paper, and reports the average (and standard deviation) over different hospitals of $b_{1,h}$ from estimating equation (2) by OLS. The panel titled "With Hospital-Specific Trend" reports the analogous results from estimating and augmented version of equation (2), which additionally controls for hospital-specific linear time trends. Specifically, we report estimates of $\beta_{4,h}$ from the regression $y_h = \beta_0 + \beta_{1,h} + \beta_{2,t} + \beta_{3,h} \times t + \beta_{4,h} \times BP_h + \epsilon_{h,t}$, where $\beta_{1,h}$ are hospital fixed effects, $\beta_{2,t}$ are calendar year fixed effects, $\beta_{3,h} \times t$ are hospital fixed effects interacted with a linear time trend, and $\beta_{4,h} \times BP_h$ are hospital fixed effects interacted with an indicator for treatment hospitals in period 2 (when bundled payment was in effect); we estimate this equation using data from 2013, 2014, 2016 and 2017. All estimates are weighted by the number of episodes in the hospital in period 2. The p-values in columns (3) and (6) are based on robust standard errors.

Table A3: Counterfactuals from Alternative Specification with Stop-Loss and Stop-Gain Provisions

	Share Selecting In	Government 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%	24,370	0	0	0	0	0
Mandatory Bundled Payment	100.0%	24,358	-13	117	130	-1,322	-1,308
Voluntary Bundled Payment	38.9%	26,057	1,941	1,728	-213	1,236	-704
Panel B: Alternative Voluntary Regimes with Different Target Prices							
Perfect targeting	19.8%	24,673	349	358	9	151	-198
Feasible targeting	28.8%	24,978	700	663	-37	343	-357
Observed targeting	35.0%	25,366	1,146	1,050	-96	638	-509
No targeting	41.1%	26,163	2,062	1,834	-229	1,335	-728
Narrow bundle, no targeting	38.1%	25,550	1,357	1,223	-134	779	-578
Narrow bundle, observed targeting	33.0%	25,180	932	860	-72	485	-447

Note: Table reports analogue outcomes to Table 4 based on an alternative specification of the model with stop-loss and stop-gain provisions for reconciliation payments. See Appendix Section C for more details.