

VOLUNTARY REGULATION: EVIDENCE FROM MEDICARE PAYMENT REFORM*

LIRAN EINAV

AMY FINKELSTEIN †

YUNAN JI

NEALE MAHONEY

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; however, after two years, participation was made voluntary for half of these hospitals. Using detailed claim-level data, we document that voluntary participation is more likely for hospitals that 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 stylized model of responsiveness to and selection into the program. We find that the current voluntary regime generates inefficient transfers to hospitals, and that alternative (feasible) designs could reduce these inefficient transfers and raise welfare. Our analysis highlights key design elements to consider under voluntary regulation. *JEL* Codes: H51, I13, I18.

*We thank Kate Ho, Parag Pathak, Jonathan Skinner, four anonymous referees, Stefanie Stantcheva (the Editor), and participants in many seminars for helpful comments. We are grateful to Yining Mo, Xuyang Xia, and Chuan Yu for outstanding research assistance. We gratefully acknowledge support from J-PAL North America’s Health Care Delivery Initiative (Finkelstein and Mahoney), the National Institute of Aging grant P01AG019783-15, the Laura and John Arnold Foundation (Einav, Finkelstein and Mahoney), the Becker Friedman Institute at the University of Chicago (Mahoney), and the National Science Foundation SES-1730466 (Mahoney).

†The author for correspondence. Mailing Address: 50 Memorial Drive, E52, Room 442, Cambridge MA 02142. Phone: (617) 253-4149. Email: afink@mit.edu.

I. 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.

Conceptually, this idea is not new. Indeed, it is a core tradeoff highlighted in theoretical analyses of optimal contract design and government regulation, generally (e.g., Laffont and Tirole, 1993) and in the health care sector (e.g., Newhouse, 1996, 2004). This paper’s key contribution is to leverage a unique, mid-course reform that changed a mandatory-participation, randomized controlled trial of a particular incentive scheme into a voluntary participation program; this provides a rare opportunity to estimate “slope” effects for both those who choose to participate and those who do not. This allows us to estimate empirically the core components of the canonical model and to illustrate – via our particular context and estimates – how the framework can be used for applied analysis.

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 few exceptions, provider participation in these payment models has been voluntary (GAO, 2018). There is an ongoing and active policy debate over whether

these programs should be made mandatory, but this has focused primarily on concerns about selection on levels, and has ignored the potential salutary benefits from selection on slopes (e.g., Gronniger et al., 2017; King, 2019; Frakt, 2019b; Levy et al., 2018; Liao et al., 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 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 bundle 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 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.

We begin by providing descriptive evidence on the mandatory and voluntary regimes. In the mandatory regime, we closely follow prior analyses and find that, consistent with them, 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; The Lewin Group, 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 also explore and document heterogeneity across hospitals in the impact of payment reform, as well as 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 remain in the program. Consistent with selection on slopes, we also find that hospitals that achieved larger reductions in claims under mandatory bundled payments are less likely to opt out 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 bundle 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 \$25,000 and that bundled payments reduced claims, on average, by about \$250 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 \$5,000. Observed bundle 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 alternative counterfactual 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; The Lewin Group, 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 bundle 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 descriptive work, indicate substantial selection on levels: hospitals that opt

into the voluntary bundled payment regime would have had (counterfactual) average episode claims under the FFS status quo that were substantially lower (relative to the bundle price) than hospitals which do not opt in. This selection on levels is offset by a small amount of favorable selection on slopes and social savings from a “too low” bundled price. On net, we estimate that the voluntary bundled payments regime modestly raises social surplus relative to the FFS status quo.

We also compare the surplus from the observed bundle prices to counterfactuals with better and worse targeting. Within the range defined by no targeting at the bottom and the best feasible targeting at the top, we find that the observed bundled prices generate approximately two-thirds of the feasible gains. We also discuss how defining the bundle more narrowly could effectively allow for better targeting of the bundle price.

While the mid-stream regulatory change to the program we study is unique from a research perspective, the regulator’s problem of whether or not to allow for voluntary participation is ubiquitous. Our empirical estimates are naturally specific to our setting, but the payment reform we analyze is quite similar in nature to a host of alternative payment models that Medicare has been introducing over the last decade (CMS, 2020), and is continuing to introduce through a mixture of voluntary and mandatory models.¹ Incentive programs with voluntary participation are also widespread 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 et al., 2020), by private schools into whether or not to accept public vouchers (DeAngelis et al., 2019), and by residential electricity consumers into whether to face a constant or time-varying regulated electricity price schedule (Ida et al., 2020).

Our paper relates to several distinct literatures. In addition to our conceptual debt to the theoretical literature on optimal regulation discussed above, our empirical analysis 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). Within health economics, our work contributes to the growing literature on the impact and optimal design of financial incentives for healthcare providers (e.g., Cutler, 1995; Gaynor et al., 2004; Clemens and Gottlieb, 2014; Ho and Pakes, 2014; Einav et al., 2018; Eliason et al., 2018; Gaynor et al., 2020). It also relates to work on so-called “selection on moral hazard” – i.e., consumer selection of health insurance plans

¹For instance, the announced End-Stage Renal Disease (ESRD) Treatments Model is mandatory, while the Kidney Care Choices (KCC) Model is voluntary. See the CMS Innovation Center website for more on these and other models: <https://innovation.cms.gov/innovation-models/kidney-care-choices-kcc-model>.

based not only on levels but on slopes (Einav et al., 2013, 2016; Shepard, 2020; Marone and Sabety, 2021); here, we examine selection on moral hazard from the provider side rather than the consumer side.

Most narrowly, the paper contributes to the literature on the impact of Medicare bundled payment programs. This includes several recent evaluations of the program we study focusing on random assignment under mandatory participation (e.g., Finkelstein et al., 2018; The Lewin Group, 2018; Barnett et al., 2019; Haas et al., 2019; Einav et al., 2020a; Wilcock et al., 2021). 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.² It is well understood that non-random selection into voluntary models can bias the evaluation of these programs (e.g., Gronniger et al., 2017; Levy et al., 2018). Our focus here is on proposing (and applying) a framework that allows us to quantify the impact of such programs, accounting for this selection.

The rest of the paper proceeds as follows. Section II provides background on our setting. Section III describes the data and presents 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 IV presents a stylized model of selection into a voluntary bundled payment program. Section V presents the econometric specification of the model and describes its identification and estimation. Section VI presents our main results. The last section concludes. All appendix material can be found in the Online Appendix.

II. SETTING

II.A. 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

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

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.³

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 costs in the local medical market. The contracts may also be structured to limit risk exposure for the hospital.

Proponents 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 low-cost patients (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 et al., 1997; Newcomer et al., 2014; Doran and Zabinski, 2015; Froemke et al., 2015; Dummit et al., 2016; Navathe et al., 2017; Carroll et al., 2018). However, voluntary participation makes separating treatment from selection difficult, and the small number of participating hospitals raises concerns about generalizability (Gronniger et al., 2017; King, 2019).

³One exception to this system is hospital reimbursements. Starting in 1982, Medicare adopted the Prospective Payment System in which it makes a fixed payment for the hospital stay based on the patient's diagnosis.

II.B. 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 to 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 *bundle 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 bundle price are key design elements in a bundled payment program. Let b_h denote the average per episode bundle 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 b_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 $b_h - y_h$ per episode, so that the gross Medicare payment is b_h .

Under CJR, Medicare tried to set the bundle price before each program year to be slightly lower than expected per-episode claims under FFS.⁴ To do so, the bundle 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 years. The "discount factor" was designed to reflect Medicare's portion of expected savings from CJR.

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. However, in practice the quality standard was not

⁴In particular, Medicare set hospital-specific bundle 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.

binding for the vast majority of the hospitals.⁵ 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 to both parties, the reconciliation payment was subject to stop-loss and stop-gain provisions. Specifically, if $b_h - y_h$ was less than (the negative of) the stop-loss amount, the hospital “only” has to pay Medicare the stop-loss amount. Similarly, if $b_h - y_h$ is greater than the stop-gain amount, Medicare “only” has to pay the hospital the stop-gain amount.⁶ 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 Einav et al., 2020b), so we abstract from them throughout.

II.C. 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, 2015b). 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.⁷ 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 III for more details).

Participation was mandatory in treatment MSAs: eligible hospitals had no choice but to be reimbursed

⁵For 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. See <https://innovation.cms.gov/Files/x/cjr-qual-sup.pdf> for more details on the quality standard.

⁶The stop-loss and stop-gain amounts increased over time. In the first year, the stop-gain amount was set as 5% of b_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 b_h .

⁷After 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.

under the new bundled payment model. This mandatory participation feature was immediately controversial, with then-US representative Tom Price spearheading a letter in September 2016, signed by 179 members of Congress, complaining that mandatory participation was unethical and unauthorized (The Hill, 2016). Subsequently, as the new Secretary of Health and Human Services – the federal agency charged with overseeing Medicare – Price led the effort to roll back mandatory participation bundled payment models. As a result, in a rule finalized in December 2017, Medicare 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 with the lowest episode claims under FFS (using 2011-2014 data, before random assignment was announced). 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. About a quarter of the hospitals in the voluntary bundled payment MSAs (73 out of 279) chose to remain under bundled payments, which is typical of participation rates in other voluntary Medicare payment programs, where rates have ranged from about 15% to one third (The Lewin Group, 2015). 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. Period 3 is defined as the final 3 years of the program when the program was voluntary for some hospitals.⁸

Figure I shows a flow chart of the experimental design. The middle 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

⁸At the start of period 3, Medicare also began to cover (in all hospitals) outpatient knee replacement as well as inpatient knee replacement. Since we find no statistically or quantitatively significant impact of treatment assignment on period 3 knee replacement volume or setting (not reported), we abstract from it in what follows.

analysis, we will compare mandatory treatment to mandatory control, and voluntary treatment to voluntary control.

III. DATA AND DESCRIPTIVE EVIDENCE

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

III.A. Data and Sample

Our main data are the 100% Medicare enrollment and claims files from 2013 to 2018. These contain basic demographic information (age, race, sex, and Medicaid enrollment) and claims for inpatient, outpatient, and post-acute care. 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 bundle price, annual reconciliation payment, and whether the hospital opted into bundled payments when it became voluntary in 2018 (CMS Innovation Center). When participation is voluntary (in 2018), we only observe bundle prices for hospitals that select in or are mandated to take part in the bundle payments. In Appendix A we describe how we impute bundle prices for the hospitals that do not select into the program.⁹ 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. Episodes were excluded if

⁹Among hospitals where we observe prices, the correlation between the imputed prices and the observed prices is 0.98. Because of the strength of this correlation, for the rest of the paper we abstract from any potential measurement error associated with the imputation procedure.

the patient did not have Medicare as the primary payer, was readmitted during the episode for LEJR, or died during the episode. Finkelstein et al. (2018) provide more detail on these eligibility criteria.

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 episodes would end by December 31, 2017, the close of the second year. 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). Similarly, we define period 3 to include all episodes admitted between January 1, 2018 and September 15, 2018. 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, 2015a).

To construct our baseline sample, we start with the universe of 1,584 hospitals in the 171 treatment and control MSAs that had a CJR episode during period 2; these hospitals had a total of 396,643 CJR episodes during period 2. So that we can observe outcomes in all three periods, we restrict the sample to the 1,416 hospitals that have at least one CJR episode in each period. These 1,416 hospitals constitute our baseline sample, out of which 647 hospitals are located in a treatment MSA and 769 hospitals are in a control MSA. A total of 379,150 CJR episodes in period 2 fall into these 1,416 hospitals.

III.B. Average Treatment Effects

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

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

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.

Table I 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 I 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 II.B, 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 patients’ discharge destination following their index admission. In the control group, patients are discharged to institutional PAC, 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. This can be either because the patients who would have been sent to

institutional PAC are being sent home *without* home health, or because there is a cascading effect where the patients who would have been sent to institutional PAC are being sent home *with* home health, and patients who would have been sent home with home health are now being sent home without home health support. A cascading effect seems more likely (to us), but we cannot differentiate between these two mechanisms. These experimental estimates are consistent with qualitative evidence on how hospitals respond to CJR. In a survey of hospital executives and administrators, 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 bundle price, b_h , under bundled payments) does not change. The point estimate indicates a statistically insignificant increase of \$33 (standard error of \$208) in average government spending per episode. The lack of a reduction in government spending – despite modest reductions in submitted claims – reflects the design feature of bundled payments, according to which bundle 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 I 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. Of course, in interpreting this evidence, it is important to bear in mind that the quality measures available in claims data are limited. We cannot, for example, measure outcomes such as morbidity, mobility, or activities of daily living in our data. However, surveys of patients at the end of their episode of care show similar improvements in functional status and pain for patients in treatment and control groups (The Lewin Group, 2020), which is consistent with clinical trial evidence that patient mobility following knee surgery is not improved by inpatient (as opposed to outpatient) rehabilitation

(Buhagiar et al., 2017). Given this lack of 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.

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. We examine patient composition by estimating effects on the Elixhauser Comorbidity Score of the patient pool, which is constructed as the sum of indicators for 31 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. In principle, 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 (2020) documents that physicians strategically direct patients across hospitals within a local area to maximize revenue. We therefore looked at whether hospitals that opted to remain under the bundled payment model experienced changes in their patient volume or patient composition relative to hospitals within the same market that opted out of the bundled payment regime. We might expect physicians to direct more complex patients towards hospitals that are no longer under bundled payment while steering less complex (and hence lower expected cost) patients towards hospitals that remained under bundled payment. However, Appendix Table A.1 finds no clear evidence of such effects, at least during the first year of the voluntary regime, which is covered by our data.

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 of the average episode. Appendix Table A.2 shows that the main results from Table I 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 A.2.

(Lack of) Dynamic Effects. Appendix Table A.3 explores the dynamics of these treatment effects. We focus on the outcomes where the treatment effects are noticeable: total episode claims, institutional PAC claims, and the share discharged to institutional PAC. For these outcomes, we report treatment effects for four, roughly biannual time periods in period 2 (2016-2017) for all MSAs (Panel A) and Voluntary MSAs (Panel B). In addition, when we analyze the time pattern of treatment effects for Mandatory MSAs (Panel C), we further extend the analysis to add two additional time periods in 2018 (the first year of period 3). The results indicate that the treatment effects occur immediately, and are largely stable over time, albeit with a slight (but statistically insignificant) upward trends for total episode and institutional PAC claims.

The immediate and relatively stable treatment effects suggest that dynamics (e.g., learning by doing) are unlikely to be first order. The similar time patterns for the treatment effects across mandatory and voluntary MSAs (compare Panels B and C of Appendix Table A.3) suggest that there was little anticipatory response to the opportunity to opt out of CJR in period 3 (finalized in December 2017). That is, there is no evidence that the treatment effects fade out in voluntary MSAs in anticipation of the opt out of many of these hospitals in 2018.

More generally, the patterns are consistent with qualitative evidence from Zhu et al. (2018) and The Lewin Group (2018, 2019a,b, 2020) on how hospitals responded to the incentives from bundled payments. Based on semi-structured interviews with hospital executives and administrators, these authors report that hospitals endeavored to reduce spending on institutional PAC through a mix of fixed cost and variable cost activities. To reduce discharges to institutional PAC, hospitals employed pre-surgery patient education and physical therapy (so-called “pre-habilitation”), both primarily variable costs. To control spending conditional on discharge to institutional PAC, hospitals worked to form networks of preferred SNFs and to monitor patients at these SNFs. While the formation of networks is primarily a fixed cost, building data platforms and hiring care coordinators to follow up with PAC facilities and track patient suitability for discharge involves a mix of variable and fixed costs.

III.C. Heterogeneity in Levels and Slopes

The average treatment effects reported in the last section mask substantial heterogeneity in the levels and slopes across hospitals. Table II examines this heterogeneity, focusing again on total episode claims, institutional PAC claims, and the share discharged to institutional PAC.

To examine hospital-specific levels, in principle we would want to examine the level for each outcome

in period 2. Because treated hospitals change behavior in response to bundled payments (see Table I), we do not observe unaffected spending levels. The model developed in the following sections will allow us to formally recover counterfactual spending levels. To provide model-free evidence, for now we simply look at outcomes in period 1, when all hospitals were paid under FFS. Since levels are strongly auto-correlated within hospitals over time, period 1 levels are a good proxy for levels in period 2.¹⁰

To construct hospital-specific slopes (i.e., the behavioral changes in response to bundled payments in period 2), we estimate a modified version of Equation 1 that allows the treatment effect of bundled payments to vary by hospital. Letting $outcome_{h2}$ denote the average episode outcome for hospital h in period 2, we estimate:

$$(2) \quad outcome_{h2} = \beta_0 + \beta_{1h}BP_h + \beta_2outcome_{h,2014} + \beta_3outcome_{h,2013} + \delta_{s(h)} + \epsilon_h,$$

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 rather than the MSA 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 I).

The top row of Table II reports the mean and standard deviation of the hospital-specific levels and slopes. The means of the hospitals-specific levels and slopes estimates are similar to control mean and average effects reported in Table I, while the standard deviations indicate substantial heterogeneity in levels and slopes across hospitals.¹¹ The remaining rows of Table II reports coefficients from bivariate regressions of these hospital-specific levels and slopes on hospital characteristics. Panel A shows that levels for each outcome are lower at hospitals with more CJR episodes, a higher quality index, and for-profit facilities, but higher at larger hospitals, teaching hospitals, and non-profits. The patterns are similar in Panel B, where we examine associations with hospital slopes. That is, the hospital characteristics that are associated with low levels are also associated with larger decreases in outcomes in response to the incentives from bundled payments.¹²

¹⁰Specifically, 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.

¹¹Heterogeneity in the hospital-specific slopes may partially reflect idiosyncratic changes within hospitals over time. In the model we specify below, we will explicitly model such idiosyncratic variation, which may be partially driven by measurement error.

¹²To examine sensitivity of these estimates, we estimate slopes using a modified version of Equation 2 that controls for hospital-

While there are clear patterns in these associations, hospital characteristics only explain a small share of the overall variation in levels and slopes. In Appendix B, we describe a variance decomposition exercise that quantifies the explanatory power of the hospital characteristics. The accompanying Appendix Table A.5 shows that the hospital-level characteristics explain little of the cross-hospital variation. Specifications with all of the characteristics, along with strata and MSA fixed effects, explain only a quarter of the variation in levels and even less of the variation in slopes. Fully saturated specifications that additionally control for all available patient characteristics still leave at least half of the variation unexplained.

In addition to elucidating the specific activities that hospitals engage in to respond to the incentives from bundled payments, the interviews conducted in Zhu et al. (2018) and The Lewin Group (2018, 2019a,b, 2020) also shed light on some of the sources of the remaining unexplained heterogeneity in slopes (i.e., treatment effects). Some hospitals reported that prior experience with bundled payments (e.g., with private insurance) or with other value-based performance reforms in Medicare left them well-posed to make changes, while others noted that their prior efforts left less scope for further improvement (The Lewin Group, 2018). There was more general agreement across hospitals on the importance of having a “physician champion” who would coordinate and lead efforts for change in response to the bundled payment (The Lewin Group, 2019b). Hospitals also noted the importance of obtaining buy-in from surgeons, which was generally easier when surgeons were employed by the hospital but could also be achieved when surgeons were not employees (The Lewin Group, 2020, page 114).

III.D. 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 bundle 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 bundle 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 (levels) and behavioral responses to bundled payments in period 2 (slopes). Table III presents the results.

specific linear time trends using hospital level data going back to 2010. Appendix Table A.4 reproduces the analysis in Table II using these alternative estimates of the hospital level slopes.

Panel A of Table III 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 I). 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 and out of bundled payments. The three rows report results for the three outcomes where we observed a statistically significant impact of bundled payments in Table I: episode claims, claims for institutional PAC, and share of patients discharged to institutional PAC.

The results – which are similar to those in Wilcock et al. (2021) – are consistent with selection on levels. In the last section (Table II) we showed substantial heterogeneity across hospitals in period 1 levels, indicating potential scope for selection. Columns (2) and (3) of Table III show that, as expected, hospitals who select into bundled payments have, on average, about \$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.

To assess selection on slopes, we use the estimated hospital-specific slopes from the previous section (i.e., the estimated 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. Panel B of Table III 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 \$791, compared to a lower reduction (\$665) 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 (\$518) and those who select out (\$176) in period 3 is statistically distinguishable ($p=0.05$).¹³

¹³In Appendix Table A.6 we examine the sensitivity of these estimates to a specification that allows for hospital-specific (linear) time trends using data going back to 2010. The general results remain qualitatively similar, though differences between select-in and select-out hospitals remain statistically insignificant and the difference between the point estimates are smaller.

Panel C of Table III 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 V. 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.

IV. MODEL OF VOLUNTARY SELECTION

IV.A. 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 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 $f_i^{OTH} = c_i^{OTH}$.¹⁴ In what follows, for each variable x_i , we focus on hospital-level averages, defined as $x_h = \frac{1}{n_h} \sum_{i=1}^{n_h} x_i$, 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 payments, Medicare reimburses the admitting hospital the fixed bundle price b_h for the entire episode, so average government spending per episode is b_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 un-

¹⁴This 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.

der 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 $\phi''_h > 0$.¹⁵ Hospitals, thus, choose effort to maximize

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

and optimal effort is pinned down by $\phi'_h(e_h^*) = 1$. Since the hospital internalizes both the social marginal cost and benefit of effort, bundled payments results in the first-best effort level.

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 $\pi_h^{BP} = b_h - (c_h^{HOSP} + f_h^{OTH} - \frac{\omega_h}{2})$.

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

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

where the right hand side is the sum of a “level” effect ($b_h - \lambda_h$) and a “slope” effect ($\frac{\omega_h}{2}$). The level effect ($b_h - \lambda_h$) 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?” (ArborMetrix, n.d.)

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

¹⁵For 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.

surplus (S) and producer profits (π), minus government spending (G) weighted by the marginal cost of public funds $\Lambda > 0$:

$$(5) \quad W = S + \pi - (1 + \Lambda)G.$$

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 the government (or consumers) rather than in the hands of hospitals.¹⁶ Consistent with the descriptive results in Section III, we assume that the hospital's effort does not affect patient welfare (S).

Government spending per episode (G) is b_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

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

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 $b_h - \lambda_h$, which is associated with a social cost of Λ per dollar transferred. 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.

IV.B. Graphical Intuition

We illustrate the setting graphically in Figure II, 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

¹⁶Recall 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.

on the financial incentive, b_h , the tradeoff is represented in Figure II.

To draw the figure, we hold the bundle price b 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 ($b - \lambda$) 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 $b_h - \lambda_h > 0$), 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 II represents the set of hospitals for which social welfare is the same whether they participate in bundled payment 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 ($b - \lambda$) 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 II 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).

IV.C. 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 $b_h = \lambda_h - \frac{\omega_h}{2}$ for each hospital. Under these bundle 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 III illustrates this tradeoff in our setting. In Panel (A) we start with Figure II and superimpose on it the participation and social welfare indifference sets that are associated with a higher bundle price $b' > b$. The blue (solid and dashed) indifference lines that correspond to b' are analogous to the gray lines, which correspond to b . 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 $\Lambda(b' - b)$, 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 bundle prices. This is shown in the remaining panels of Figure III, 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 payment 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 payment. However, when ω_h is low (i.e., the cost of effort is higher and optimal effort is lower), as in Panel (C), selection is primarily driven by levels, and it is difficult to generate social-welfare enhancing participation by hospitals. In this case, it requires much 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 payment. 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 key in assessing the potential social welfare gains from alternative payment models. It is therefore the object of interest in our econometric exercise in the next section.

V. SPECIFICATION AND RESULTS

V.A. 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 payment has not been announced. In period 2, a subset of the hospitals are randomly assigned to bundled payment. In period 3, a subset of these latter hospitals endogenously choose to remain under bundled payment while the rest switch back to FFS.

For each hospital in the sample, we observe three periods of claims data $\{y_{h1}, y_{h2}, y_{h3}\}$ and three periods of bundled payment participation indicators $\{BP_{h1}, BP_{h2}, BP_{h3}\}$, such that in period 1 $BP_{h1} = 0$ for all hospitals, in period 2 BP_{h2} depends on the treatment assignment, and in period 3 BP_{h3} stays the same as in period 2 for all hospitals, except for those in the voluntary treatment group, which endogenously choose their period 3 participation status. For hospitals randomly assigned to bundled payment in period 2, we also observe the hospital-specific bundle prices, $\{b_{h2}, b_{h3}\}$ in period 2 and period 3 (even if they select out of the program).

The model in Section IV defines each hospital by two hospital-specific parameters (level and slope). To accommodate the panel nature of the data, we maintain the assumption that hospitals are associated with these level and slope parameters, $\{\lambda_h, \omega_h\}$. However, while we assume that ω_h is fixed over time (consistent with the evidence presented in Appendix Table A.3) and is known to the hospital, we allow the realized level parameter to vary over time, such that

$$(7) \quad \ln \lambda_{ht} = \ln \lambda_h + \gamma_t + \epsilon_{ht} \quad \text{for } t = 1, 2, 3,$$

where $\{\gamma_1, \gamma_2, \gamma_3\}$ are period-specific indicators (with a normalization of $\gamma_2 = 0$), and ϵ_{ht} is drawn (iid) from $N(0, \sigma_\epsilon^2)$. Thus, using the model of Section IV, these assumptions imply that claims are given by

$$(8) \quad y_{ht} = \lambda_{ht} - BP_{ht} \omega_h \quad \text{for } t = 1, 2, 3.$$

We also assume that hospitals know their level type λ_h and the period-specific shifters (γ_1 and γ_3), but do not have any ex-ante information about the random shock ϵ_{ht} .

We further assume that $\{\lambda_h, \omega_h\}$ are drawn from a joint lognormal distribution with a correlation parameter ρ , so that

$$(9) \quad \begin{pmatrix} \ln \lambda_h \\ \ln \omega_h \end{pmatrix} \sim N \left(\begin{pmatrix} x'_h \beta^\lambda \\ x'_h \beta^\omega \end{pmatrix}, \begin{pmatrix} \sigma_\lambda^2 & \rho \sigma_\lambda \sigma_\omega \\ \rho \sigma_\lambda \sigma_\omega & \sigma_\omega^2 \end{pmatrix} \right),$$

where x_h is a vector of hospital characteristics as well as strata fixed effects.¹⁷ Because claims for index admission are fixed, we restrict ω_h from being infeasibly large by truncating it from above at $0.71\lambda_h$.¹⁸

Finally, we use the model of Section IV to specify the participation decision:

$$(10) \quad BP_{h3} = 1 \iff (b_{h3} - \lambda_h \exp(\gamma_3)) + \frac{\omega_h}{2} + v_h > 0.$$

The participation decision (which only applies for period 3) follows the model we described in Section IV, with two changes. First, because hospitals do not know ϵ_{h3} at the time that they make their participation decision, they do not observe λ_{h3} and instead base their decision on the components they do observe (namely, λ_h and γ_3).¹⁹ The second change is that we introduce a hospital-level choice shifter, v_h , into the participation equation, which we assume to be drawn (iid) from $N(x'_h \beta^v, \sigma_v^2)$. For now, we remain agnostic as to what this choice shifter represents and, in particular, whether it only affects hospital choice or is also relevant for social welfare. In our counterfactual analyses in Section VI we consider both possibilities.

To summarize, our econometric model is fully characterized by Equations 7, 8, and 10, and the parametric (normal and lognormal) distributional assumptions we described. It can be thought of as three equations, one for λ , one for ω , and one for v , each associated with mean-shifter parameters β and a variance parameter σ . The remaining model parameters are γ_1 , γ_3 , σ_ϵ , and ρ .

¹⁷Since 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.

¹⁸The 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.

¹⁹This specification captures a setting where hospitals base their participation decision on their point estimate for λ_{h3} without accounting for the realization noise. It can also be viewed as an approximation of a setting where hospitals make their participation decision based on expected profits and integrate over the realization noise. In this latter case, given the lognormal distribution of λ_{h3} , part of expectation based on the variance of σ_ϵ will be captured by the mean (across hospitals) of the choice-shifter v_h .

V.B. 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: $\beta^\lambda, \sigma_\lambda, \gamma_1, \gamma_3, \sigma_\epsilon$; a set of parameters that correspond to the reduction in claims under bundled payment: $\beta^\omega, \sigma_\omega$; a correlation parameter that relates the levels and slopes: ρ ; and choice shifter parameters: β^v, σ_v . The intuition for identification follows in three steps.

First, using data from the control group alone, in which we observe $\lambda_{h1}, \lambda_{h2},$ and λ_{h3} for the same set of hospitals, we can identify $\beta^\lambda, \gamma_1, \gamma_3, \sigma_\lambda,$ and σ_ϵ . Intuitively, these would be identified from a standard 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 the mandatory assignment part of the data (that is, periods 1 and 2), we can identify $\beta^\omega, \sigma_\omega,$ and ρ . 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. Because of random assignment we know that γ_1 , which is identified off 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_1 identifies β^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 ρ 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 remaining choice shifter parameters β^v and σ_v . 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_h 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 systematic deviation from this “predicted” participation rate identifies β^v and σ_v .

V.C. Estimation

We estimate the model using a Markov Chain Monte Carlo (MCMC) Gibbs sampler. Without the choice equation, the model is linear and resembles a random effects model, so estimation using traditional methods such as maximum likelihood would be fairly straightforward. However, the choice equation makes estimation more difficult, as it requires us to numerically integrate over two-dimensional random effects, which combined with the large number of parameters introduced to capture observed heterogeneity across hospitals, makes optimization infeasible in our computing environment.

In contrast, the Gibbs sampler is tractable because our econometric model is fully parametrized, and we can rewrite our econometric model in a hierarchical manner, augmenting λ_h , ω_h , and ν_h as (pseudo) parameters of the model. This means that we can break up simulation from the posterior distribution into smaller steps. Conditional on draws of λ_h , ω_h , and ν_h the model is linear, and pretty standard. And drawing values from the posterior distribution of each of the components of λ_h , ω_h , and ν_h , conditional on the other two components, is reasonably simple. Appendix C provides complete details of the estimation procedure.

We verified the estimation method using simulated data. Reassuringly, the model also generated similar estimates for a simpler version of the model from a previous version of the paper that we estimated by maximum likelihood (Einav et al., 2020b). The Gibbs sampler converges to a stable (posterior) distribution after several thousands of simulations. We thus ran the Gibbs sampler for 100,000 iterations, and used the last 90,000 to construct our parameter estimates. For each parameter, we report the posterior mean and posterior standard deviation, which we will refer to interchangeably as the parameter estimate. For other quantities of interest, we compute each quantity simulation-by-simulation, and then report the posterior mean and posterior standard deviation, thus capturing any underlying correlation among parameters. Appendix D provides more details on how we map the MCMC estimates to the results reported below.

V.D. Results

We present the results in Table IV, Table V, and Figure IV.²⁰ We start in Table IV by presenting the parameter estimates. Because the parameter estimates are not always economically interpretable, in Table V we present summary statistics on the distribution of key economic objects. Finally, in Figure IV, we present

²⁰Recall that we define period 2 to include data from 2016 and 2017. To address the potential concern that hospitals might have adjusted their behavior in anticipation of the switch to voluntary participation that was finalized in December 2017, we re-estimate the model and re-conduct the counterfactuals using data from 2016 only (and not 2017) for period 2. The results, which are essentially unchanged, are shown in Appendix Tables A.7, A.8, and A.9.

empirical analogues of the selection figures we used to illustrate the model in Section IV.

Panel A of Table IV presents parameter estimates associated with the three hospital-specific components of the model ($\ln \lambda_h$, $\ln \omega_h$, and ν_h). We demean the hospital characteristics so that the constants can be interpreted as the mean of that component of the model. While we estimate separate constants for each stratum, for the sake of brevity we report the episode-weighted mean of the strata-specific constants in the table.

The parameter estimates for the hospital characteristics largely match the patterns from the descriptive analysis shown in Table II. First, there is a distinctive role for large hospitals in terms of their number of beds relative to “experienced” hospitals in terms of their CJR episode volume. Large hospitals are associated with higher levels and slopes, while hospitals with a high volume of CJR episodes are associated with a smaller level and a smaller slope. Large hospitals also have higher propensity to select into the bundled payment program, although all of the parameters associated with ν are imprecisely estimated. Second, high-quality hospitals have slightly lower episode spending, are more responsive to the incentives from bundled payments, and based on the point estimate, are more likely to select into the bundled payment programs. Third, for-profit hospitals respond to incentives more strongly and, based on the point estimates, are more likely to select into the bundled payment program relative to teaching hospitals or government-owned hospitals. Finally, via the observables, the choice shifter ν is negatively correlated with the level of spending (correlation of -0.136) and positively correlated with the slope (correlation of 0.108).²¹ Since the choice equation already accounts for the level and slope incentives, this suggests that hospitals are perhaps even slightly overweighting the level and slope incentives in their selection decisions. We defer discussion of the constants and the variance parameters (σ_λ , σ_ω , σ_ν) to our discussion of Table V below.

Panel B shows estimates of the time trend and other non-hospital specific parameters. The γ_t parameters capture the time trends in the level of spending. The time trend is u-shaped with a large 6.7% decline in spending between period 1 and period 2 ($\gamma_1 = 0.067$ relative to $\gamma_2 = 0$) followed by a small 1.5% increase from period 2 and period 3 ($\gamma_3 = 0.015$). The correlation between the level and slope is positive but small ($\rho = 0.14$). We discuss σ_ϵ below.

Panel A of Table V presents summary statistics on the distribution of key economic objects across hospitals, weighting each hospital by the number of episodes per hospital so that the statistics are representative of

²¹To estimate this correlation, we use the correlation (across hospitals) between ν and λ and between ν and ω in each iteration of the Gibbs sampler, and then average across iterations. Recall that ν is assumed to be independent of other parts of the model, so this correlation is solely driven by observables.

the average episode. Among all hospitals, the mean level of claims ($E[\lambda_{h3}]$) in period 3 is \$27,028, and the mean slope ($E[\omega_h]$) is \$485, which is qualitatively similar but somewhat smaller than the estimate of \$797 in Table I. Although not statistically significant, the difference in point estimates presumably reflects some combination of the log specification and richer structure of the estimated model. The 0.19 standard deviation of $\ln(\lambda_h)$ is large relative to the 0.073 standard deviation of the idiosyncratic component ($\sigma_\epsilon = 0.073$ in Table IV), indicating that the cross-hospital component is the primary determinant of the level of claims.

Panel B shows summary statistics for hospitals in the voluntary treatment group. This is the group of hospitals for which we can observe bundle prices and will be the focus of our counterfactuals in the next section. Among voluntary treatment hospitals, the mean level of claims is unsurprisingly lower for all hospitals (recall that the voluntary treatment group was specified as treatment group hospitals from MSAs with below-median historical spending) and the slope is lower as well.

The model emphasizes the importance of heterogeneity in levels relative to the slope in determining the nature of selection and social welfare under voluntary bundled payments. The standard deviation of the level of claims (\$4,527) is much larger than the mean or standard deviation of the slope (\$246 and \$611, respectively). While the observed bundle prices reduce the heterogeneity in levels by roughly half (standard deviation of $b_h - \lambda_h \exp(\gamma_3)$ of \$2,640), the reduction is not large enough to “undo” the result that most of the selection is on levels. Finally, the choice-shifter term ν has a large negative mean and substantial variation.

One question is whether additional observables could be used to better set the bundle prices. A comparison of the conditional standard deviation of $\ln \lambda$ (0.139 in Table IV) to the “unconditional” standard deviation of $\ln \lambda$ (0.19 in Table V) is not encouraging, as it indicates that observables account for less than one third of the cross-hospital variation in levels.²² This underscores the difficulty in customizing bundle prices appropriately to specific hospitals, a point we will return to in the next section.

To further examine the role of bundle prices, Figure IV produces empirical analogues of the selection figures (Figure III) we used to illustrate the model in Section IV, again focusing on the set hospitals in the voluntary treatment group where we can observe bundle prices. To provide a baseline, Panel (A) of Figure IV plots simulated hospitals from the joint distribution of ω_h (vertical axis) and λ_{h3} (horizontal axis), without netting out the bundle price. Consistent with the discussion of Table V, Panel (A) indicates that selection on levels is a primary concern, with a large mass of hospitals selecting bundled payments inefficiently or

²²In contrast, the analogous exercise indicates that observables account for more than half of the cross-hospital variation in slopes.

selecting FFS. Panel (B) of Figure IV examines the role of targeting by plotting $\lambda_{h3} - b_h + \bar{b}$ on the horizontal axis, which nets out the hospital’s bundle price b_h and adds in the average bundle price \bar{b} so that the axis remains on the same scale as in Panel (A). Netting out bundle prices shrinks the heterogeneity along the horizontal axis, but not enough, so that a sizable mass of hospitals continue to select bundled payments inefficiently or select FFS.

VI. COUNTERFACTUALS

We use the estimated model to perform two sets of counterfactual exercises. First, we compare social welfare under mandatory FFS, mandatory bundled payments, and voluntary bundled payments. Second, within the voluntary bundled payments regime, we consider how alternative bundle prices affect social welfare.

The current bundled payment option offers the choice of only one bundle price to each hospital, 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 bundle 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 IV, 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.

Throughout, we focus on 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 I), since this is the group of hospitals for which we observe bundled prices.

VI.A. 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 canceling the program entirely or keeping it mandatory. Throughout, we assume a social cost of funds of $\Lambda = 0.15$. Appendix D provides more details about how

we map the estimation results to the reported quantities.

Panel A of Table VI 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 (G in the definition of social welfare from Equation 5) averages \$25,517 per episode, which corresponds to the mean λ_{h3} reported in Panel B of Table V. 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 payments, hospitals receive a transfer of $(b_h - \lambda_{h3})$ and are residual claimants on the ω -related savings they generate ($\frac{\omega_h}{2}$). If bundle prices had been calibrated to equal counterfactual claims under FFS incentives ($\mathbb{E}_h b_h = \mathbb{E}_h \lambda_{h3}$), government spending would have been unaffected relative to the baseline. However, as seen in Panel B of Table V, bundle prices b_h were on average nearly \$2,000 lower than FFS claims, so government spending is \$1,859 lower and (multiplied by $1 + \Lambda = 1.15$) social costs decrease by \$2,137 (column (3)).

We consider two different versions of the welfare analysis, depending on whether we treat the choice shifter term, ν_h , as 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) of Table VI, where we assume that ν_h is not welfare relevant, hospital profits relative to FFS decline by \$1,736 (column (4)). This decline reflects the difference between the ω -related savings of \$123 (half of the expected value of ω_h from Panel B of Table V) and the \$1,859 reduction in government payments. Social surplus rises by \$402 (column (5)). In other words, the incentive effects of bundled payment (which generate $\frac{\omega_h}{2} = \$123$ in social savings on average) represent about one third of the social gain, with the remainder coming from the reduction in government spending. Naturally, if ν_h is taken into account (as in Appendix Table A.10), both hospital profits and relative social surplus are lowered by its average value of more than \$7,700 (see Panel B of Table V).

The third row of Panel A considers the voluntary selection scenario that actually took place. (To be consistent with the other counterfactuals, outcomes are still simulated from the model.) 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 III), providing assurances about the in-sample fit of the model. As we discussed in Section V, the much greater heterogeneity in $b_h - \lambda_h$, relative to $\frac{\omega_h}{2}$, suggests that selection into bun-

dled payments is primarily on levels. Consequently, voluntary participation raises government spending per episode by \$1,396 relative to the mandatory bundled payment regime, but lowers government spending per episode by \$463 relative to the mandatory FFS regime.

Since hospitals are given a choice, their profits must be weakly higher under the voluntary regime relative to a mandatory regime. However, because we treat the ν_h term as non-welfare relevant, hospital profits (net of ν_h) decline by \$405 relative to the mandatory FFS benchmark, which is \$1,330 higher than under mandatory bundled payments. Ignoring ν_h , social surplus under voluntary bundled payments is higher than under the mandatory FFS benchmark (by \$127, column (5)) but lower than the mandatory bundled payment regime (by \$275); 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 (in Appendix Table A.10), hospital profits and social welfare are much higher under the voluntary regime, as hospitals sort on the value of their choice shifter, and the standard deviation in ν_h across hospitals is large. 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 generates social surplus relative to the mandatory FFS status quo. Yet, the gains (when we ignore the ν_h term) are relatively modest. This is because voluntary bundled payments generate positive ω -related efficiency gains, which are partially offset by 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 bundle prices to better reflect underlying hospital-specific costs and thus capture more of the \$60 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 \$123).

VI.B. Targeting

In order to explore price targeting under voluntary participation in a systematic fashion, we approximate the observed bundle prices using a parametric distribution, and then examine the impacts of shifting its parameters. Specifically, we assume that hospital-level bundle prices, b_h , are lognormally distributed and are correlated with hospital costs. We then explore voluntary participation under different parameter values. Appendix E provides more details.

Figure V summarizes the outcomes from this exercise, plotting social surplus relative to the mandatory FFS benchmark (y-axis) against government spending (x-axis) for different bundle price counterfactuals. In the plot, we focus on the social surplus values that do not consider ν_h (column (5) of Table VI). Panel B of Table VI reports outcomes associated with each exercise (and Appendix Table A.10 reports the parallel exercise when we treat ν_h as welfare relevant). The black dot in Figure V corresponds to the observed distribution of b_h , and serves as a baseline. Because of the lognormal parameterization of b_h , the outcomes for observed targeting in Panel B of Table VI are slightly different from 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 bundle prices so that they perfectly correlate with the realized claims under FFS (λ_{h3}), while maintaining the mean of the log bundle price at the level we estimate in the data. This improved targeting leads to lower government spending, and raises average social surplus by \$155 per episode, or \$22 more per episode (17%) than observed targeting.

Perfect targeting is a useful benchmark, but not feasible in the context of our model. To see this, recall that we model $\ln \lambda_{h3} = \ln \lambda_h + \gamma_3 + \epsilon_{h3}$, 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 labeled “feasible targeting” in which Medicare sets a bundle price that is perfectly correlated with $\ln \lambda_h$ but is uncorrelated with the ϵ_{h3} . We estimate that relative to the observed targeting, feasible targeting generates average welfare gains of \$17 per episode, or 77% ($= (\$150 - \$133) / (\$155 - \$133)$) of the social welfare gains that could be achieved with perfect targeting.

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 bundle prices are uniform across hospitals at a value equal to the average of bundle price. Relative to the observed targeting, the no targeting case leads to slightly greater participation in bundled payments, higher government spending, and welfare gains of \$91 per episode relative to the FFS benchmark. Within \$91 to \$155 the range defined by no targeting at the bottom and feasible targeting at the top, the observed targeting generates approximately 66% ($= (\$133 - \$91) / (\$155 - \$91)$) of the feasible gains.

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, we decompose average per-episode claims into two additive separable components: hospital claims and other claims, which include post-acute care. Because the hospital claims portion was already reimbursed by a predetermined, DRG-based amount under FFS, we do not expect hospital claims to respond to the incentives from bundled payments; in effect, the hospitals already faced a bundled payment for the services they provided. Indeed, the empirical evidence show that most of the ω -related savings come from non-hospital 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 eliminating hospital claims from the bundle. Under this counterfactual payment regime, labeled “narrow bundling, no targeting,” the bundled payment for hospitals in treatment MSAs now covers only the non-hospital costs; in both treatment and control MSAs, hospitals continue to be paid a lump sum DRG-based “bundle” to cover the within-hospital costs. As shown in the last row of Panel B, this bundle achieves almost identical social surplus to the observed targeting. This suggests that a narrowly targeted bundle with even a modest amount of targeting of non-hospital costs would raise surplus relative to the observed regime.

Overall, our findings suggest that the observed targeting performs fairly well, but that there are also feasible improvements in targeting which could generate meaningful reductions in government costs and higher total surplus. These could arise through tailoring bundle 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.

VII. 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 treatment markets, with hospitals in 104 other randomly selected control markets paid under the status quo. Two years into this five-year experiment, however, hospitals in half of the treatment markets were allowed to choose whether to remain under bundled payments or revert back to the status quo payment model. This provided an opportunity to estimate the impact of the alternative model on all hospitals' behavior 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 (selection on levels). 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 only a modest increase in social welfare relative to imposing the status quo fee-for-service (FFS) payment regime on all hospitals. However, we also estimated that alternative (feasible) voluntary 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 et al., 1985).

Our quantitative results are, of course, specific to our setting. They are, however, likely to be fairly representative of Medicare's experience with alternative payment models, which comprise over 30% of Medicare spending (Shatto, 2016). In these other models, the estimated impacts (slopes) have also tended to be modest – "singles not home runs" in the words of Frakt (2019a). This suggests that, as in our setting, there may be limited scope for social-welfare improving selection on slopes, and that voluntary models should be designed with particular attention toward limiting selection on levels.

In addition, while the cancellation of a mandatory participation RCT provides us with a unique opportunity as researchers to estimate treatment effects and then observe participation decisions, from the participant perspective, our setting is similar to other Medicare programs in allowing one-time retroactive withdrawal from the program. For example, the BPCI-Advanced models which are voluntary bundled payment models for certain inpatient and outpatient episodes allowed participating hospitals the opportunity to leave the program – i.e. retroactively withdraw with no financial consequences – during the first six months of the

program. Thus, in these voluntary participation programs, as in the one we study, there is opportunity for participants to learn something about their levels and slopes before making a final participation decision (CMS, 2018).

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 few exceptions, 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, electricity, and environmental regulation. In these sectors, in addition to the tradeoff between selection on slopes and levels, firm exit may be an important margin of adjustment. Exploring the impact and optimal design of such voluntary programs in such settings – with heterogeneity in levels, slopes, and exit propensity – is an important and fruitful direction for further work.

STANFORD UNIVERSITY AND NATIONAL BUREAU OF ECONOMIC RESEARCH, UNITED STATES
MASSACHUSETTS INSTITUTE OF TECHNOLOGY AND NATIONAL BUREAU OF
ECONOMIC RESEARCH, UNITED STATES
HARVARD UNIVERSITY, UNITED STATES
STANFORD UNIVERSITY AND NATIONAL BUREAU OF ECONOMIC RESEARCH, UNITED STATES

REFERENCES

- Alexander, Diane (2020), “How do doctors respond to incentives? unintended consequences of paying doctors to reduce costs.” *Journal of Political Economy*, 128, 4046–4096.
- ArborMetrix (n.d.), “A data-driven approach to bpci advanced: Methods for selecting appropriate risk.” Technical report.
- 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, 252–262.
- Buhagiar, Mark A, Justine M Naylor, Ian A Harris, Wei Xuan, Friedbert Kohler, Rachael Wright, and Renee Fortunato (2017), “Effect of inpatient rehabilitation vs a monitored home-based program on mobility in patients with total knee arthroplasty: The hiho randomized clinical trial.” *JAMA*, 317, 1037–1046.

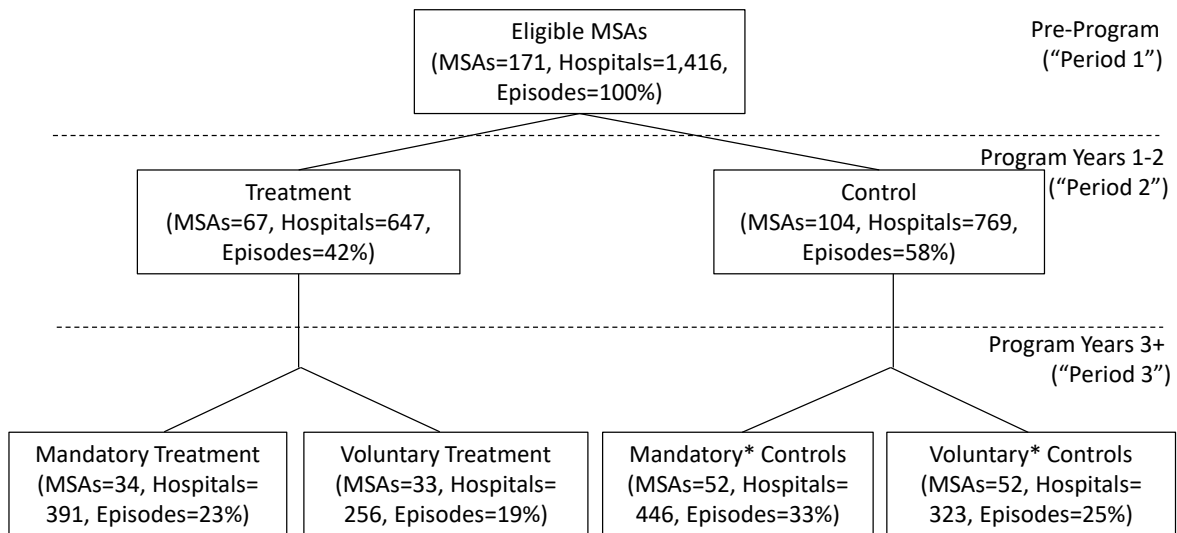
- Carroll, Caitlin, Michael Chernew, A Mark Fendrick, Joe Thompson, and Sherri Rose (2018), “Effects of episode-based payment on health care spending and utilization: Evidence from perinatal care in arkansas.” *Journal of Health Economics*, 61, 47–62.
- Cicala, Steve, David Hémons, and Morten Olsen (2020), “Adverse selection as a policy instrument: Unraveling climate change.” https://home.uchicago.edu/~scicala/papers/unraveling/unraveling_climate_change_draft.pdf, Working Paper.
- Clemens, Jeffrey and Joshua D Gottlieb (2014), “Do physicians’ financial incentives affect medical treatment and patient health?” *American Economic Review*, 104, 1320–49.
- CMS (2015a), “Medicare program; comprehensive care for joint replacement payment model for acute care hospitals furnishing lower extremity joint replacement services. final rule.” *Federal Register*, 80, 73273–73554.
- CMS (2015b), “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, 49325–49886.
- 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 (2018), “CMS statement regarding the retroactive withdrawal policy and the new participation agreement for bpci advanced.” Technical report, Centers for Medicare and Medicaid Services.
- CMS (2019), “CMS fast facts.” Technical report, Centers for Medicare & Medicaid Services.
- CMS (2020), “CMS innovation center episode payment models.” Technical report, Centers for Medicare and Medicaid Services, <https://innovation.cms.gov/files/reports/episode-payment-models-wp.pdf>.
- CMS Innovation Center (n.d.). <https://innovation.cms.gov/initiatives/CJR>.
- 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, 41.
- Cutler, David M (1995), “The incidence of adverse medical outcomes under prospective payment.” *Econometrica*, 63, 29–50.
- Cutler, David M and Kaushik Ghosh (2012), “The potential for cost savings through bundled episode payments.” *New England Journal of Medicine*, 366, 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, 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, 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, 1267–1278.
- Einav, Liran, Amy Finkelstein, Yunan Ji, and Neale Mahoney (2020a), “Randomized trial shows healthcare payment reform has equal-sized spillover effects on patients not targeted by reform.” *Proceedings of the National Academy of Sciences*, 117, 18939–18947.
- Einav, Liran, Amy Finkelstein, Yunan Ji, and Neale Mahoney (2020b), “Voluntary regulation: Evidence from medicare payment reform.” Working paper 27223, National Bureau of Economic Research.
- Einav, Liran, Amy Finkelstein, Raymond Kluender, and Paul Schrimpf (2016), “Beyond statistics: the economic content of risk scores.” *American Economic Journal: Applied Economics*, 8, 195–224.
- Einav, Liran, Amy Finkelstein, and Neale Mahoney (2018), “Provider incentives and healthcare costs: Evidence from long-term care hospitals.” *Econometrica*, 86, 2161–2219.
- 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, 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, 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, 892–900.
- Fisher, Elliott S (2016), “Medicare’s bundled payment program for joint replacement: Promise and peril?” *JAMA*, 316, 1262–1264.
- Frakt, Austin (2019a), “‘value’ of care was a big goal. how did it work out?” *The New York Times*, September 23.
- Frakt, Austin (2019b), “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, 173–191.
- Fromke, 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, 1676–1682.
- GAO (2018), “Voluntary and mandatory episode-based payment models and their participants.” Technical Report GAO-19-156, United States Government Accountability Office.
- Gaynor, Martin, Nirav Mehta, and Seth Richards-Shubik (2020), “Optimal contracting with altruistic agents: A structural model of medicare payments for dialysis drugs.” Technical report, National Bureau of Economic Research.

- Gaynor, Martin, James B Rebitzer, and Lowell J Taylor (2004), “Physician incentives in health maintenance organizations.” *Journal of Political Economy*, 112, 915–931.
- 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, 924–931.
- Heckman, James J and Bo E Honore (1990), “The empirical content of the roy mode.” *Econometrica*, 58, 1121–1149.
- Ho, Kate and Ariel Pakes (2014), “Hospital choices, hospital prices, and financial incentives to physicians.” *American Economic Review*, 104, 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, 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, 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, 305–309.
- Marone, Victoria R and Adrienne Sabety (2021), “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, 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, 322–326.
- Newhouse, Joseph P (1996), “Reimbursing health plans and health providers: Efficiency in production versus selection.” *Journal of Economic Literature*, 34, 1236–1263.
- Newhouse, Joseph P (2004), *Pricing the Priceless: AHhealth Care Conundrum*. MIT press.
- 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.” Technical report, Centers for Medicare and Medicaid Services.
- Shepard, Mark (2020), “Hospital network competition and adverse selection: Evidence from the Massachusetts health insurance exchange.” Working Paper 22600, National Bureau of Economic Research.
- Thaler, Richard H and Cass R Sunstein (2003), “Libertarian paternalism.” *American Economic Review*, 93, 175–179.
- The Hill (2016), “Lawmakers call for end to medicare ’experiments’.” Technical report, <https://thehill.com/policy/healthcare/298769-lawmakers-call-for-end-to-medicare-experiments>.
- The Lewin Group (2015), “CMS bundled payments for care improvement (bpci) initiative models 2-4: Year 1 evaluation & monitoring annual report.” Technical report, The Lewin Group.
- The Lewin Group (2018), “CMS comprehensive care for joint replacement model: Performance year 1 evaluation report.” Technical report, The Lewin Group.
- The Lewin Group (2019a), “CMS comprehensive care for joint replacement model: Performance year 2 evaluation report.” Technical report, The Lewin Group.
- The Lewin Group (2019b), “CMS comprehensive care for joint replacement model: Performance year 2 evaluation report: An in-depth look: Hospital case studies.” Technical report, The Lewin Group.
- The Lewin Group (2020), “CMS comprehensive care for joint replacement model: Performance year 3 evaluation report.” Technical report, The Lewin Group.
- Wilcock, Andrew D, Michael L Barnett, J Michael McWilliams, David C Grabowski, and Ateev Mehrotra (2021), “Hospital responses to incentives in episode-based payment for joint surgery: A controlled population-based study.” *JAMA Internal Medicine*.
- 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, 1282–1289.

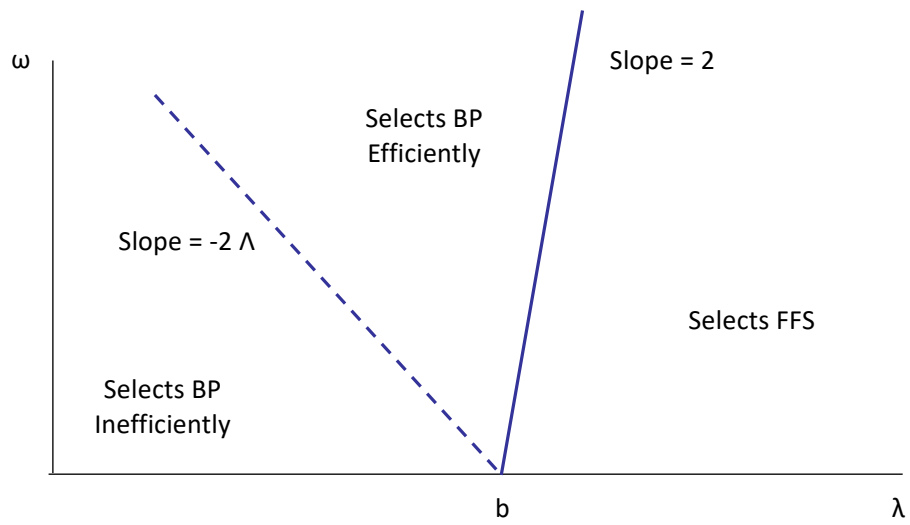
FIGURE I
Experimental Design



Notes: Figure shows the design of the bundled payments experiment. The top part shows the pre-program period. The middle part shows the initial mandatory design in program years 1-2. The bottom part shows the partially voluntary design in program years 3-5. Episode shares are based on data from program years 1-2.

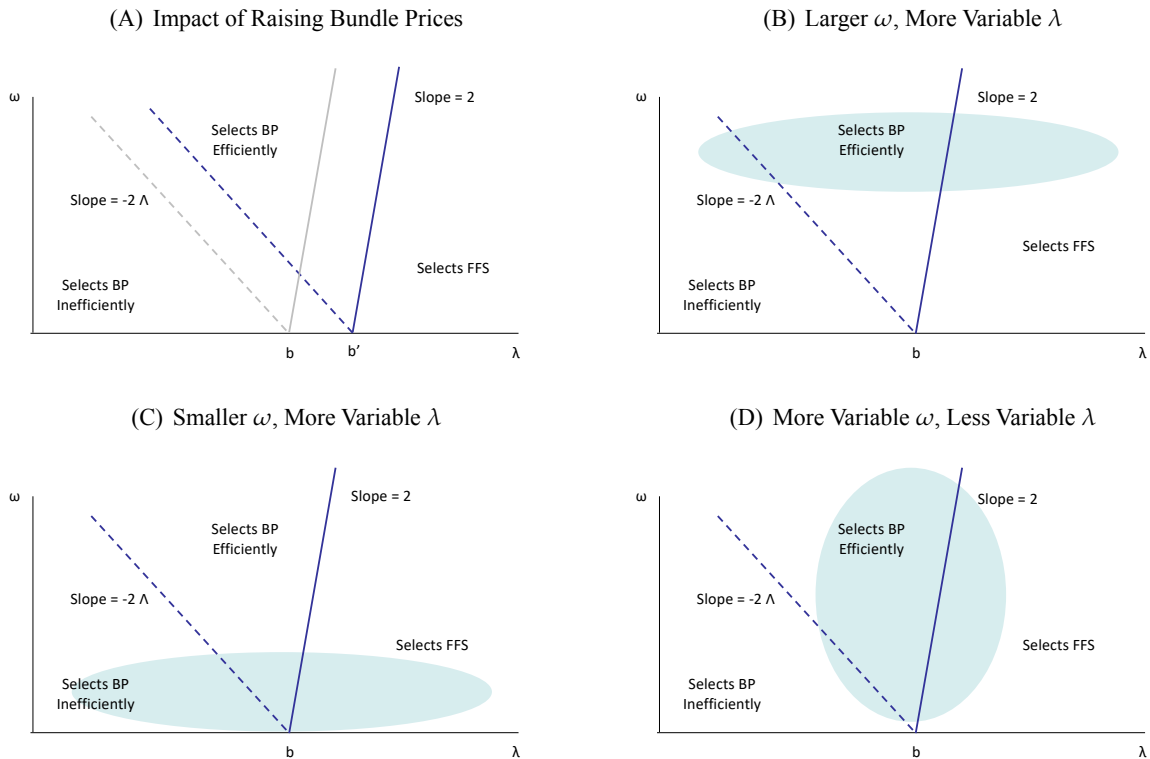
*Control group MSAs were assigned to mandatory vs. voluntary by authors, using the same determination that CMS used to assign Treatment group MSAs to mandatory vs. voluntary. Specifically, the bottom half of the MSAs in the control (or treatment) group – based on historical spending over the period July 1 2011 through June 30 2014 – were assigned to voluntary.

FIGURE II
 Hospital Selection Into Bundled Payment and Social Welfare Implications



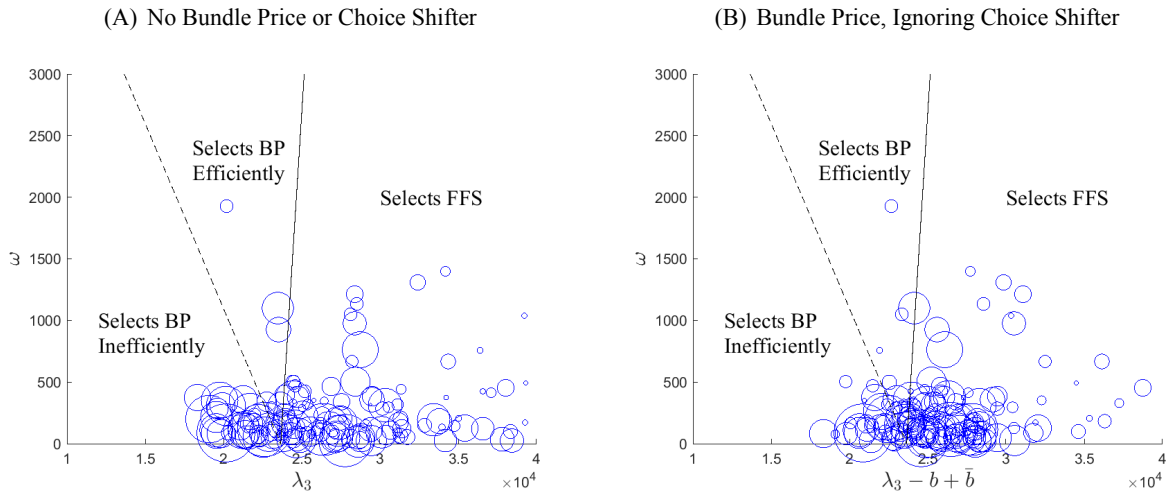
Notes: Figure shows, for a given bundle price b , the hospital participation decision and social welfare implications as a function of the hospital's level (λ), shown on the horizontal axis, and slope (ω), shown on the vertical axis.

FIGURE III
Model Illustration



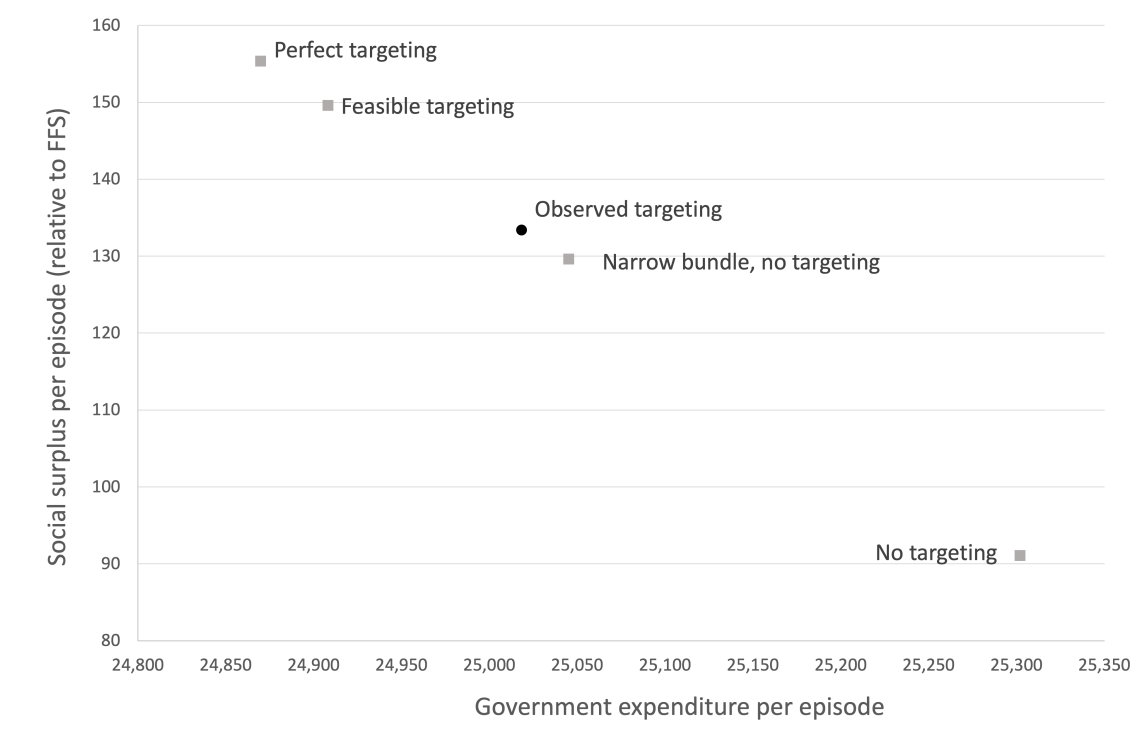
Notes: Figure illustrates some of the key analytics in voluntary bundled payment design as a function of the hospital's level (λ), shown on the horizontal axis, and slope (ω), shown on the vertical axis. Panel (A) illustrates the tradeoffs involved in setting higher bundle prices $b' > b$; Panels (B) 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 IV
Model Estimates



Notes: Figure reports the empirical analogues of Figures II and III. Specifically, it reports simulated hospitals based on our estimates for the 256 hospitals in the voluntary treatment group (see Figure I), with circles proportional to the number of episodes. Panel (A) reports results assuming there are no bundle prices and the choice shifter v_h in the hospital selection equation is not decision-relevant. Selection decisions are characterized by levels (λ_{h3}), shown on the horizontal axis, and slopes (ω_h), show on the vertical axis. Panel (B) considers the role of targeted bundle prices by plotting $\lambda_{h3} - b_h - \bar{b}$ on the horizontal axis. In this panel, in addition to netting out the bundle price b_h , we add the average bundle price \bar{b} so that the axis remains on the same scale as in Panel (A).

FIGURE V
 Medicare Costs and Social Surplus Under Alternative Bundle Prices



Notes: 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 VI. See that table note for more details.

TABLE I
Experimental Estimates During Mandatory Participation Period

	Control Mean	Control SD	Average Treatment Effect	Standard Error	p-value
Panel A: Claims, Utilization, and Gov't Spending (per Episode)					
Claims	25,294	3,603	-790	204	0.001
Claims for Index Admission	13,542	2,389	-169	89	0.06
Claims for Institutional PAC	4,119	1,378	-499	128	0.001
Claims for Home Health	1,800	918	-89	59	0.13
Other Claims	5,832	532	28	55	0.61
Utilization					
Number of Days in Index Admission	2.6	0.4	-0.1	0.0	0.22
Number of Days in Institutional PAC	7.7	2.3	-0.6	0.2	0.01
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	0.019	0.032	-0.004	0.002	0.05
Government Spending	25,294	3,603	40	208	0.85
Panel B: Quality Measures					
Complication Rate	0.011	0.005	0.001	0.001	0.26
ER Visit During Episode	0.198	0.027	0.003	0.003	0.40
90-day All Cause Readmission Rate	0.102	0.015	-0.001	0.002	0.73
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	1.00

Notes: 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 II
Correlates of Levels and Slopes

	Panel A: Heterogeneity in Levels						Panel B: Heterogeneity in Slopes					
	Claims		Claims for Institutional PAC		Probability of Discharge to PAC		Claims		Claims for Institutional PAC		Probability of Discharge to PAC	
Mean (S.D.)	28,357	(5,998)	5,814	(3,021)	0.455	(0.187)	-1,154	(3,054)	-1,011	(1,809)	-0.057	(0.105)
Coefficient (S.E.) from Bivariate Regression												
Number of CJR Episodes	-5.31	(1.56)	-3.39	(0.69)	-0.0001	(0.0001)	-0.85	(0.57)	-1.10	(0.37)	-0.00003	(0.00002)
Quality	-441	(41)	-197	(18)	-0.010	(0.001)	-123	(24)	-77	(15)	-0.004	(0.001)
Number of Beds	3.90	(1.21)	0.62	(0.32)	0.0001	(0.00003)	0.53	(0.46)	0.09	(0.20)	0.000004	(0.00001)
Teaching	4,528	(599)	561	(258)	0.049	(0.021)	819	(525)	-212	(241)	-0.025	(0.017)
For-Profit	-3,030	(660)	-387	(304)	-0.064	(0.025)	-1,012	(651)	-414	(292)	-0.030	(0.020)
Non-Profit	-219	(596)	369	(264)	0.008	(0.023)	-32	(617)	-160	(246)	-0.012	(0.018)

Notes: Panel A reports hospital-specific levels (hospital-specific period 1 outcomes) and coefficients from separate regressions of hospital-specific levels on hospital characteristics. Panel B reports hospital-specific slopes (obtained from estimating Equation 2) and coefficients from separate regressions of hospital-specific slopes on hospital characteristics. In both panels, the coefficients on for-profit and non-profit are obtained from the same regression, where government-owned is the omitted category. All regressions are weighted by the number of episodes in period 2. Robust standard errors are shown in parentheses.

TABLE III
Selection

	Voluntary Control	Voluntary Select-In	Voluntary Select-Out	<i>p</i> -value of Select- In vs. Select-Out Difference
	(1)	(2)	(3)	(4)
Number of Hospitals	323	73	183	
Number of Episodes in Period 1	51,469	14,664	24,777	
Percent of Episodes in Period 1		37.2%	62.8%	
Panel A: Selection on Levels (Period 1 Outcomes per Episode)				
Claims	26,524 (5,367)	26,146 (4,176)	27,776 (5,497)	0.03
Claims for Institutional Post Acute Care	4,811 (2,402)	4,681 (2,054)	5,551 (2,446)	0.01
Share Discharged to Institutional Post Acute Care	36.9% (16.0%)	37.3% (16.2%)	41.6% (14.5%)	0.10
Panel B: Selection on Slopes				
Impact on episode claims		-791 (1,931)	-665 (2,826)	0.73
Impact on Institutional PAC claims		-518 (973)	-176 (1,474)	0.05
Impact on Share Discharged to Institutional PAC		-3.3% (7.8%)	-1.2% (9.2%)	0.12
Panel C: Selection on Hospital Characteristics				
Mean Number of CJR Episodes	306	320	252	0.13
Mean Number of Beds	339	320	362	0.43
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%	0.56
Government-owned	4.0%	10.0%	18.9%	0.20
Mean Quality Score	11.7	13.1	10.6	0.001

Notes: Table reports means (and standard deviations in parentheses). In Panel A, all outcomes are defined in period 1 (although the bundle price is defined in period 2). Panel B reports the average (and standard deviation) of the hospital-specific slopes (obtained from estimating Equation 2). In Panel C, mean number of CJR episodes are based on period 1 data. Mean number of beds and percentage of hospitals that are teaching, for-profit, non-profit, and government-owned are taken 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 323, 73, and 182, respectively. Finally, mean quality score is based on a modified version of the hospitals' composite quality scores from period 2, which is based on the first 18 points of the score (see footnote 7). The *p*-values of the differences are computed based on a simple *t*-test of equality of the means.

TABLE IV
Parameter Estimates

Panel A: Equation-Specific Parameters						
	ln(λ) equation		ln(ω) equation		v equation	
	Mean	Std. Err.	Mean	Std. Err.	Mean	Std. Err.
Constant*	10.165	0.005	4.895	0.293	-7,984	12,616
ln(CJR Episodes)	-0.066	0.004	-0.559	0.145	4,930	7,601
ln(Beds)	0.050	0.006	0.502	0.247	1,172	5,065
Quality score	-0.169	0.022	4.895	1.217	43,473	54,507
Teaching	0.017	0.002	-0.034	0.084	-2,275	3,939
For-profit	-0.008	0.002	0.078	0.066	3,912	4,517
Government-owned	-0.002	0.001	-0.102	0.060	-621	1,322
Non-profit	omitted category		omitted category		omitted category	
Strata fixed effects	yes		yes		yes	
σ	0.139	0.003	0.727	0.117	24,669	27,548
Panel B: Additional Model Parameters						
	Mean	Std. Err.				
γ_1	0.067	0.004				
γ_2	normalized to 0					
γ_3	0.015	0.003				
σ_ϵ	0.073	0.001				
ρ	0.143	0.196				

Notes: Table reports posterior means and posterior standard deviations of the model parameters (parameter estimates), estimated using a Gibbs sampler. Panel A reports parameter estimates associated with the three hospital-specific components of the model (λ_h , ω_h , and ν_h). Hospital characteristics are demeaned so that the constant term can be interpreted as the mean of that component of the model. Panel B shows estimates of the time trend (γ_t) and other non-hospital specific parameters.

*Constant is episode-weighted average of strata specific constants.

TABLE V
Posterior Distributions

	E(x)	SD(x)	P5	P25	P50	P75	P95
Panel A: All hospitals							
$\ln(\lambda_h)$	10.17	0.19	9.89	10.04	10.14	10.27	10.51
$\ln(\omega_h)$	4.895	1.747	1.835	3.717	5.020	6.179	7.528
λ_{h3}	27,028	5,960	19,621	23,043	25,921	29,786	38,238
ω_h	485	1,180	7	44	160	500	1,902
Panel B: Hospitals in the Voluntary Treatment Group Only							
λ_h	25,247	4,527	19,532	22,207	24,390	27,370	33,723
λ_{h3}	25,517	5,109	19,333	21,949	24,597	27,841	35,277
b_h	23,659	3,744	19,530	21,174	22,785	25,110	31,141
$b_h - \lambda_h \exp(\gamma_3)$	-1,961	2,640	-6,728	-3,224	-1,706	-358	1,390
ω_h	246	611	5	24	73	218	1,014
$(b_h - \lambda_h \exp(\gamma_3)) + \omega_h/2$	-1,838	2,623	-6,593	-3,074	-1,584	-258	1,477
v	-7,719	31,103	-59,838	-28,472	-7,227	13,454	42,939

Notes: Table presents summary statistics on the distribution of economic objects, weighting each hospital by the number of episodes per hospital so that the statistics are representative. Panel A shows statistics for all hospitals. Panel B shows statistics for hospitals in the voluntary treatment group, which is the group for which we can observe bundle prices. See Appendix D for more details on how the model estimates are reported.

TABLE VI
Counterfactuals

	Share Selecting In (1)	Government Spending (2)	Relative Social Costs (3)	Ignoring Choice Shifter	
				Relative Hospital Profit (4)	Relative Social Surplus (5)
Panel A: Mandatory vs. Voluntary					
Mandatory FFS (Benchmark)	0.0%	25,517	0	0	0
Mandatory Bundled Payment	100.0%	23,659	-2,137	-1,736	402
Voluntary Bundled Payment	38.8%	25,055	-532	-405	127
Panel B: Alternative Voluntary Regimes with Different Bundle Prices					
Perfect targeting	38.7%	24,870	-745	-589	155
Feasible targeting	38.5%	24,908	-700	-551	150
Observed targeting	38.7%	25,018	-574	-440	133
No targeting	39.1%	25,302	-248	-157	91
Narrow bundle, no targeting	38.5%	25,045	-543	-413	130

Notes: All counterfactuals are conducted on the 259 hospitals in the voluntary treatment group (See Figure I). 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 bundled payments 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 bundle 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 v_h is not welfare relevant; therefore hospital profits relative to FFS are given by $(b_h - \lambda_h) + \frac{\omega_h}{2}$, and social surplus relative to FFS is given by $-\Lambda(b_h - \lambda_h) + \frac{\omega_h}{2}$.