

Provider Incentives and Healthcare Costs: Evidence from Long-Term Care Hospitals*

Liran Einav, Amy Finkelstein, and Neale Mahoney[†]

January 2017

Abstract. We study the design of provider incentives in the post-acute care setting – a high-stakes but under-studied segment of the healthcare system. We focus on long-term care hospitals (LTCHs) and the large (approximately \$13,000) jump in Medicare payments they receive when a patient’s stay reaches a threshold number of days. The descriptive evidence indicates that discharges increase substantially after the threshold, and that the marginal patient discharged after the threshold is in relatively better health. Despite the large financial incentives and behavioral response in a high mortality population, we are unable to detect any compelling evidence of an impact on patient mortality. To assess provider behavior under counterfactual payment schedules, we estimate a simple dynamic discrete choice model of LTCH discharge decisions. When we conservatively limit ourselves to alternative contracts that hold the LTCH harmless, we find that an alternative contract can generate Medicare savings of about \$2,100 per admission, or about 5% of total payments. More aggressive payment reforms can generate substantially greater savings, but the accompanying reduction in LTCH profits has potential out-of-sample consequences. Our results highlight how improved financial incentives may be able to reduce healthcare spending, without negative consequences for industry profits or patient health.

JEL classification numbers: D22, I11, L21

Keywords: Healthcare, post-acute care, financial incentives, nonlinear contracts

*We thank our discussant Mark Shepard for helpful comments. We are grateful to Yunan Ji, Abby Ostriker, and Yin Wei Soon for excellent research assistance, and to Jeremy Kahn, Hannah Wunsch, and numerous seminar participants for their comments. Einav and Finkelstein gratefully acknowledge support from the NIA (R01 AG032449). Mahoney acknowledges support from the Becker Friedman Institute at the University of Chicago.

[†]Einav: Department of Economics, Stanford University, and NBER, leinav@stanford.edu; Finkelstein: Department of Economics, MIT, and NBER, afink@mit.edu; Mahoney: Booth School of Business, University of Chicago, and NBER, Neale.Mahoney@chicagobooth.edu.

1 Introduction

Healthcare spending is one of the largest fiscal challenges facing the U.S. federal government. In 2014, the U.S. federal government spent \$1.1 trillion on public healthcare programs (BEA, 2015) and the CBO projects that spending will grow to \$2.0 trillion by 2026 (CBO, 2016). Within the healthcare system, post-acute care (PAC) is an under-studied sector, with large stakes for both spending and patient health.

Post-acute care is the term used for formal care provided to help patients recover from a surgery or some other acute care event. Medicare spending on PAC is substantial, about \$60 billion in 2013, or about 20% more than the much-studied Medicare Part D program. PAC spending is growing faster than overall Medicare spending. PAC spending accounts for almost three-quarters of the unexplained geographic variation in Medicare spending (Newhouse et al., 2013). The quality of post-acute care has potentially important consequences for patient health; over 40% of hospital patients are discharged to PAC, and PAC patients are high-risk, with 15% of Medicare deaths involving a PAC stay in the prior 30 days.¹

In this paper, we study the impact of provider financial incentives in determining patient flows and government spending in the Medicare PAC system. The PAC setting is attractive for several reasons. First, given its fiscal importance, understanding the effects of financial incentives is a natural area for inquiry. Second, the institutional setting – involving multiple interlocking and potentially substitutable settings that operate under different reimbursement regimes – suggests that financial incentives may have first order consequences. Third, inefficiencies in the PAC sector have potentially important implications for public health, given that PAC patients are disproportionately high-risk and might be more vulnerable to inefficiencies in the delivery of care.

Our analysis focuses on patients whose point of entry into the PAC system is an LTCH.² We focus on LTCH patients because of the sharp variation in provider incentives at this type of facility. This is illustrated in Figure 1: providers are reimbursed a daily amount (of approximately \$1,300) up to a threshold number of days, at which point there is a large (approximately \$13,000 on average) increase in payments for keeping a patient an additional day beyond the threshold, but no payments for any days beyond it. We investigate the effects of this “jump” in payments at the threshold using detailed Medicare claims and administrative data on the universe of LTCH stays over the 2007-2012 period, when this non-linear payment schedule was in effect, as well as the 2000-2002 period, when LTCHs were instead reimbursed under a linear (i.e., constant per-diem) payment schedule.

We start by presenting descriptive evidence on the effect of the jump in payments. Discharges respond strongly to the payment increase, with the share of stays discharged increasing from 2% to 9% at precisely the day of the jump. The marginal patient discharged at the threshold appears to be much healthier than the average dischargee prior to the threshold. At the threshold, pa-

¹These statistics are taken from MedPAC (2004), MedPAC (2015a), and MedPAC (2015b), with the exception of the statistic on deaths which we calculate using the data described in Section 2.

²The acronym LTCH is typically pronounced “el-tack”, presumably reflecting the fact that LTCHs are sometimes referred to as long-term acute care hospitals (LTACs), which is pronounced in this manner.

tients are disproportionately more likely to be discharged to a less intensive PAC facility or home (“downstream”) than to an acute care hospital (“upstream”); they also have substantially lower post-discharge mortality than patients discharged on earlier days.

A natural question raised by this evidence is whether distortions in the timing of discharge have an impact on patient health. Given the high baseline mortality rate for LTCH patients (30% die within 90 days of LTCH admission), if the distortions are harmful, it seems plausible that we could detect an effect. Empirical analysis is challenging, however, because unlike discharge behavior, mortality effects may not appear right “at” the threshold.

The available evidence shows no compelling evidence of any mortality effects from the distortions in discharge behavior. We find no evidence of a change in the level or the slope of the mortality hazard in the vicinity of the threshold. We provide an additional informative contrast by comparing for-profit and non-profit LTCHs. We show that while both types of hospitals have similar payment schedules, the behavioral response in discharges is much stronger for for-profit hospitals. Despite the larger behavioral response at the threshold, we find no indication of a mortality impact even for the for-profit hospitals.

While these results provide no “smoking gun” evidence of patient harm, they do not allow us to comprehensively rule out a mortality effect. We cannot, for instance, rule out an effect for every type of patient or at each and every hospital, and we cannot rule out adverse health effects that would not manifest in higher mortality rates. However, at minimum, we do not detect robust enough mortality effect that would allow us to use the jump in payments to estimate how LTCHs trade off profits and patient health. Indeed, the results suggest that the marginal patients are able to receive similar care whether they are located in an LTCH or in their alternative setting, which empirically is usually a less intensive PAC facility such as a Skilled Nursing Facility (SNF).

Another natural question raised by the descriptive evidence on discharge behavior at the threshold is how providers respond to the payment schedule on days further from the threshold. Likewise, the descriptive analysis does not provide a natural way to gauge the magnitude of provider response, or to estimate how treatment patterns and Medicare payments would be affected by counterfactual payment schedules.

To address these questions, we specify and estimate a dynamic model of LTCH discharge behavior. In our model, patients are characterized by their health, which evolves stochastically over time. LTCHs face a (daily) decision of whether to retain the patient or discharge her to another facility. The LTCH’s objective function includes both net revenue (Medicare payments net of costs) and patient utility. If the patient is discharged from the LTCH, the provider receives zero net revenue, but internalizes at least a portion of the patient utility from being treated in an alternative location. If the LTCH keeps the patient, it receives net revenue that depends on Medicare’s payment schedule, while also accounting for the patient’s utility from being treated in the LTCH and the option value of making a similar discharge decision the following day. The provider therefore faces a standard dynamic discrete choice problem.

We estimate the model by simulated method of moments to match the observed discharge and mortality patterns under the linear and non-linear payment schedules. To take advantage of the

variation provided by the sharp jump in payments, we assign greater weights to moments that are close to the jump. The estimated model fits the data reasonably well.

We use the model and the estimated parameters to investigate the effects of alternative contracts that – like the observed contract – have a daily reimbursement rate up to a cap but that – unlike the observed contract – do not have a jump in payments at a threshold day. We find, for example, that if we were to lower the fixed payment to eliminate the jump in payments at the threshold, we would reduce total payments per admission for the episode of care by 25% on average, or about \$13,000 per admission. However, such a payment schedule substantially reduces LTCH revenue and estimated profits, and therefore may have out-of-sample impacts on LTCH behavior that our estimates would not capture, such as induced LTCH exit or lower service quality.

We therefore also engage in a more conservative set of counterfactuals in which we restrict attention to alternative contracts that would hold the LTCH harmless if their behavior did not change. Specifically, we consider the set of contracts that hold LTCHs revenue constant under their observed discharge schedule. Thus, if we apply this schedule and it triggers behavioral response by LTCHs then they must be better off. Using our estimated model, we are able to identify a broad set of “win-win” payment schedules that reduce Medicare payments and, by construction, leave LTCHs (weakly) better off.³ The contract that generates the largest savings reduces Medicare payments for the episode of care by 4.5%, and increases LTCH profits by 5.1%. Interestingly, this reduction in Medicare payments is slightly larger than the reduction from changes in organization form; our model estimates imply that if all LTCHs behaved like non-profits – rather than all LTCHs behaving like for-profits – Medicare payments per episode of care would be reduced by 2.9%.

Our paper relates to a large literature examining how healthcare spending responds to financial incentives. Given the importance of healthcare spending in the economy and in public sector budgets, the existence of this large literature is not surprising. What is surprising – and arguably unfortunate from this perspective – is that the vast majority of this literature (including much of our own work) has focused on the impact of *consumer* financial incentives, such as deductibles and co-payments, with relatively less attention to the impact of *provider* financial incentives.⁴ The majority of healthcare spending, however, is accounted for by a small share of high-cost individuals whose spending is largely in the “catastrophic” range where deductibles and co-payments no longer bind, and thus consumer cost-sharing is likely to have little impact relative to provider-side incentives.

The relative lack of research on the provider side presumably reflects the difficulties in finding clean variation in incentives to model and study. Perhaps not surprisingly, the sharp incentives created by the current LTCH payment schedule have already received some attention in both

³ Given the lack of compelling evidence of mortality effects at the threshold, it seems reasonable to assume that mortality impacts are also unlikely to be first-order under these “LTCH held harmless” alternative contracts.

⁴The literature on the impact of consumer incentives (“moral hazard”) in health insurance is too vast to try to summarize or cite here. Existing work on provider-side incentives has focused on descriptive evidence that providers do respond to incentives, with much of the evidence coming from the response to the introduction of the Inpatient Prospective Payment System in 1983 (Cutler and Zeckhauser (2000) provide a review of this evidence). More recently, Clemens and Gottlieb (2014) provide a rare look at the behavioral response of physicians to financial incentives.

academic (Kim et al., 2015) and popular (Weaver et al., 2015) spheres. Our descriptive work on discharges around the threshold is quite similar to this prior work, while our descriptive analysis of the health of the marginal discharger and our exploration of mortality effects is new. Our paper is most closely related to Eliason et al. (2016) who – in independent work – also study the impact of the LTCH payment schedule on discharge behavior descriptively and through the lens of a dynamic model. Where comparable, the results are broadly similar to ours; however they focus more on heterogeneity in provider responses and, unlike us, do not analyze patient mortality or post-discharge healthcare use.

Finally, from a more conceptual perspective, our paper is related to a growing literature that seeks to interpret descriptive evidence of the behavioral responses to non-linear payment schedules (“bunching”) through the lens of economic models that allow for assessments of behavior under counterfactual schedules (e.g., Chetty et al., 2011; Einav et al., 2015; Manoli and Weber, 2017).

The rest of the paper proceeds as follows. Section 2 provides some background on the PAC sector, LTCHs, and our data. In Section 3, we describe the discharge and mortality patterns around the jump in payments. Section 4 presents the model, discusses estimation, and presents the results from our counterfactuals. Section 5 concludes.

2 Setting and data

2.1 Post-acute care in the United States

Post-acute care (PAC) is the term for rehabilitation and palliative services provided to patients recovering from an acute care hospital stay. In the United States, the Center for Medicaid and Medicare Services (CMS) associates PAC with three types of facilities – long-term care hospitals (LTCHs), skilled nursing facilities (SNFs), and inpatient rehabilitation facilities (IRFs) – as well as care at home provided by home health agencies (HHAs) (MedPAC, 2015b). In 2013, Medicare paid \$60 billion to PAC providers, approximately 16% of the \$368 billion paid that year for Traditional Medicare (TM) claims (MedPAC, 2015a). This is about 20% larger than spending in Traditional Medicare on the much-studied Medicare Part D program.

In recent years, the geographic variation and growth rate of spending on PAC have raised concerns about the efficiency of the sector. From 2001 to 2013, Medicare spending on PAC grew at an annual rate of 6.1%, 2 percentage points higher than the rate of total spending growth for TM (The Boards of Trustees for Medicare, 2002 and 2014; MedPAC, 2015a). A recent Institute of Medicine report found that, despite accounting for only 16% of spending, PAC contributed to a striking 73% of the unexplained geographic variation in spending, suggesting that there may be substantial inefficiencies in the sector (Newhouse et al., 2013).

It is useful to think about patients generally flowing “downstream” through the healthcare system. Upon experiencing an acute health event, they enter a regular Acute Care Hospital (ACH); from there they may enter a PAC facility to recover, and eventually go home once they are sufficiently healthy and can function independently. Some ACH patients “skip” the PAC stay and

return home directly from the ACH, and some patients occasionally relapse and move “upstream” from a PAC facility back to an ACH.

The top panel of Figure 2 gives a sense of transitions among ACHs, PAC facilities (LTCHs, SNFs, and IRFs), home (including HHAs), and death (including hospice). (Throughout the rest of the paper, we use the term PAC facilities to refer to LTCHs, SNFs, and IRFs, because these are facilities that provide in-house care, in contrast to HHAs, which provide care at the patient’s home.) In our data, described below, 26% of patients who are discharged from an ACH receive follow-up care from a PAC facility.⁵ From these PAC facilities, 60% of patients continue to flow home, where they may still receive treatment from an HHA, while 33% are discharged back to an ACH. The remaining 7% of discharges are to a hospice or due to death.

Just like the natural flow of patients into and out of the PAC system, there is also a general ordering of care within it. LTCHs provide the most intensive care, SNFs and IRFs provide less intensive care, and HHAs provide the least intensive bundle of medical services. Severity of Illness (SOI) categories are a commonly used measure of intensity of care, and are constructed using the patient’s age, diagnoses, procedures, and comorbidities. The share of patients in the highest severity of illness category declines from 43% at LTCHs, to approximately 12% at SNFs and IRFs, to 4% at HHAs (AHA, 2010). Medicare payments per day follow the same declining pattern.

Our point of entry into the PAC landscape is through admission to an LTCH. The bottom panel of Figure 2 looks at patient flows from LTCHs. About 11% of LTCH patients are discharged back to an ACH, 38% are discharged to another PAC facility (SNF or IRF), and 33% are discharged to their homes, where they may continue to receive care from an HHA. The remaining 18% are discharged to a hospice (4%) or die within the LTCH (14%). In contrast, once in a SNF or IRF, patients almost never get discharged to an LTCH, die much less frequently (5%), and much more often (60%) return directly home.

Despite the interlocking nature of the PAC system, the way that Medicare reimburses post-acute care varies substantially by the setting. Historically, all providers were paid according to an administrative estimate of their costs. Since the early 2000s, however, many PAC stays are paid under a prospective payment system (PPS), yet the unit of payment varies across sites. Loosely, HHAs are paid per 60-day episode-of-care, SNFs are paid a fixed rate per day of stay, while IRFs and LTCHs are in principle paid a fixed amount per admission (like ACHs).⁶ We provide more details on LTCH payments in Section 3.

The fact that each type of facility is paid under a different system has raised concerns. From a public health perspective, there is concern that the separate payment systems do not give providers

⁵In analysis that includes HHAs in the calculation, the share of ACH patients who are discharged to PAC rises to 42% (MedPAC, 2015b).

⁶These different payment systems also have differential implications for beneficiaries’ cost-sharing requirements across types of PAC settings. Beneficiaries generally are not required to make any cost sharing for HHA services. IRF and LTCH stays are tied to the beneficiaries’ inpatient deductible, so when patients arrive from an ACH there would typically be no requirement for additional cost sharing. SNF stays are associated with separate SNF coinsurance for stays longer than 20 days.

enough incentive to coordinate care across different facilities. From a budgetary perspective, there is concern that providers may shuffle patients across facilities with the aim of increasing Medicare payments. These concerns have spurred various proposals for payment reform, including a recent bill which proposes providing a “bundled payment” to a single PAC coordinator, and letting this coordinator internalize the costs and benefits associated with the sequence of admissions and discharges for the entire episode of care (H.R.1458 – BACPAC Act of 2015).

2.2 Long-term care hospitals

Our primary focus is on patients whose point of entry into the PAC system is a long-term care hospital (LTCH). The demarcation “LTCH” describes how the provider gets paid by Medicare. It is a regulatory concept, rather than a medical one. For a hospital to get paid as an LTCH, it must have an average inpatient length of stay of 25 days or more. Naturally, there are many ways to meet this requirement, so from a medical standpoint the question of what exactly is an LTCH often results in differentiated or fuzzy answers.

The LTCH category of hospitals was created to solve a potential problem created by the 1982 Tax Equity and Fiscal Responsibility Act (TEFRA), which established the prospective payment system (PPS) for acute care hospitals. Under the new PPS, hospitals were paid per discharge, and not based on their costs, as a way to create incentives for hospitals to be efficient in their treatment decisions. Regulators who were designing the PPS realized that there was a small number of hospitals that had long average length-of-stays (LOS) and would not be financially viable under the fixed-price PPS. LTCHs were thus created as a carve-out from PPS for hospitals that had an average LOS of at least 25 days. At that point in time, there were 40 hospitals that qualified as LTCHs – mainly former tuberculosis and chronic disease hospitals in the Boston, New York City, and Philadelphia metropolitan areas. LTCH payments were based on costs measured in 1982, roughly in the spirit of the pre-1982 payment system, and adjusted for inflation in subsequent years. See Liu et al. (2001) for more on the background of the LTCH sector.

Over the last 30 years, and perhaps because of the LTCH exemption from PPS, there was rapid growth in the LTCH sector. Because new entrants did not have cost data for 1982, payments for new entrants were determined by costs in their initial years of operation. This encouraged new entrants to be inefficient when they first opened and earn profits by increasing their efficiency over time (Liu et al., 2001). From the initial 40 hospitals first designated as LTCHs in 1982, there are now over 400 such hospitals in the country.

Geographic penetration of LTCHs is extremely varied. This presumably reflects their historical roots as tuberculosis and chronic disease hospitals in the northeast, as well as certificate of need (CON) laws that restrict entry. There are only a few LTCHs in the west of the country, and three states (Massachusetts, Texas, and Louisiana) account for a third of all LTCHs. In places where there are LTCHs, these hospitals are an important part of Medicare’s PAC landscape. For instance, in hospital service areas (HSAs) with at least one LTCH, we calculate that LTCHs account for 13% of Medicare PAC facility days and 28% of Medicare PAC facility spending; nationwide, payments

to LTCHs account for 12% of Medicare PAC facility spending.⁷

LTCHs are much more likely to be for-profit than other medical providers. According to 2008 data from the American Hospital Association (AHA), 72% of LTCHs are for-profit (versus 17% for ACHs), 22% are non-profit, and 6% are government run. The LTCH market is dominated by two for-profit companies, Kindred Health Systems and Select Medical, which run about 40% of LTCHs, according to the AHA data. Company reports indicate that LTCHs are highly profitable. For their business segments that include LTCHs, Kindred’s profits have hovered between 22% and 29% of revenue and Select’s profits have ranged between 16% to 22% of revenue.⁸

Approximately half of LTCHs are known as Hospitals-within-Hospitals (HwHs), meaning that they are physically located within the building or campus of an ACH but have a separate governing body and medical staff. Regardless of an LTCH’s location (co-located or freestanding), they tend to have strong relationships with a single ACH.⁹ Because of concerns over close relationships between LTCHs and their partner ACHs, in 2005 CMS established a policy known as the “25-percent rule” that creates disincentives for admitting more than 25% of patients from a single facility; however, Congress has delayed the full implementation of the law.¹⁰

2.3 Data

Our main analysis focuses on patients who are admitted to an LTCH and follows them throughout their entire healthcare episode. Our primary data source is the Medicare Provider and Analysis Review (MedPAR) data, spanning the years 2000-2012. The dataset contains claim-level information on discharges from ACHs, LTCHs, SNFs, and IRFs. Each record is a unique stay for which a claim was submitted, and the data contain information on procedures, admission and discharge dates, admission sources and discharge destinations, hospital charges, and Medicare payments. The MedPAR data also provide us with basic demographic information such as the age, sex, and race of the beneficiary, and information about the patient’s diagnoses (DRG).

We supplement this primary source with several ancillary data sources. First, we use Medicare’s beneficiary summary file on cost and use it to approximate post-LTCH discharge payments for hospice and for HHAs, as well as post-LTCH discharge hospice days; Appendix A provides more

⁷Authors’ calculations using 2007-2012 data described below. PAC facilities include LTCH, SNF and IRF. PAC facilities represent about 70% of total PAC spending, with the other 30% due to PAC provided by HHAs (MedPAC, 2015a).

⁸Profits are defined as EBITA (earnings before interest, taxes, and amortization). Kindred’s profits are based on 2009 to 2015 company reports. Prior to 2009, Kindred did not separate out their reporting of LTCH profits from the much larger SNF category. Select’s profits are based on company reports from 2004 to 2015.

⁹MedPAC (2004) found that HwHs receive 61% of their cases from their most frequent referring hospital and freestanding hospitals receive 42% from their most frequent referring hospital.

¹⁰There is also a regulation known as the “5-percent rule” that addresses the incentive for HwH to “ping-pong” patients between the ACH and LTCH. In particular, if more than 5% of patients who are discharged from an LTCH to an ACH are readmitted to the LTCH, the LTCH will be compensated as if the patient had a single LTCH stay (42 CFR 412.532).

details.¹¹ Second, we use Medicare’s beneficiary files to determine whether the beneficiary is dually eligible for Medicare and Medicaid and the date of death (if any) through 2012. Third, we use the Medicare chronic conditions file to measure whether the individual has any of 27 chronic conditions in the calendar year prior to the LTCH stay. Finally, we use data from the American Hospital Association (AHA) survey over the same period to determine whether an LTCH is for-profit, non-profit, or government owned, and whether it is co-located with an ACH.

Our analysis focuses on the current Medicare payment schedule for LTCHs, known as LTCH-PPS. We analyze the time periods before and after full implementation of LTCH-PPS, which was phased in over a five year period starting on October 1, 2002. We define the *pre-PPS period* as discharges that occurred from January 1, 2000 to September 30, 2002. For this period, we measure post-discharge payments, days and mortality through March 31, 2003, which is six months after the last LTCH discharge. We exclude the October 2002 to September 2007 phase-in period because provider behavior during this period potentially reflects the combination of changing financial incentives and learning about the new incentive structure, complicating the interpretation of the data. We define the *PPS period* as discharges that occurred from October 1, 2007 to July 31, 2012, and analyze post-discharge payments, days and mortality through December 31, 2012, which is similarly six months after the last LTCH discharge.

Table 1 shows summary statistics on ACH, LTCH, and SNF/IRF admissions in the pre-PPS and PPS periods.¹² Since an observation is an admission, some patients (16%) show up multiple times in the data. LTCH patients are, on average, slightly younger than ACH patients and much younger than SNF/IRF patients. LTCH patients are also almost twice as likely to be black and about one-third more likely to be eligible for Medicaid, relative to ACH and SNF/IRF patients. These differences are fairly stable over time. From a health perspective, patients in an LTCH look less healthy than those in an ACH or SNF/IRF. They have more chronic conditions prior to the stay and higher mortality. For example, about 15% of LTCH patients die within 30 days of admission and 30% die within 90 days; these mortality rates are about 50% larger than mortality rates for SNF/IRF patients and about twice as large as those for ACH patients.

In terms of the intensity of medical care, LTCH stays are closer to ACH stays than stays at a SNF/IRF. The majority of LTCH and ACH patients receive at least one medical procedure versus about 5% of patients who visit an SNF/IRF. The most common LTCH procedures (blood transfusion and ventilation) are also more similar to those that occur at an ACH, relative to occupational and physical therapies, which are the most common procedures in SNF/IRF. Length of stay at an LTCH, however, is (by design) much more similar to that of a SNF/IRF. The average

¹¹In the MedPAR data, we can observe all discharge destinations, but we can only observe post-discharge payments and days for ACH and PAC facilities (SNF, IRF, LTCH) and not for home health visits or hospice. Appendix A shows how we estimate these using the summary file on cost and use which includes spending for all components of care. In practice, HHA and hospice payments are quite small as a share of the total. For example, we estimate that of individuals with an LTCH discharge, LTCH and SNF/IRF payments constitute over 90% of total PAC payments, with home health accounting for only 8%.

¹²We group SNF and IRF admissions together for convenience, as both represent post-acute care that is “less intense” than an LTCH and because IRFs only account for a small (6.4%) fraction of these admissions.

stay at an ACH is 5 days, while it is just over 25 days in LTCH and SNF/IRF.

The bottom rows of Table 1 show statistics on Medicare and out-of-pocket payments. Medicare payments in the PPS period average \$2,079 per day at an ACH, \$1,392 per day at an LTCH, and \$507 per day at a SNF/IRF. However, because LTCH stays are much longer than ACH stays, per-admission Medicare payments at LTCHs average over \$35,000, which is three times greater than per-admission ACH and SNF/IRF payments. Out-of-pocket payments at ACHs and LTCHs arise from Medicare’s Part A deductible (\$1,156 in 2012) and from co-insurance payments that apply when the patient has more than 60 hospital days in the benefit period (\$289 per day in 2012). Because patients have no out-of-pocket exposure between the deductible and their 60th hospital day, out-of-pocket payments are a modest 7.7% of Medicare payments at ACHs and 5.4% at LTCHs in the PPS period. SNFs, on the other hand, have a separate co-insurance schedule with payments of \$144.50 per day in 2012 for stays in excess of 20 days, and have a much higher out-of-pocket share.

Our analysis encompasses not only the experience of the patient in the LTCH (i.e., length of stay and payments) but also their post-discharge experience. Table 2 provides some summary statistics on post-discharge experience. We define a post-discharge episode of care as the spell of continuous days with a Medicare payment to an ACH, SNF/IRF or LTCH; the episode ends if there are two days or more without any Medicare payments being made to any of these institutions. For each post-discharge episode we report 30-day mortality, 90-day mortality, post-discharge Medicare payments, and post-discharge facility days (i.e. days in an ACH, SNF/IRF, LTCH or hospice). Focusing on the PPS period, about one-quarter of LTCH patients die in the 90-days post discharge. Average length of stay in the post-discharge episode of care is 26 days, which is similar to the average time in the LTCH (see Table 1). Average post-discharge Medicare payments is \$22,975, about 60% of Medicare payments to the LTCH (see Table 1).

In some of our analyses below, we will find it useful to classify live discharges from the LTCH as either “upstream” or “downstream” based on their discharge destination. “Upstream” discharges represent patients in worse health than “downstream” destinations. Specifically, we group LTCH discharges to hospice or ACH as “upstream” and discharges to SNF/IRF, or home (with or without home healthcare), or to other as “downstream”.¹³ Table 2 shows that most (about 85%) of LTCH discharges are downstream, and that patients initially discharged downstream have substantially lower post-discharge mortality (e.g. 18% 90-day mortality compared to 60% for upstream discharges), length of stay, and payments.

¹³Appendix Table A1 shows more granularity on the discharge destinations within upstream and downstream. In the PPS period 97% of patients discharged upstream are sent to ACH (vs hospice); of patients discharged downstream, about half are sent to SNF/IRF (which we group together because over 90% of discharges to SNF or IRF are to SNF) and another 40% are discharged to home or home health care.

3 LTCH payments, discharge patterns, and outcomes

In this section we present descriptive analysis on LTCHs’ response to financial incentives. We start by describing the LTCH budget set, including the large jump in payments that is our primary source of identification. We then show evidence on how discharge patterns and mortality rates vary with the budget set. This descriptive evidence motivates several of our key modeling choices in the dynamic model of LTCH discharge behavior, which we present in the next section.

3.1 LTCH payments

We provide a basic overview here of how LTCH payments vary with the patient’s length of stay, an object we refer to as the LTCH budget set or payment schedule. Appendix B provides more details. Figure 1 summarizes the payment schedules in the pre-PPS and PPS periods.

Prior to October 1, 2002, LTCHs were paid their (estimated) daily cost, generating a linear relationship between the length of the hospital stay and payments. As described earlier, this “cost plus” reimbursement of LTCHs was seen as potentially encouraging inefficient entry into the LTCH market. Because of this and other concerns, the 1997 Balanced Budget Act (BBA) and 1999 Balanced Budget Refinement Act (BBRA) implemented a PPS for LTCHs. LTCH-PPS was phased in over a 5-year period starting on October 1, 2002 and was fully implemented by October 1, 2007. At a broad level, LTCH-PPS is designed to operate like the PPS for acute care hospitals (IP-PPS), under which hospitals are paid a lump-sum that is based on the patient’s diagnosis (diagnosis-related group, or DRG) and does not vary with the patient’s length of stay.

Much like LTCHs were originally created to address a potential problem with the introduction of PPS for ACHs, the features of the LTCH-PPS payment schedule can similarly be thought of as attempting to address a potential problem arising from the introduction of PPS for LTCHs. In particular, in designing LTCH-PPS, officials were concerned that LTCHs might discharge patients after a small number of days but still receive large lump-sum payments intended for longer hospital stays. To address this concern, they created a short stay outlier (SSO) threshold. If stays were shorter than the SSO threshold, payments would be based on the pre-PPS cost-based reimbursement schedule and LTCHs would not receive a large lump sum. However, while reducing the incentive to cycle patients in and out of the LTCH, the SSO system creates potentially problematic incentives at the SSO threshold. At the day when payments switch from per-day reimbursement to lump-sum prospective payment amount, Medicare payments for keeping a patient an additional day “jump” by a large amount.

Figure 1 graphs the average payment schedules in the pre-PPS and PPS periods, pooling across LTCH facilities and DRGs. The y-axis shows cumulative Medicare payments, inflation-adjusted to 2012 dollars. The x-axis shows the length of the stay relative to the SSO threshold, which we normalize to be day 0. The SSO threshold is defined as five-sixths the geometric mean length of stay for that DRG in the previous year and therefore varies by DRG (and also, to a much lesser extent, by year). The average threshold is at 22.5 days; the modal threshold (accounting for 22.4%

of PPS stays) is 20 days and 99% of the sample has a SSO threshold between 16 and 39 days.¹⁴ As a result, in this and subsequent figures, we present results relative to the SSO threshold so that we can pool analyses across DRGs.¹⁵ Because the SSO threshold is undefined in the pre-PPS period, we assign pre-PPS stays the threshold for their DRG from PPS period.¹⁶

Under the pre-PPS system, average payments scale linearly with the length of stay at a rate of \$1,071 per day. Under the PPS system, payments increase linearly by \$1,386 per day to the left of the SSO threshold, “jump” by \$13,656 at the SSO threshold, and remain constant thereafter. The increase in payments at the jump is large: it is equal to 55% of the cumulative payment amount on the day prior to the threshold, or equivalent to about 10 days of payments at the pre-threshold daily rate.

This sharp jump in payments was presumably not the intention of the policymakers who designed the LTCH-PPS, but it arises naturally from the interaction of two sensible policies. As is standard in fixed price contracts, the LTCH-PPS payments were likely set to approximate average costs per stay. As noted, payments on a cost-plus basis up to the SSO threshold were introduced to avoid paying LTCHs large lump sum amounts for relatively short stays. While reducing the incentive to discharge patients from an ACH to an LTCH, the transition unavoidably creates potentially problematic incentives. Particularly concerning is where the threshold was set: we estimate that under the pre-PPS payment scheme, 44% of stays would have been below the subsequent short stay outlier threshold, which is a large fraction for a policy that is at least ostensibly designed to target “outlier” events.

In Section 4 we explore the impact of alternative, counterfactual payment schedules. To eliminate the jump in payments, our counterfactuals alter the payment prior to the SSO threshold (so that it does not approximate per-day costs), alter the fixed PPS amount (so that it does not approximate average costs), or alter both segments of the payment schedule.

3.2 Discharge patterns

We present a number of descriptive results on discharge patterns from the LTCH around the threshold. The results can be summarized as follows. First, there is a large spike in discharges at precisely the day of the jump in payments, indicating a strong response to financial incentives. Second, the marginal patients discharged at the threshold are in relatively better health: they are disproportionately discharged “downstream” and they have lower post-discharge mortality rates than patients discharged at other times. Third, among patients discharged downstream, the marginal patients discharged at the threshold are relatively sicker, with higher post-discharge payments

¹⁴ Across all DRGs, the SSO threshold ranges from 14 to 56 days.

¹⁵ We start the x-axis range at -15 days because nearly all SSO thresholds occur after 16 days. If we extended the x-axis range to -16, for example, there would be a change in the composition of DRGs between days -16 and -15 due to the entry of new DRGs into the sample. We end the x-axis range at 45 days because there are relatively few (2.1%) patients who are kept at the LTCH more than 45 days beyond the SSO threshold.

¹⁶ Since the thresholds occasionally vary over time, we use the SSO threshold from 2007, which is the first year in the PPS period.

than pre-threshold discharges. Below we discuss how these patterns can be rationalized with a simple model of LTCH behavior.

Figure 3 shows the aggregate pattern of discharges by length of stay in the pre-PPS and PPS periods. A discharge occurs when the patient is transferred to another facility, sent home, or dies at the LTCH. The y-axis shows discharges as a share of the total number of stays at the LTCH. The x-axis plots the length of stay relative to the DRG-specific SSO threshold, defined in the same manner as in Figure 1. In the PPS period, there is a sharp increase in discharges at the SSO threshold, with the share of discharges increasing from about 2% to 9% per day. Discharge rates remain elevated over the subsequent 7 to 10 days before reverting to baseline. In the pre-PPS period, there is no evidence of any bunching at the SSO threshold. This discharge pattern is consistent with a strong response to the financial incentives.

Relative to the pre-PPS constant, per-day payment schedule, the increase in discharges under PPS at the threshold could be drawn either “from the left” or “from the right” of the distribution. In other words, the excess discharges at the threshold could reflect patients who under the pre-PPS schedule would have been discharged before the threshold but are “retained” in order to get the lump sum payout at the threshold, or patients who would have been discharged after the threshold but are now discharged earlier since there is no longer a marginal financial payment associated with keeping them for additional days.

Because the share of discharges to the left of the threshold is lower in the PPS period relative to the pre-PPS period, it is tempting to infer that the excess mass of discharges is primarily “drawn from the left” of the distribution. However, we caution that differences in the discharge rate might not only reflect the change in financial incentives but also changes in patient health and other secular trends between the pre-PPS and PPS periods. For example, a simple reweighting of the pre-PPS admissions to match the DRG composition of the PPS period eliminates approximately 30% of the difference between the pre-PPS and PPS discharge rates prior to the SSO threshold (Appendix Figure A2). In Section 4, we show how we can use our model to compare discharge patterns under the observed PPS payment schedule to discharge patterns under alternative, counterfactual payment schedules.

Figure 4 decomposes the discharge pattern by the location of discharge in order of descending patient health: downstream, upstream, and death. The figure shows increases at the threshold in discharges both upstream and downstream, but the proportional increase is substantially larger on the downstream margin. Moreover, because the pre-threshold discharge rate is much higher downstream, the sharp change in discharge rate at the threshold (shown in Figure 3) is almost entirely driven by downstream discharges. We defer our discussion of the bottom panel on mortality to the subsection below.

Appendix Figure A3 plots the 30-day post-discharge mortality rate, defined as death within 30 days of a (live) discharge, by length of stay. The graph shows a sharp drop in post-discharge mortality at the SSO threshold, again suggesting that the patients who are discharged at the threshold are healthier than the patients who are discharged immediately beforehand. Of course, the decline in mortality not only reflects changes in the composition of patients discharged at the

threshold, but could in principle reflect a treatment effect of discharge on health. We address this in the next section.

Finally, Figure 5 plots Medicare payments for the episode of care that occurs after the LTCH discharge, by length of stay at the LTCH. We show these post-discharge payments separately for patients discharged “upstream” and discharged “downstream” and view them as a proxy for the patient’s health at the time of discharge. For patients discharged downstream, there is a sharp increase in post-discharge payments at the threshold, with average post-discharge payments increasing from approximately \$10,000 to \$20,000. There is a small change in the opposite direction for patients initially discharged upstream. For longer lengths of stay, the figure becomes noisy due to the small number of discharges.

Figure 5 suggests a simple model of LTCH behavior, which motivates the model we present in Section 4. Prior to the threshold, retaining patients is profitable, and only the healthiest patients are discharged to SNF/IRF or their home and the sickest patients discharged to an ACH or a hospice. After the threshold, on the downstream margin, LTCHs work “down the distribution” and discharge less healthy patients, increasing post-discharge payments on average. Similarly, on the upstream margin, LTCHs work “up the distribution,” discharging patients who are in better health, and decreasing post-discharge payments on average. The marginal patient discharged downstream at the threshold is therefore sicker than the average patient discharged prior to the threshold, while the marginal patient discharged upstream is slightly healthier than the average patient discharged upstream in prior days.

3.3 (Lack of) Mortality effects

A natural question raised by the discharge patterns is whether the distortions in the timing of discharges have an impact on patient health and in particular mortality. Since the 90-day mortality rate of LTCH patients is approximately 30% (see Table 1), if these distortions are harmful to health, it seems plausible that we might be able to pick up an effect with our data.

Empirical identification of mortality effects from the distortion in patient location at the threshold is challenging, however. Health evolves according to a stochastic process, with sicker patients having a higher probability of death. Distortions to the location of care might impact the level of someone’s health, generating an on-impact effect on the probability of death analogous to the on-impact effect on discharges we detected. However, distortions to the location might also affect the stochastic process for health, which would be associated with a longer-run change in mortality rate, but might not have an immediate mortality effect. We therefore attempt to examine not only whether there is an immediate impact on mortality at the threshold, but whether we can detect any longer-run changes.

We have already seen in the bottom panel of Figure 4 some suggestive evidence that mortality rates are declining over the course of the LTCH stay with little difference around the SSO threshold.¹⁷ However, the interpretation of the bottom panel of Figure 4 – which plots mortality

¹⁷The downward trend results from “natural selection”: as the sickest patients die, the remaining patients become

rates for LTCH patients by length of stay – is complicated by selection concerns. Since LTCHs are differentially discharging healthier patients at the SSO threshold, the composition of patients who remain at the LTCH is changing, making it tricky to disentangle any potential treatment effects on mortality from effects due to changes in the selection of LTCH patients.

In Figure 6 we circumvent this issue by taking advantage of the fact that our data allow us to track mortality outcomes for patients even after their LTCH discharge. Conceptually, our mortality analysis follows the logic of a reduced form regression, where the mortality hazard is the outcome, discharge patterns are the endogenous variable, and the SSO threshold is the instrument. In particular, since we know there is a sharp jump in discharge patterns at the threshold (analogous to a large first stage), if there is a change in the level or slope of the mortality hazard at the threshold (that is, non-zero reduced form), we can infer that the distortion in discharge location has an impact on mortality.

The top panel of Figure 6 is thus similar to the bottom panel of Figure 4, but uses the full set of LTCH patients (unconditional on their location) rather than only those who have yet to be discharged. As before, natural selection leads mortality rates to decline over time, but we now can interpret more cleanly the mortality pattern around the SSO threshold. The plot shows no obvious evidence of a change in the level of mortality hazard in the vicinity of the threshold during the PPS period. In Appendix C, we examine this mortality pattern more formally using a regression discontinuity design and similarly fail to reject the null of a smooth mortality hazard around the SSO threshold. These findings are consistent with no mortality effect but do not allow us to rule out a gradual effect that would not appear sharply in the data.

If distortions in the location of care affected the stochastic process for health, we might not observe an immediate effect, but would see a change in mortality over a longer time horizon. The bottom panel of Figure 6 attempts to look for a more gradual effect by plotting a 30-day mortality rate, by days since LTCH admission, where the 30-day mortality hazard measures the share of patients who are alive on a given day but die during the subsequent 30 days. The plot once again shows no effect around the threshold, suggesting that there are no gradual effects of the distortion in location on mortality. In Appendix C, we present a regression discontinuity analysis that more formally confirms this result.

Figure 6 suggests little evidence of a quantitatively large effect on mortality that is created by the sharp discharge incentives at the SSO threshold. This (lack of) reduced form effect should be judged in relation to the size of the first stage effect on the location of care. And as we will show in the analysis in the next section, the jump in payments at the threshold only shifts the location of care for a relatively small number of days, so perhaps the null result is not surprising.

Finally, we provide some additional traction on potential mortality impacts by exploiting differences in the response to incentives between for-profit and non-profit LTCHs. Figure 7 shows the payment schedules (top panel), discharge shares (middle panel), and mortality rates (bottom panel), separately by for-profit status. While the payment schedules are almost the same across

gradually healthier.

the two groups of hospitals, the behavioral response to the jump in payments is substantially larger at for-profit LTCHs. Prior to the threshold, for-profit hospitals have a slightly lower discharge rate, but at the SSO threshold their discharge share rises by about twice as much as that of non-profit hospitals, suggesting that, perhaps not surprisingly, for-profit hospitals are relatively more responsive to financial incentives.

Given that for-profits LTCHs exhibit much larger behavioral response to the jump in payments, if the PPS payment schedule affected mortality, we would expect a more pronounced mortality effect at for-profit hospitals. The bottom panel of Figure 7 provides no suggestion of any differences in the mortality pattern by for-profit status. Mortality hazards in the PPS period are remarkably similar across both groups in the immediate vicinity of the SSO threshold; this suggests no on-impact effect of the location distortion on mortality.¹⁸

While these results provide no “smoking gun” evidence of patient harm, they do not allow us to comprehensively rule out a mortality effect. We cannot, for instance, rule out an effect for every type of patient or at each and every hospital, and we cannot rule out adverse health effects that would not manifest in higher mortality rates. However, combined with the earlier results which indicate that the patients who are most affected by the SSO threshold are disproportionately healthy, the results suggest that the marginal patients affected by the PPS payment schedule are able to receive similar care whether they are located in an LTCH or in their alternative setting, which empirically is usually a SNF.

4 Quantifying the importance of financial incentives

The results in the last section provide descriptive evidence of the response of LTCHs to the sharp financial incentives associated with the SSO threshold. One way to quantify the importance of the financial incentives in determining LTCH discharge patterns is to assess how these patterns would change in response to counterfactual financial contracts that exhibit weaker incentives. Doing so requires a dynamic model, which is the focus of this section.

4.1 A simple model of LTCH discharge decisions

Consider a patient i who is admitted at day $t = 0$ to LTCH l . We index patient i 's health at the time of admission by $h_{i,0}$, and assume that $h_{i,t}$ (conditional on patient i staying at LTCH l) evolves stochastically from day to day. In our baseline specification, we assume that $h_{i,t}$ follows a monotone Markov process, such that $h_{i,t} \sim F(\cdot|h_{i,t-1})$ with $F(\cdot|h)$ stochastically increasing in h .¹⁹ We use higher values of h to indicate better health and thus assume that daily mortality hazard $m(h)$ is strictly decreasing in h .

¹⁸The mortality pattern for non-profit hospitals is slightly noisier, but this is presumably driven by sampling variation due to the smaller samples size of non-profit admissions (113,154 versus 487,988 for for-profits in the PPS period).

¹⁹In sensitivity analysis, we examine the robustness to our findings by allowing the health process to vary with time since admission.

Hospital l 's flow (daily) payoff from keeping patient i (whose health is given by h) during the t th day since admission is given by

$$u_l(h, t|keep) = p(t) - c_l(h) + \beta w_l(h) + \sigma_\epsilon \epsilon_{ilt}, \quad (1)$$

where $p(t)$ is the hospital's revenue, which depends on CMS' reimbursement schedule for patient i , $c_l(h)$ is the hospital's daily cost of treating a patient with health index h , and the third term captures the patient's utility from staying at LTCH l , $w_l(h)$, multiplied by the hospital's weight on it β . Finally, ϵ_{ilt} is an error term, which is distributed i.i.d. type I extreme value and scaled by the parameter σ_ϵ . The error term presumably captures idiosyncratic considerations associated with the patient and/or the hospital.

Our focus is on the hospital's discharge decision. Consider a set of J alternative destinations for patient i , each indexed by j . Conditional on discharging the patient to destination j , LTCH l 's revenue and cost are both zero, and its flow payoffs are given by the patient's utility, again multiplied by the hospital's weight on it β , and a logit error, scaled by σ_ϵ :

$$u_j(h, t|discharge\ to\ j) = \beta w_j(h) + \sigma_\epsilon \epsilon_{ijt}. \quad (2)$$

Moreover, because hospital l loses control over the patient upon discharge, it will be convenient to denote by $W_j(h)$ the present value to hospital l of the patient's utility from being discharged to alternative j .

This setting lends itself to a simple dynamic programming problem, which can be represented by the following Bellman equation:

$$W_l(h, t) = E \left(\max \left\{ \begin{array}{l} u_l(h, t|keep) + \delta (1 - m(h)) \int W_l(h', t + 1) dF(h'|h), \\ \max_{j \in J} (u_j(h, t|discharge\ to\ j) + \delta (1 - m(h)) \int W_j(h') dG_j(h'|h)) \end{array} \right\} \right), \quad (3)$$

where δ is the LTCH's (daily) discount factor. That is, the two state variables are the health of the patient and the number of days since LTCH admission. Every day the hospital makes a binary decision whether to discharge or keep the patient, and in the event of a discharge the hospital also decides about the discharge location. Of course the model can allow the patient to actually "decide" about the discharge destination by having the hospital place a large weight on patient utility. While we did not find a mortality effect in our descriptive analysis, by allowing the health process outside the LTCH to evolve according to $G_j(\cdot|h)$, instead of $F(\cdot|h)$ within the LTCH, our model allows patient health to evolve differentially across alternative locations of care

It is convenient to benchmark the patient's utility against her utility at the LTCH. That is, we normalize $w_l(h) = 0$ for all h , $v_j(h) = w_j(h) - w_l(h)$, and normalize $V_j(h)$ accordingly. Applying these adjustments and using the well-known expression for the logit's inclusive value, we can write the problem as

$$V_l(h, t) = \sigma_\epsilon \ln \left\{ \exp \left(p(t) - c_l(h) + \delta (1 - m(h)) \int V_l(h', t + 1) dF(h'|h) \right) + \sum_{j \in J} \exp(V_j(h)) \right\}. \quad (4)$$

Finally, we note that the state variable t only affects the problem through the hospital revenue function $p(t)$, and $p(t) = 0$ for all $t > SSO$, so the problem becomes stationary after the SSO threshold, and the solution is simply a fixed point of

$$V_l^{t>SSO}(h) = \sigma_\epsilon \ln \left\{ \exp \left(-c_l(h) + \delta (1 - m(h)) \int V_l^{t>SSO}(h') dF(h'|h) \right) + \sum_{j \in J} \exp(V_j(h)) \right\}. \quad (5)$$

We can therefore solve for the dynamic problem by first solving for the fixed point associated with the post-SSO stationary part of the problem given by equation (5), and then iterating backwards until $t = 0$ using equation (4).

4.2 Parameterization, estimation, and identification

Parameterization. We make several additional assumptions before we take the model to the data. First, we restrict the set of alternative discharge destinations J to include only two options, $J = \{D, U\}$. Motivated by the summary statistics described in Section 3, option D covers the collection of downstream destinations (SNF, IRF, and home with or without HHA) that are appropriate for LTCH patients who are of better health or require lower levels of medical monitoring. In contrast, option U covers upstream discharge destinations (ACH and hospice) which would be natural discharge destinations for patients who are of worse health.

Because, conditional on discharge, financial incentives do not affect the LTCH's discharge destination, having a richer set of discharge options is unlikely to affect our counterfactual predictions. By focusing on two options, we essentially restrict the LTCH to consider two types of marginal LTCH patients. One set of marginal patient are those who are healthier, and for whom the hospital must consider whether to keep them or discharge them to location D . The second set of marginal patients are sicker, and for whom the hospital must consider whether to keep them in the LTCH or transfer them to location U .

The second assumption is about the health process. Given that mortality is monotone in h , it is convenient to normalize the health index by mortality risk. We do so by assuming that h is defined by its associated mortality hazard using the following relationship:

$$m(h) = 1 - \Phi(h), \quad (6)$$

where $\Phi(\cdot)$ is the standard normal CDF. We note that h is an index and thus has no cardinal meaning, and the above is simply a normalization, which endows h with a cardinal measure. Equipped with this normalization, we then make parametric assumptions about the initial (as of LTCH admission, $t = 0$) health distribution of newly admitted patients, and about how the health process evolves over time within the LTCH. Specifically, we assume that $h_{i,0}$ is drawn from $N(\mu_0, \sigma_0^2)$ and that $F(\cdot|h_{i,t-1})$ follows a simple AR(1) process:

$$h_{i,t} = \mu + \rho h_{i,t-1} + \varepsilon_{i,t}, \text{ where } \varepsilon_{i,t} \sim N(0, \sigma^2). \quad (7)$$

In our baseline specification, we allow the health process to be different inside the LTCH in the pre-PPS and PPS periods to accommodate potential differences in the LTCH patient mix, which may result from the growth of the LTCH sector, time trends in medical technology and practice, or directly from the change in financial incentives. Note that we do not need to model the health process outside the LTCH since any effect of the post-discharge location on health would affect the discharge decisions through its (unidentified) effect on the continuation values, $V_j(h)$, which we do not model explicitly.

The third assumption is associated with the LTCH’s costs, $c_l(h)$, which we assume are given by

$$c_l(h) = \gamma c^{reported}. \quad (8)$$

That is, as described in Section 2 and reported in the bottom row of Table 1, we observe the cost associated with each hospital l , which enters the formula by which it is paid by CMS. We use the averages reported there (separately for the pre-PPS and PPS periods) as our estimates of $c^{reported}$. We do not treat these reported costs as the “true” costs, but we use it to guide our model of costs in two ways. First, and importantly, we assume that the hospital’s cost do not vary with the patient health h , which is consistent with CMS’s treatment of costs and also seems natural given that LTCH patients are generally stable. Second, we assume that the reported costs are true up to a monotone transformation, which we assume to be linear. This assumption means that if hospital l has higher reported costs than hospital k , we will also assume that this also reflects the ranking of their true underlying costs. This seems (to us) reasonable, and could be driven by a variety of factors, including geographic location. We expect $\gamma \leq 1$, but do not impose this constraint in estimation.

The fourth assumption is to parameterize $V_U(h)$ and $V_D(h)$. We approximate each using a linear function in h , so that

$$V_j(h) = v_{0j} + v_{1j}h \text{ for } j = D, U. \quad (9)$$

Recall from Section 3 that healthier patients (higher h), who are associated with lower mortality, are discharged to D , while sicker patients (lower h), associated with higher mortality, are discharged to U . It is therefore natural to expect $v_{1,D} > 0$ and $v_{1,U} < 0$. That is, all else equal, facility D becomes a more attractive discharge destination as health gets better (h is higher) and facility U becomes a more attractive discharge destination as patients’ health worsens (h is lower). As explained below, one of the intercept terms $v_{0,D}$ and $v_{0,U}$ needs to be normalized, so we set $v_{0,U} = 0$.

Fifth, as we discussed in Section 2, LTCHs are part of an interlocking post-acute care system, with changes in LTCH incentives potentially affecting Medicare spending throughout the patients’ entire episode of care. In particular, Figure 5 shows sharp changes in both upstream and downstream post-discharge payments at the SSO threshold, indicating a relationship between patients’ health at discharge and total Medicare spending. To account for this type of spillover effects in our counterfactuals, we model the relationship between health at discharge and post-discharge payments as

$$P_j(h) = \exp(\zeta_{0,j} + \zeta_{1,j}h) \text{ for } j = D, U, \quad (10)$$

where $P_j(h)$ represents the post-discharge payments for a patient initially discharged to location $j = D, U$ with health status h at the time of discharge. We allow this relationship to vary by whether the patient is initially discharged upstream or downstream and use an exponential specification so that predicted post-discharge payments are strictly positive.

Finally, as is typical in these types of models, we set (rather than estimate) the daily discount factor to $\delta = 0.95^{1/365}$.²⁰ Thus, overall we are left with 19 parameters to estimate: five parameters $(\mu_0, \sigma_0, \mu, \sigma, \rho)$ that are associated with the health distribution and the way it evolves over time in the pre-PPS period, five corresponding parameters in the PPS period, the cost parameters (γ) , four parameters $(v_{1U}, v_{0D}, v_{1D}, \sigma_\epsilon)$ associated with the relative value of patients at facilities U and D , and four parameters associated with post-discharge payments in the PPS period $(\zeta_{0,D}, \zeta_{1,D}, \zeta_{0,U}, \zeta_{1,U})$.

Estimation. An important decision is how to treat heterogeneity across patients, observable health conditions, and LTCH hospitals. In our baseline specification, we abstract from such heterogeneity and instead model the “average” discharge decision as it pertains to the “average” LTCH patient and the “average” payment schedule. That is, we pool all payment schedules observed in the data, separately for the pre-PPS and PPS periods, measure each day in the schedule relative to the DRG-specific SSO threshold in the PPS period (which is normalized to day zero), and construct the average payment schedule for each day, as shown in Figure 1. We then apply an analogous exercise to the discharge pattern, and construct the distribution of discharge patterns in a 61-day window around the SSO threshold (from day -15 to day 45), as shown in Figure 3 and Figure 4. We then estimate our model in an attempt to match these average patterns. The dramatic difference in the payment schedules between the pre-PPS and PPS periods will assist in the identification of some of the parameters and it is an important component of our research design.

An advantage of this approach of focusing on the average pattern rather than the heterogeneous pattern is that it only requires us to solve the dynamic problem once (for each period, pre-PPS and PPS), which is computationally attractive. Given that our primary focus is on the aggregate effect of financial incentives across the entire LTCH sector, abstracting from the heterogeneity across patients and hospitals is likely to be inconsequential. That said, motivated by the descriptive evidence in how for-profit and non-profit LTCHs differentially respond to financial incentives, we separately conduct a counterfactual analysis to explore the heterogeneous effects of financial incentives for these two types of hospitals, which reveals some intriguing patterns. Heterogeneity in the response is also the focus of the related exercise reported by Eliason et al. (2016).

We estimate the model using simulated method of moments, by trying to match the daily mortality and discharge patterns presented in Figure 4, as well as post-discharge payment moments that are based on Figure 5. Specifically, we use two sets of moments. First, we use 183 moments for the pre-PPS payment schedule, reflecting the daily discharge and mortality risks within the 61-day window around the SSO threshold. One set of moments is associated with discharge rates

²⁰While having a discount factor so close to 1 might generally create convergence issues, in our Bellman equation the probability of survival enters jointly with the discount factor so that the “effective discount factor” $\delta(1 - m(h))$ is not too close to 1.

to D , another with discharge rates to U , and a third with mortality rates. We then construct another set of 183 corresponding moments for the PPS period. Because much of the identification is driven by the sharp change to discharge incentives at the SSO threshold, we assign greater weights to moments that are closer to day zero (the SSO threshold) by making the weights decrease by a constant amount (1/61) per day away from the threshold. The second set of moments uses the data on post-discharge payments to allow us to capture spillover effects in our counterfactuals. Specifically, we average post-discharge payments for each discharge destination (U or D), separately for discharges before and after the SSO threshold. We then match them to the model prediction regarding the health status distribution of discharged patients, thus allowing us to link health and post-discharge payments.²¹

Generating the model predictions requires us to solve the dynamic problem described in the previous section for each set of parameters. To ease with computation, we approximate the health process $F(\cdot|h_{i,t-1})$ with a discrete health space that evolves according to a Markov transition matrix (Tauchen, 1986).²² This eases the solution of the dynamic problem, and at the same time allows us to read the discharge probabilities directly off the solution, without any need to integrate (presumably by simulation) over potentially intractable integration regions.

Intuition for Identification. To see the intuition for the identification of the parameters, it is easiest to consider first a case where health is homogeneous across patients and over time. Under this assumption, the data can be characterized by daily observations of discharge shares to U and to D (s_{Ut} and s_{Dt} , respectively), with the remaining patients staying at the LTCH ($s_{lt} = 1 - s_{Ut} - s_{Dt}$). The problem resembles a repeated static discrete choice problem, where the mean utility of each discharge destination is given by the continuation values V_{Ut} , V_{Dt} , and V_{lt} . As is usual in multinomial logit models, the observed daily shares can then be inverted to recover the values of V_{Ut} , V_{Dt} , and V_{lt} , subject to a required level and scale normalizations.

Let us start with the level normalization, and consider discharge shares after the SSO threshold, where the model is in a steady state. Setting $V_{Ut} = 0$ allows us to estimate V_D from the average values of V_{Dt} and V_{lt} , up to a scale normalization. To identify the scale parameter σ_ϵ , consider the discharge shares the day before the SSO threshold: the jump in payments introduces a sharp change in the continuation values, which identifies the effect of payments on V 's, or equivalently identifies σ_ϵ . Finally, the cost parameter γ is identified off the non-stationarity of the model prior to the SSO threshold, and the way discharge patterns are changing as the SSO threshold gets closer. As the SSO threshold nears, the expected (remaining) length of stay of patients shortens, making the expected life-time costs associated with a given patient gradually decline in a way that is not proportional to expected payments, allowing us to separately identify the impact of cost. Loosely, V_{lt} varies over time as a function of the payment schedule $p(t)$, reported costs scaled by

²¹Because there is no variation in payments in the pre-PPS period, we do not have the variation in discharges' health that we need to identify the pre-PPS post-discharge payments model. Since we focus on the PPS period in our counterfactuals, and therefore do not need these estimates, we do not estimate pre-PPS post discharge payments.

²²In particular, we approximate the health distribution with a grid of 250 evenly spaced values that span a range of +/- three standard deviations around the mean of the steady state health process.

parameter γ , and the relevant time horizon, which depends on the mortality rate and subsequent, endogenous discharge decisions. So one can think of the identification of the scale parameter σ_ϵ and γ as a projection of the values of V_{it} on these observables. The sharp change in payments at the SSO threshold provides a sharp change in the present value of payments and identifies the scale parameter (or equivalently the coefficient on payments when the variance of the error term is standardized), and the differential change in the present value of payments versus costs as the patient approaches the SSO threshold identifies γ . This identification can be achieved from the PPS moments alone, but given that we restrict these parameters to be time-invariant, it is also aided by variation in discharge patterns between the pre-PPS and PPS periods.

If health status h were observed, we could make the argument above conditional on health, and thus identify each object as a non-parametric function of h . In practice h is unobserved, but identifying the health process is conceptually easy given our assumptions. If there are no discharges, which is roughly the case during the first week or so of the LTCH stay, the only attrition from the sample is due to mortality. With only five parameters that determine the initial health distribution and how it evolves from day to day, mortality rates over five days are sufficient to identify the health process parameters, separately in the pre-PPS and PPS periods. Once the unobserved health distribution is identified, we can integrate over h and apply a similar intuition to the one we described above for the homogenous h case. Moreover, once the health process is identified, the cross-sectional distribution of h varies over time in “known” ways, so we can also identify how the key parameters – in particular the V ’s – vary as a function of h . Finally, the relationship between the health status h and the post-discharge payments is directly identified off the data given that the model does not impose any restrictions on this relationship.

Finally, the parameters of the post-discharge payments model ($\zeta_{0,D}, \zeta_{1,D}, \zeta_{0,U}, \zeta_{1,U}$) are identified by the sharp change in health of patients discharged on different sides of the SSO threshold. That is, they are essentially mapping the model’s prediction regarding the average health status of discharges before and after the SSO threshold to the corresponding average of observed post-discharge payments, which also exhibit a sharp change around the SSO threshold (as shown in Figure 5).

Obviously, as is typically the case, the intuition for identification requires us to have substantial variation in the data. In practice, some of the variation is not as large, and statistical power issues require us to impose more parametric structure, so the estimable model is not as flexible – especially in terms of the extent to which parameters vary with h – as the identifiable structure would be.

4.3 Parameter estimates and model fit

Table 3 presents the parameter estimates. We estimate $\gamma = 0.78$ implying that LTCH’s actual costs are 22% lower than their reported value. This is consistent with our prior that reported costs are inflated.

The $v_{1,U}$, $v_{0,D}$, and $v_{1,D}$ parameters capture the value the LTCH places on the patient’s utility from being discharged to U or D relative to remaining at the LTCH, as well as any potential effect

on patient health evolution in the discharge location (relative to remaining at the LTCH). The estimates imply that LTCHs are indifferent between U and D for a patient with $h = 2.0$, which is a fairly low health level. For instance, $h = 2.0$ is the 4.2th percentile of the steady state PPS health (normal) distribution (with estimated parameters $\mu = 6.2$ and $\sigma = 2.4$) and corresponds to a daily mortality hazard of 2.1%. Consistent with our description of patients flowing “downstream” as their health improves, D is relatively better for healthier patients and U is better for sicker patients. The magnitude of the slope parameter $v_{1,D}$ is about one-fifth as large (in absolute value) as the slope parameter $v_{1,U}$, which indicates that a given change in financial incentives will have a much larger effect on discharges on the downstream D margin. These estimates are consistent with the descriptive evidence that shows a substantially larger response on the downstream margin at the SSO threshold.

Relative to the slope parameters $v_{1,U}$ and $v_{1,D}$, the scale parameter σ_ϵ on the logit error is fairly small. The estimates imply that a tenth of a standard deviation increase in the error term increases the value of discharging a patient to a given location by \$71, the product of $\sigma_\epsilon = 550$ and a tenth of a standard deviation of the logit error ($\pi/\sqrt{6}$). In contrast, a tenth of a standard deviation increase in steady state health index ($\sigma = 2.4$) raises the value of discharging a patient downstream by \$1,523 ($= 0.1 \times 2.4 \times 6,330$) and lowers the value of discharging a patient upstream by \$7,274 ($= 0.1 \times 2.4 \times 30,230$), indicating that health status is capturing most of the unobserved heterogeneity in discharge behavior.

The $\zeta_{0,U}$, $\zeta_{1,U}$, $\zeta_{0,D}$, and $\zeta_{1,D}$ parameters capture the relationship between health at discharge and post-discharge payments in the PPS period. Consistent with our interpretation of Figure 5, the estimates indicate that post-discharge payments are declining as the patient gets healthier, with semi-elasticities of $\zeta_{1,U} = -0.79$ and $\zeta_{1,D} = -0.27$ on the upstream and downstream margins, respectively.

We are cautious not to over-interpret the change between the pre-PPS and PPS periods in the health process parameters. Because they are the only parameters that are allowed to vary across the time periods, they capture not only differences in the health of admitted patients but may also reflect other factors that vary over time, such as changes in medical technology or the administrative capacity of providers.

The model fits the data reasonably well. Figure 8 presents our discharge moments and the simulated moments from the estimated model. The left column shows values in the PPS period and the right column shows values in the pre-PPS period. The top row shows the share of discharges to U by day relative to the SSO threshold, the middle row shows the share of discharges to D , and the bottom row shows the share of patients who die at the LTCH. The model does a very good job fitting the “spike” in discharges to U and D in the PPS period. This is particularly encouraging because this variation is our key source of identifying variation. The model fit for the mortality patterns in the pre-PPS and PPS periods is good over the initial days, but less good at longer time horizons. This is likely due to our fairly parsimonious parameterization of the health process. The model fit is also poorer for discharges to U in the pre-PPS period.

Figure 9 provides some intuition for how the model operates, illustrating the “policy function”

at the estimated parameters.²³ Healthy patients (higher h) are discharged to D , while sick patients (lower h) are discharged to U . Consistent with the descriptive evidence, LTCHs work “down the distribution” at the jump and lower their discharge threshold on the D margin and conversely work “up the distribution” on the U margin and increase the discharge threshold. The larger shift on the D margin relative to the U margin relates directly to our discussion above on the magnitude of the slope parameter estimates ($v_{1,D}$ and $v_{1,U}$). The relatively small outward shift in the policy function just before the SSO threshold is consistent with the descriptive results which show limited evidence on “missing mass” immediately to the left of the SSO threshold.

4.4 The effects of counterfactual financial incentives

We use our model to simulate discharge patterns and Medicare payments under a variety of counterfactual payment schedules. Throughout this section, we assume that the initial distribution of health of admitted patients stays the same but that the subsequent discharge decisions reflect the incentives provided by the counterfactual payment schedules.

We limit our attention to alternative schedules that maintain the current practice of a cap on payments after a certain number of days. We do this both because it respects the current policy approach toward LTCH payments and because an “uncapped” schedule would lead to a small number of long hospital stays, which is outside of the variation in our data. Specifically, we consider three main types of counterfactuals: payment schedules that remove the jump while holding the threshold day constant, payment schedules that eliminate the jump (and allow the threshold day to vary) while holding LTCHs harmless, and cost-based reimbursement at a constant per diem, capped at 60 days. We discuss each in turn below.

Removing the jump. We start by considering two simple modifications of the baseline payment schedule that eliminate the jump in payments at the SSO threshold, but, like the baseline PPS payment schedule, provide no payments on the margin for stays in excess of the SSO threshold. Figure 10 plots these counterfactual payment schedules and the baseline schedule for comparison. The top panel shows a counterfactual we call “higher rate per day,” which eliminates the jump by increasing the per-diem payment from \$1,384 to \$2,151 prior to the SSO threshold but holds the post-threshold payment fixed. The bottom panel shows a counterfactual schedule we call “lower cap,” which eliminates the jump in payments at the SSO threshold by reducing the PPS payment from \$34,419 to \$22,144 but holds the pre-SSO per diem payment fixed. The “higher rate per day” contract is weakly more generous than the baseline schedule, while the “lower cap” contract is weakly less generous.

We use our model to simulate discharge patterns and Medicare payments under these two counterfactuals. Figure 11 shows the policy functions under each payment schedule, Figure 12 shows the discharge patterns, and Table 4 summarizes the impact of each of these payments schedules on

²³The policy function is not a deterministic function of h ; given the ϵ 's in the LTCH's keep and discharge flow payoff functions (see equations 1 and 2), h is related to discharge stochastically. The policy lines in Figure 9 are drawn so that at that given level of h , 50% of the patients are discharged to the relevant destination.

Medicare payments to the LTCH and to other facilities.

The black dashed lines in Figure 11 show the policy function under the “higher rate per day” counterfactual. During the first few days, the policy function is similar to that under the observed schedule. However, as the length-of-stay increases, the elimination of the jump reduces the incentive to retain patients, and the policy function shifts inwards on the upstream and downstream margins. The black dashed lines in Figure 12 shows discharge patterns under this counterfactual. Mirroring the changes in the policy function, the “higher rate per day” counterfactual increases discharges in the 10 days prior to the SSO threshold, relative to the observed schedule. As shown in column (2) of Table 4, these higher discharge rates reduce average length of stay by 1.9 days or 7.2%.²⁴

Despite the significantly higher per-day payments prior to the jump (\$2,151 versus \$1,384), Medicare payments to LTCHs increase by a very small amount of \$369 or 1%. The small increase is due to a large behavioral response to the incentives. Holding discharge patterns fixed, LTCHs would get paid about 50% more per day for stays below the SSO threshold, but in response to the elimination of the jump, patients are now discharged earlier, so overall payments are lower. Holding discharge patterns fixed, we calculate that the mechanical effect of this counterfactual is a \$1,767 increase in Medicare payments to the LTCH, implying that the behavioral response to the removal of the jump reduces Medicare payments to the LTCH by \$1,398 per admission. LTCH profits per admission rise by about \$2,483 or 37% relative to the observed schedule.

We next explore the effects of the “lower cap” payment schedule. The grey lines in Figure 11 show the policy function under this counterfactual. The elimination of the jump in payments shifts the policy function inwards during the entire pre-threshold period, relative to that under the observed schedule. The grey lines in Figure 12 show that discharges correspondingly rise, with the daily share of discharges to D increasing four-fold and the share of discharges to U increasing more modestly over most of the pre-threshold period. As shown in column (3) of Table 4, average length of stay is reduced by 4.5 days, and payments to the LTCH are reduced by \$11,954 or 43%. The mechanical effect (holding observed discharge patterns fixed) of the “lower cap” payment schedule is a reduction in payments of \$8,844 or about 74% of the overall reduction, with the remaining 26% due to the behavioral response. LTCH profits per admission fall by over \$6,954 per admission and are estimated to be negative, a point we return to below.

The remaining rows of Table 4 consider the impact of these counterfactual payment schedules on Medicare payments throughout the rest of the episode of care. For these counterfactuals, the spillovers on post-discharge payments are small. For the “higher rate per day” counterfactual, post-discharge payments for patients discharged U and D are affected by a few hundred dollars. For the “lower cap” counterfactuals, the decrease in post discharge payments is larger but still small compared to decrease in LTCH payments. Combining the effects on LTCH payments and post-discharge payments, our estimates indicate that the “higher rate per day” has virtually no effect on total Medicare payments (\$70 increase) and the “lower cap” reduces total Medicare payments

²⁴Length of stay is measured from day -15. However, because the average SSO threshold across admissions in our sample is 22.5 days, values for length of stay need to be increased by 7.5 (=22.5-15) days to match the summary data.

by a substantial \$12,467 or 26%.

While interesting, neither of the above counterfactuals is fully satisfactory. While the “lower cap” counterfactual suggests that alternative payment schedules could substantially reduce Medicare payments, the large decrease in LTCH revenue (and in estimated profits) might have unintended consequences. For instance, under this payment schedule, LTCHs might cut back on socially valuable fixed investments or even exit the market. In contrast, the “higher rate per day” counterfactual, while clearly making LTCHs better off, has virtually no effect on Medicare payments. Yet, these two exercises suggest that there might be “intermediate” contracts that generate cost savings without the risk of unintended consequences. We explore such counterfactuals below.

“Win-win” payment schedules. With the above considerations in mind, we now consider a set of counterfactuals that hold LTCH revenue fixed under the observed discharge patterns. Faced with these contracts, if LTCHs do not change their behavior, they will have identical revenue, costs, and profits as they would under the observed payment schedule. If the LTCH change their behavior, by revealed preferences, they must be (weakly) better off. Therefore, by design, these contracts should not have a negative impact on LTCHs.

In the same spirit as the previous counterfactuals, we consider contracts that pay a constant per-diem amount up to a threshold length of stay, at which point the payments are capped and per diem payments drop to zero (obviously, with no jump). We consider contracts where the payment is capped at thresholds in a +/- 10 day range on either side of the current SSO threshold day. Since the generosity of the contract is strictly increasing in the per diem rate, for a given length of stay at which payments are capped, there is a unique payment schedule that holds LTCH revenue fixed under the baseline discharge patterns.

Figure 13 plots a set of these contracts, where to avoid overcrowding the figure, we show payment schedules for only a subset of the 21 contracts considered. The top panel of Figure 14 plots LTCH payments against total Medicare payments (including estimated post-discharge payments) for each of the contracts we consider. The bottom panel of Figure 14 plots estimated LTCH profits against Medicare payments for each contract. For comparison, both plots also show outcomes under the observed payment schedule.

The figures indicate that there is a broad set of “win-win” payment schedules that reduce total Medicare payments for the episode of care while leaving LTCHs (weakly) better off. Every counterfactual contract with a threshold between -8 and 8 days reduces Medicare spending, although there is substantial heterogeneity in the reduction. LTCH revenues increase for every contract with a threshold of 6 to 10 days and declines for contracts with a threshold of -10 to 5 days. Because LTCHs value both profits and patient utility, LTCH profits under counterfactuals do not necessarily increase. LTCH profits are higher than their baseline level for contracts with a threshold -10 to 2 days and are lower for thresholds of 3 to 10 days. Counterfactuals that decrease profits do not lower them by a substantial amount.

The counterfactual with payment threshold of 1 day more than the current SSO threshold results

in the largest reduction in Medicare spending and is a natural contract to focus on.²⁵ Column (4) of Table 4 shows outcomes for this contract. Under this payment schedule, Medicare payments to LTCHs are reduced by \$1,660 or 5.9%. Accounting for Medicare payments across the entire episode of care leads to somewhat higher savings of \$2,145, or 4.5% of total episode payments. Despite the reduced payments, LTCH profits rise by \$344 per stay or 5.1%: the decline in LTCH revenue is offset by lower costs, as length of stay is almost 2 days (9%) shorter.

Cost-based payment schedules. The last two columns of Table 4 report results from two additional counterfactuals that explore cost-based reimbursement as a constant per diem rate. The first counterfactual pays LTCHs a constant per diem of their estimated costs, which is \$1,107 per day. The second counterfactual pays LTCHs a constant per diem of \$507, which is the average per day payment to SNFs during the post-discharge period. We think about this counterfactual as a form of “reference based pricing” where Medicare pays LTCHs the opportunity cost to Medicare of the patient – i.e., the amount Medicare would have incurred for the patient at a location that provides fairly similar care (at least for the marginal patient). To avoid extrapolating too far outside of our data, for both of these counterfactuals we cap payments after 60 days, which leads LTCHs to discharge virtually all of their patients within 90 days of the current SSO threshold.

Paying LTCHs their estimated costs leads to a substantial \$7,010 increase in payments to LTCHs and a smaller \$3,530 overall increase in total Medicare payments. Payments increase because LTCH retain patients for longer time periods rather than discharging them to SNFs where Medicare payments would be lower, with average length of stay increasing from 19.2 to 32.4 days. Paying LTCHs the average per-diem for SNFs leads to a massive decrease in LTCH payments and total Medicare payments, accompanied by a sharp reduction in length of stay. Of course, concerns about unintended consequences, which we discussed in the context of the “lower cap” schedule, are also relevant here.

For-profit versus non-profit LTCHs. At the end of Section 3, we presented descriptive evidence where we split the sample by whether the LTCH was a for-profit or non-profit hospital. In particular, we showed that while payment schedules were similar for for-profit and non-profit LTCHs, for-profits exhibited a larger increase in discharges at the SSO threshold, suggesting that these facilities are more responsive to financial incentives. Our model allows us to quantify the magnitude of the differences in for-profit and non-profit behavior, and to adjust for potential differences in the health of the patients admitted to these facilities to more precisely isolate heterogeneity in their response to incentives.

To conduct this analysis, we estimate the model allowing all of the parameters except those governing post-discharge payments to vary by for-profit status. We force the post-discharge payments parameters to be the same because, conditional on h at discharge, post-discharge payments should not depend on whether the patient was discharged from a for-profit or non-profit hospital. In addition, requiring the parameters to be equal helps ensure that the health process has the same

²⁵By contrast, we achieve the highest Medicare dissavings with the counterfactual with the payment threshold of -10 days, where Medicare spending rises by over 2%.

cardinal interpretation across these different types of LTCHs.

Appendix Table A4 shows the parameter estimates. The health process parameters are broadly similar across the two types of hospitals. The point estimate for costs is somewhat lower for for-profit ($\gamma = 0.81$) than non-profit ($\gamma = 0.91$) hospitals, which is consistent with higher operational efficiency by the private sector. The point estimates on the slope parameters ($v_{1,U}$ and $v_{1,D}$) are also smaller in absolute value for for-profits than non-profits, indicating that for-profits are more elastic to financial incentives in their discharge decisions. We note, however, that the non-profit parameters are less precisely estimated so that the differences are only sometimes statistically significant.

Comparing for-profits and non-profits parameter by parameter is not particularly insightful. Therefore, to quantify the difference between for- and non-profits over the full set of parameters, Table 5 conducts the counterfactual of assigning every LTCH for-profit versus non-profit status. For reference, column (1) shows outcomes under the observed organizational form. To construct column (2), we assign all LTCHs the γ , $v_{1,U}$, $v_{0,D}$, $v_{1,D}$, and σ_ϵ estimates for for-profit hospitals but maintain the heterogeneous health process parameters so that there is “no change” in the health distribution of the patients. For column (3), we assign all LTCHs the non-profit estimates for γ , $v_{1,U}$, $v_{0,D}$, $v_{1,D}$, and σ_ϵ and hold the health process fixed.

Consistent with the increased responsiveness to incentives, payments to LTCHs are \$458 (or about 1.7%) higher under the “all for-profit” than the “all non-profit” counterfactual. We estimate that LTCH profits are more than twice as large (\$5,681 versus \$2,590) under the counterfactual where they are all for-profit. Interestingly, the \$458 higher payments only accounts for about one-fifth of the higher profits, with the majority of the difference being due to lower estimated costs. Accounting for post-discharge payments, the estimates imply that the counterfactual of “converting” all LTCHs from for-profit to non-profit status would reduce total Medicare payments across the entire episode of care by \$1,376 or about 2.9% of their current level, holding the distribution of patient health fixed.

Robustness. In our baseline model, we made a number of parametric assumptions. In order to assess the sensitivity of our main results to these assumptions, Table 6 reports the main results from a subset of the (many) alternative specifications that we examined. The results appear to be qualitatively robust.

In our first alternative specification, we relax the assumption that the health process is stationary by allowing the auto-correlation parameter ρ to vary with length-of-stay according to $\rho = \rho_0 + \rho_1 \ln(t + 1)$, where the time index is defined such that the patient is admitted on date $t = 0$.²⁶ While in the pre-PPS period ρ_1 is very close to zero, in the PPS period the estimate is slightly negative $\rho_1 = -0.005$, which is consistent with health becoming less stable over the course of the stay. However, as shown in Panel B of Table 6, enriching the specification in this manner has virtually no effect on the counterfactuals. We also specified other models of health processes, such as a random walk and a random walk with a drift, but the ability of these models to fit the data

²⁶We assume that the correlation parameter becomes fixed after 45 days at a value of $\rho = \rho_0 + \rho_1 \ln(46)$ so that the dynamic programming problem becomes stationary for $t > 45$, allowing us to solve the $t > 45$ problem by value function iteration and earlier periods by backwards induction.

was much worse than our baseline specification.

The second specification reported in Table 6 fits the model using only the PPS moments. The counterfactuals, shown in Panel C of Table 3, are very similar to the baseline estimates, suggesting that the sharp jump in payments at the SSO threshold, relative to over-time variation from the implementation of LTCH-PPS, is the key driver of the results. The limited importance of the over-time variation presumably stems from the fact that we allow distinct health process parameters in each period, thereby soaking up much of the over-time variation.

5 Conclusions

In this paper, we examined the impact of provider financial incentives in post-acute care (PAC), a setting with large stakes both for the government budget and patient health that has received scant attention in the academic literature. Within the context of PAC, we examined the impact of a jump in Medicare payments to long term care hospitals (LTCHs) that occurs after a pre-specified length of stay, when reimbursement shifts from a per diem rate to a lump sum payment. At the threshold, the payment for keeping an individual another day jumps by about \$13,000, with no marginal payment for additional days beyond the threshold.

The descriptive evidence shows a large response by LTCHs to the jump in payments. At the threshold, there is a large spike in discharges. The marginal patient affected by the payment threshold is relatively healthy. We are unable to detect any impact on patient mortality at the threshold, even in this high-mortality population.

To examine the implications of alternative payment schedules, we specified and estimated a stylized dynamic model of LTCH discharge behavior which allows us to model LTCH behavior, payments, and profits, as well as total Medicare payments for the episode of care, under alternative payment schedules. We used the estimated model to search for “win-win” contracts that hold LTCHs (and presumably its patients as well) harmless, while reducing Medicare payments. The contract with the largest Medicare savings reduced total Medicare payments by 5% while increasing LTCH profits by a similar percentage.

We also considered more aggressive payment schedules that resulted in substantially higher Medicare savings but raised the possibility of unintended consequences due to a large reduction in LTCH profits. In particular, in our model we take admission to the LTCH as given and focus on the impact of counterfactual policies on length of stay and payments to the LTCH and post-discharge costs for this fixed set of patients. However, the large reduction in profits brought about in our more aggressive counterfactuals may affect which patients are admitted to an LTCH and might have even broader effects on the market, for instance through entry and exit decisions. We consider an important area for further work to move “up” the healthcare pathway and model the decision to discharge from an acute hospital stay to an LTCH or other PAC facilities.

More broadly, our results indicate how economic models and data can be combined to better inform contract design. Little other than common sense was needed to see that *some* alternative payment schedule should be better for both the Medicare payer and the LTCH. Data and descriptive

evidence, however, were important to move beyond this conceptual “existence proof” to demonstrate a quantitatively meaningful behavioral response by the LTCH to the current payment schedule. An economic model and its estimates were necessary to identify specific contracts that could create a “win-win” for the LTCH and Medicare payments given the LTCH’s optimizing behavioral response to alternative contracts. While naturally our results are specific to our particular setting, we hope that this type of approach can inform future work examining the impact of provider financial incentives not only for the directly affected provider but throughout the healthcare system.

References

Adda, Jerome, and Russell W. Cooper. *Dynamic Economics: Quantitative Methods and Applications*. MIT press, 2003.

American Hospital Association (AHA). 2010. “TrendWatch: Maximizing the Value of Post-Acute Care.” Available at <http://www.aha.org/research/reports/tw/10nov-tw-postacute.pdf>.

Boards of Trustees for Medicare. 2002. “The 2002 Annual Report of the Boards of Trustees of the Federal Hospital Insurance Trust Fund and the Federal Supplementary Medical Insurance Trust Fund.”

Boards of Trustees for Medicare. 2014. “The 2014 Annual Report of the Boards of Trustees of the Federal Hospital Insurance Trust Fund and the Federal Supplementary Medical Insurance Trust Fund.”

Bureau of Economic Analysis (BEA). 2016. “Download NIPA Tables: Government Current Receipts and Expenditures (Section 3).” National Economic Accounts. Available at <http://www.bea.gov/national/nipaweb/DownSS2.asp>. Accessed August 31, 2016.

Chetty, Raj, John Friedman, Tore Olsen, and Luigi Pistaferri. 2011. “Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records”. *Quarterly Journal of Economics* 126 (2): 749-804.

Clemens, Jeffrey and Joshua D. Gottlieb. 2014. “Do Physicians’ Financial Incentives Affect Medical Treatment and Patient Health?” *American Economic Review* 104(4): 1320-1349.

Congressional Budget Office (CBO). 2016. “Updated Budget Projections: 2016 to 2026,” Congressional Budget Office Publication #51384.

Cutler, David, and Richard Zeckhauser. 2000. “The Anatomy of Health Insurance.” *Handbook of Health Economics* Volume 1, Chapter 11: 563-643.

Einav, Liran, Amy Finkelstein, and Paul Schrimpf. 2015. “The Response of Drug Expenditure to Nonlinear Contract Design: Evidence from Medicare Part D.” *Quarterly Journal of Economics* 130(2): 841-899.

Einav, Liran, Amy Finkelstein, and Paul Schrimpf. forthcoming. “Bunching at the Kink: Implications for Spending Responses to Health Insurance Contracts.” *Journal of Public Economics*.

Eliason, Paul, Paul Grieco, Ryan McDevitt, and James Roberts. 2016. “Strategic Patient Discharge: The Case of Long-Term Care Hospitals”. NBER Working Paper 22598.

Kim, Yan, Eric Kleerup, Patricia Ganz, Ninez Ponce, Karl Lorenz, and Jack Needleman. 2015. “Medicare Payment Policy Creates Incentives for Long-Term Care Hospitals to Time Discharges for Maximum Reimbursement.” *Health Affairs* 34(6): 907-915.

Liu, Korbin, Cristina Baseggio, Douglas Wissoker, Stephanie Maxwell, Jennifer Haley, and Sharon Long. 2001. “Long-Term Care Hospitals Under Medicare: Facility- Level Characteristics.” *Health Care Financing Review* 23(2): 1-18.

Manoli, Dayanand, and Andrea Weber. 2017. “Nonparametric Evidence on the Effects of Financial Incentives on Retirement Decisions.” *American Economic Journal: Economic Policy*, forthcoming.

Medicare Payment Advisory Commission (MedPAC). 2004. “A Data Book: Health Care Spending and the Medicare Program.”

Medicare Payment Advisory Commission (MedPAC). 2015a. “A Data Book: Health Care Spending and the Medicare Program.”

Medicare Payment Advisory Commission (MedPAC). 2015b. “Report to the Congress: Medicare Payment Policy, Chapter 7: Medicare’s Post-Acute Care: Trends and Ways to Rationalize Payments.”

Newhouse, Joseph, Alan Garber, Robin Graham, Margaret McCoy, Michelle Mancher, and Ashna Kibria. 2013. “Variation in Health Care Spending: Target Decision Making, Not Geography.” Institute of Medicine. Available at http://www.nationalacademies.org/hmd/~ /media/Files/Report%20Files/2013/Geographic-Variation2/geovariation_rb.pdf

Tauchen, George. 1986. “Finite State Markov-Chain Approximations to Univariate and Vector Autoregressions.” *Economics Letters* 20(2): 177–181.

Weaver, Christopher, Anna Wilde Mathews, and Tom McGinty. 2015. “Hospital Discharges Rise at Lucrative Times.” *The Wall Street Journal*, February 17, 2015.

Appendix

Appendix A. Post-discharge payments and days

Our starting point of the analysis is an admission to an LTCH. We can observe all discharge destinations from an LTCH. Appendix Table A1 shows the share of discharges to different locations. In the PPS period, 14% of patients die during their LTCH stay, another 11% are discharged “upstream” (with the vast majority going to inpatient care) and 75% are discharged “downstream” (with approximately half of these patients going to a SNF/IRF and 40% going home, where they may receive care from an HHA).

We define a post-discharge “episode of care” as the spell of almost continuous days following discharge from an LTCH with Medicare payments to an ACH, SNF/IRF or LTCH. In particular, the episode ends if there are at least two days without Medicare payments to these institutions.

Although we can observe all discharge destinations within the episode of care, we only observe claim-level Medicare payments and days at ACHs, SNF/IRFs, and LTCHs. To address the fact that we do not observe payments or days at HHAs or hospices, we supplement the MedPAR data with annual spending and utilization from the Beneficiary Summary File (BSF) Cost & Utilization file. For every stay in the MedPAR data, we observe whether the patient was discharged to an HHA or hospice at some point in the episode of care. For patients who were discharged to an HHA or hospice, we impute the patient’s payments and days using the annual BSF data.

Since these annual amounts include some payments and days that occur before or after the “episode of care,” our imputation likely leads us to overestimate post-discharge Medicare payments and days. However, we think that the approach provides a reasonable approximation. Appendix Table A2 shows that our estimates of post-discharge payments and facility days are not affected much if we instead impute 0 costs for HHA and hospice, and 0 days for hospice.

Appendix B. LTCH payment systems

Prior to fiscal year 2003 (i.e. October 2002), CMS reimbursed LTCHs on a cost-based system. At the start of fiscal year 2003, CMS began transitioning LTCHs to a prospective payment system (PPS). The PPS, which was fully phased in by the start of fiscal year 2008 (i.e. October 2007), is the focus of our study. This appendix describes it in more detail, drawing heavily on Kim et al. (2015), Medicare Reimbursement Reference Guide (2015), and MedPAC (2014).

B.1. LTCH PPS rules

In contrast to the cost-based system, which had reimbursed hospitals based on the estimated cost of each patient’s case, the PPS outlined a fixed reimbursement amount for each patient, based on the patient’s DRG. These DRG-based lump-sum payments were meant to reflect the typical resources consumed by each type of patient. However, in order to discourage short stays in hospitals which

were meant to provide long-term care, the PPS includes a short stay outlier (SSO) threshold, with reduced payments below the full DRG payment for LTCH patients who are discharged before a DRG-specific threshold.

Full DRG payment

The full DRG payment is computed as

$$\text{Full DRG Payment} = \text{Adjusted Federal Rate} \times \text{DRG Relative Weight}, \quad (11)$$

where

$$\begin{aligned} \text{Adjusted Federal Rate} = & (\text{Unadj. Federal Rate} \times \text{Labor-Related Share} \times \text{Wage Index}) + \\ & + (\text{Unadj. Federal Rate} \times \text{Nonlabor-Related Share}). \end{aligned} \quad (12)$$

This payment structure is similar to the much-studied Inpatient PPS used for (regular) acute care hospitals (ACHs) that was introduced in 1983, but differs in two ways. First, although the DRGs are defined in the same way for the LTCH and Inpatient PPS, the relative weights associated with DRGs have different values in LTCH-PPS. Second, the LTCH-PPS unadjusted federal rate is larger than the corresponding Inpatient PPS value. The result is that LTCH-PPS payments are substantially greater than Inpatient PPS payments for the same DRG, presumably to reflect the greater costs at an LTCH relative to an ACH.²⁷

Short stay outlier (SSO) payment

If an LTCH stay has a length of stay (LOS) shorter than or equal to five-sixths of the geometric average length of stay (ALOS) for the DRG, it is paid as a short stay outlier. We call the smallest integer greater than five-sixth of the geometric ALOS the SSO Threshold. The SSO threshold is constant within a DRG-PPS Rate Year (with the exception of 2009).

A short stay outlier is paid the *lowest* of the following:

1. Full DRG Payment.
2. 120% of the DRG per diem amount times the length of stay, where the DRG per diem amount is defined as the ratio of the full DRG payment to the geometric average of the LOS for that

²⁷Also, like Inpatient PPS, LTCH PPS offers a High Cost Outlier (HCO) payment for particularly costly stays. Specifically, an LTCH can receive a HCO payment if the cost of the case exceeds the HCO Threshold. The HCO payment is made in addition to the regular payment amount. Importantly, for our purposes, HCO payments can be made regardless of whether the LTCH stay is considered an SSO outlier or eligible for the full DRG payment. We therefore exclude HCO payments from our analysis and model. About 9% of LTCH stays in our baseline sample have HCO payments, and the median HCO payment in our baseline sample is \$12,428.

DRG. This option is roughly equivalent to a linear interpolation of the full DRG payment between Day 0 and the SSO Threshold.²⁸

3. 100% of the cost of the case, which is computed as total charges multiplied by the facility-specific cost-to-charge ratio.
4. A blend of the Inpatient PPS amount (used at ACH) and 120% of the DRG per diem amount. Note that this option converges to option 2 as LOS increases.

B.2. Empirical payment schedules

PPS payment schedule

We use a commercial software offered by the company 3M (the product is called “Core Grouping Software” (CGS)) to compute counterfactual Medicare payments for each post-PPS period stay.²⁹ Specifically, for each stay in the PPS period, we compute the PPS payment for the actual discharge day and each possible counterfactual discharge day. The inputs into this calculation are the admission date, estimated hospital charges, principal and secondary diagnoses, procedures, discharge status, age, and sex of the patient. For counterfactual lengths of stay, we assume that hospital charges scale linearly with the observed length of stay.

With this information, the software produces the DRG code, the SSO threshold day, and the total Medicare payment for each length of stay. To validate the software, we compare the predicted DRG against the DRG we observe in the data, and the predicted payment against the observed payment for the observed length of stay. The predicted DRG matches the observed value in 99.9% stays and the predicted Medicare payment is within one dollar of the observed Medicare payment in 90% of stays.

Appendix Figure A1 illustrates the resultant, estimated payment schedules for both the pre-PPS and PPS periods. Note that this figure differs slightly from Figure 1 in the paper, which depicts a stylized model of the post-period payment schedules in which the pre-threshold payments are constant per diem. In practice, the pre-threshold payments appear to be slightly bowed downwards; we abstract from this in Figure 1 which we use in our model estimates, where we used the average payment per day for stays discharged before the threshold to construct the slope of the payment schedule prior to the threshold.

What features of the payment rule created the jump in payments at the SSO threshold? Recall that right of the SSO Threshold, short-stay outlier rules do not apply and the payment is just the full DRG payment, which means the cumulative payment schedule is always a flat line to the right of the threshold. To the left of the SSO Threshold, each stay is paid the minimum of four alternative

²⁸To see this, note that 120% of the DRG per diem amount times the length of stay is approximately equal to $120\% \times (\text{Full DRG payment}) / ((6/5)\text{SSO Threshold}) \times \text{LOS}$, which is equal to $(\text{Full DRG payment}) / (\text{SSO Threshold}) \times \text{LOS}$.

²⁹For more information about this software, see: http://solutions.3m.com/wps/portal/3M/en_US/Health-Information-Systems/HIS/Products-and-Services/Products-List-A-Z/Core-Grouping-Software/

payments; the shape of the payment schedules therefore depends on which of the four alternatives is binding. If options 1, 2, or 4 were binding, we would not see a jump at the threshold. Therefore, we conclude that cost of the case must be binding in most cases because we observe a jump on average. Note that the cost of case being binding is necessary rather than the sufficient condition for creating a jump in the payment schedule; the costs could theoretically be such that the payment schedule only has a kink at the SSO threshold rather than a jump. In practice, however, the cost of the case is on average lower than the other options, and we see a jump at the threshold.

Pre-PPS payment schedule

In the pre-PPS period, LTCHs were paid their (estimated) costs, up to facility-specific per-day limit (MedPAC 2014). For most facilities, this limit was binding. For these facilities, we calculate the LTCH payment schedule as the per-day limit multiplied by the length of the stay. For a small number of facilities, the payment limit does not appear to bind. For these facilities, we assume that reported costs are linear in the patient’s length of stay, and we calculate the payment schedule as the (imputed) per-day cost multiplied by the length of stay. When we analyze discharge patterns in the pre-PPS period, we assign each stay the SSO threshold it would have had in the first year PPS period, based on the DRG assignments made using the CGS software described above.

Appendix C. Mortality analysis

We formally test for a mortality effect using a regression discontinuity (RD) design. Let i index individuals and t index days relative to the SSO threshold. Let y_{it} be a mortality indicator. For our analysis of the 1-day mortality hazard, y_{it} takes a value of 1 if the individual dies on day t and takes on a value of 0 if the individual is alive. For the 30-day mortality analysis, y_{it} takes on a value of 1 if the individual dies in the subsequent 30 days and takes on a value of 0 if the individual does not die over this period. Individuals who have already died are excluded from the analysis.

In our baseline RD specification, we allow for a linear trend in the running variable t and permit this linear trend to vary on different sides of the SSO threshold:

$$y_{it} = \alpha_0 + \alpha_1 t + \mathbf{1}_{t \geq 0}(\beta_0 + \beta_1 t) + \varepsilon_{it}. \quad (13)$$

The coefficient of interest β_0 captures the change in mortality at the SSO threshold, conditional on the linear controls. To confirm the robustness of our findings, we also estimate a specification with a quadratic time trend that, as before, is also allowed to vary on different sides of the SSO threshold:

$$y_{it} = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \mathbf{1}_{t \geq 0}(\beta_0 + \beta_1 t + \beta_2 t^2) + \varepsilon_{it}. \quad (14)$$

In both specifications, we restrict our analysis to observations close the threshold, focusing on bandwidths of 3, 5, and 10 days within the threshold. We cluster our standard errors at the DRG level, which allows for correlation in the health process not only within an individual over time but

also within the set of individuals who have the same DRG and therefore may exhibit correlated mortality profiles. We focus the mortality analysis on the post-PPS period.

Appendix Table A3 shows the parameter estimates with Panel A focusing on the 1-day mortality hazard. Column (1) of Panel A, which shows our baseline specification with a linear time trend and a 3-day bandwidth, indicates that 1-day mortality increases by less than 0.01 percentage point at the threshold. This estimate is tiny in absolute magnitude, small relative to the baseline daily mortality rate of 0.6%, and is statistically indistinguishable from zero. Columns (2) to (6) show that this finding is robust to alternative bandwidths and a quadratic time trend. Panel B shows effects on the 30-day mortality hazard. Column (1) of Panel B, which again shows the baseline specification with a linear time trend and a 3-day bandwidth, indicates an economically tiny and statistically insignificant 0.003 percentage point decline in 30-day mortality at the threshold (relative to a baseline 30-day mortality rate of 13.5%). As before, the effect is robust to alternative bandwidths and a quadratic time trend.

Because our standard errors in the regression discontinuity analysis rely on difficult-to-test assumptions about the correlation structure of the error term, we assess the robustness of our statistical inference using permutation inference (Rosenbaum 1984, 2002; Abadie et al. 2014). Specifically, we estimate equation (13) with a bandwidth of 3, replacing the dummy variable for being to the right of the SSO threshold with a dummy variable for being to the right of placebo thresholds defined at $t = -12$ and $t = 42$ in the pre- and post-PPS periods. That is, we estimate an RD effect for a placebo threshold at each day starting 3 days after the start of our sample and ending 3 days before the end (to allow for a 3-day bandwidth); we also exclude days -3 to 3 in the post-PPS period since these days might be contaminated by a potential treatment effect.

Appendix Figure A4 plots the actual effect and the distribution of placebo estimates for the 1-day and 30-day mortality hazards. The plots show that the actual change in mortality at the SSO threshold is not particularly large relative to the typical day-to-day variation in the mortality hazard. The distributions of placebo estimates imply a p-value of 0.796 for the 1-day mortality hazard and a p-value of 0.757 for the 30-day mortality hazard.

Appendix references

Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey M. Wooldridge. 2014. “Finite Population Causal Standard Errors.” Mimeo.

3M Health Information Systems. 2015. “Inpatient Medicare Reimbursement Reference Guide GRP-017.”

Medicare Payment Advisory Commission (MedPAC). 2014. “March 2014 Report to the Congress: Medicare Payment Policy.” Ch. 11.

Kim, Yan, Eric Kleerup, Patricia Ganz, Ninez Ponce, Karl Lorenz, and Jack Needleman. 2015. “Medicare Payment Policy Creates Incentives for Long-Term Care Hospitals to Time Discharges for Maximum Reimbursement.” *Health Affairs* 34(6): 907-915.

Rosenbaum, Paul R. 1984. "Conditional Permutation Tests and the Propensity Score in Observational Studies." *Journal of the American Statistical Association* 79(387): 565-574.

Rosenbaum, Paul R. 2002. "Covariance Adjustment in Randomized Experiments and Observational Studies." *Statistical Science* 17(3): 286-327.

Figure 1: LTCH payment schedules before and after PPS

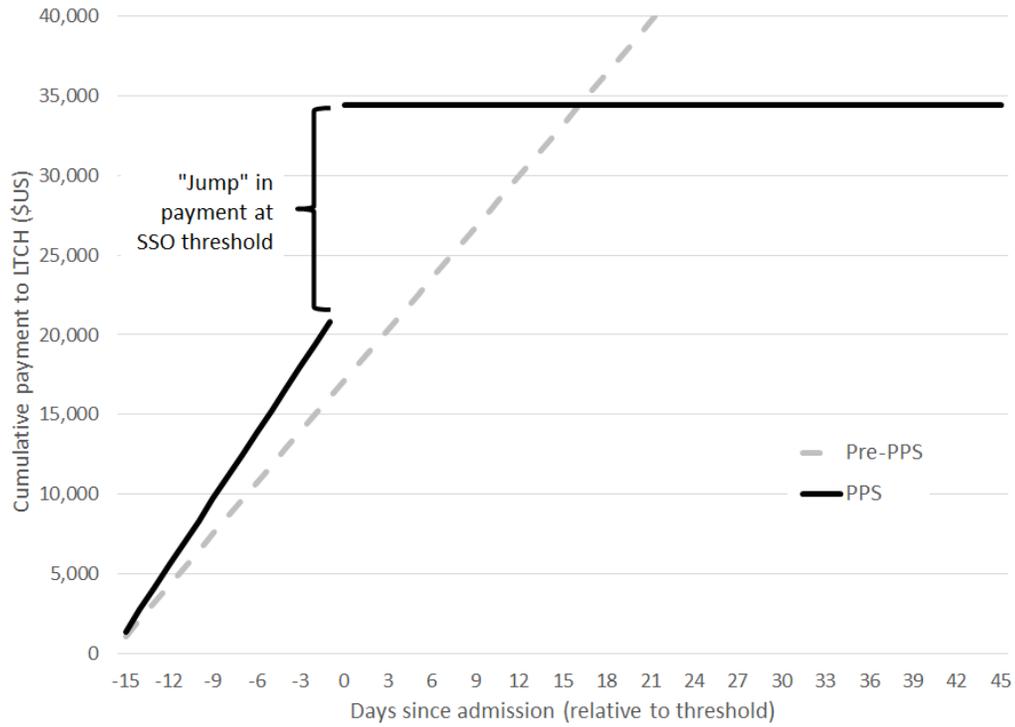
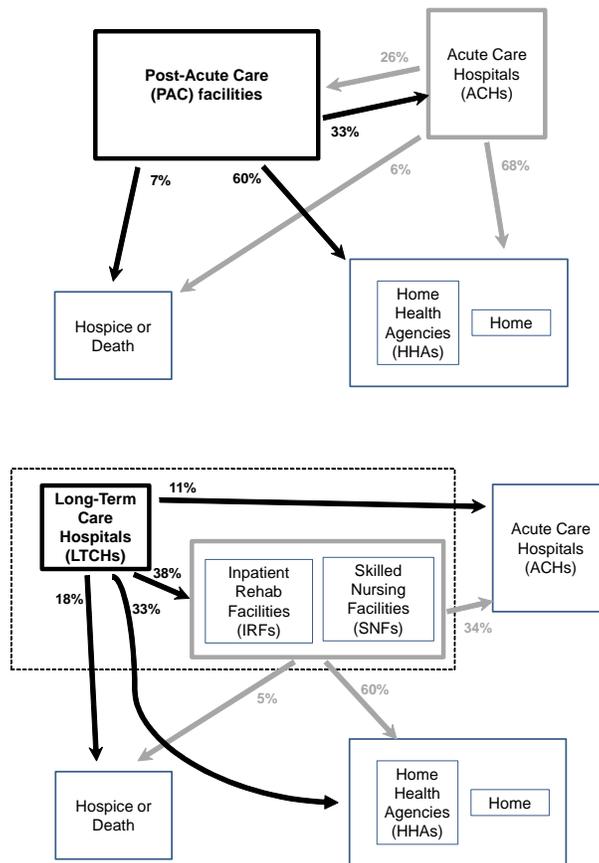


Figure presents the payment schedule in both the pre-PPS and PPS periods. Sample pools admissions that are associated with different short-stay outlier (SSO) thresholds, and x-axis is normalized by counting days relative to the threshold. The linear payment schedule begins with the first day of admission, and the y-axis is normalized to zero for day -16.

Figure 2: Patient flow into and out of Post-Acute Care



Top panel shows patient flow from acute care hospitals (ACHs) to the different destinations: post-acute care (PAC) facilities; home and home health agencies; and death or hospice. Post-acute care facilities include Long-Term Care Hospitals (LTCHs), Skilled Nursing Facilities (SNFs) and Inpatient Rehabilitation Facilities (IRFs). Bottom panel shows how the patient flow pattern is different, within PAC, between Long-Term Care Hospitals (LTCHs) and other PAC facilities (SNFs and IRFs). All numbers for this figure use the universe of Traditional Medicare admissions during the PPS period (October 2007 to December 2012). Numbers are shares of total discharges from each type of facility, excluding a small share of discharges (never greater than 5 percent) that are more difficult to classify. See Appendix for more details.

Figure 3: Discharge patterns by length of stay

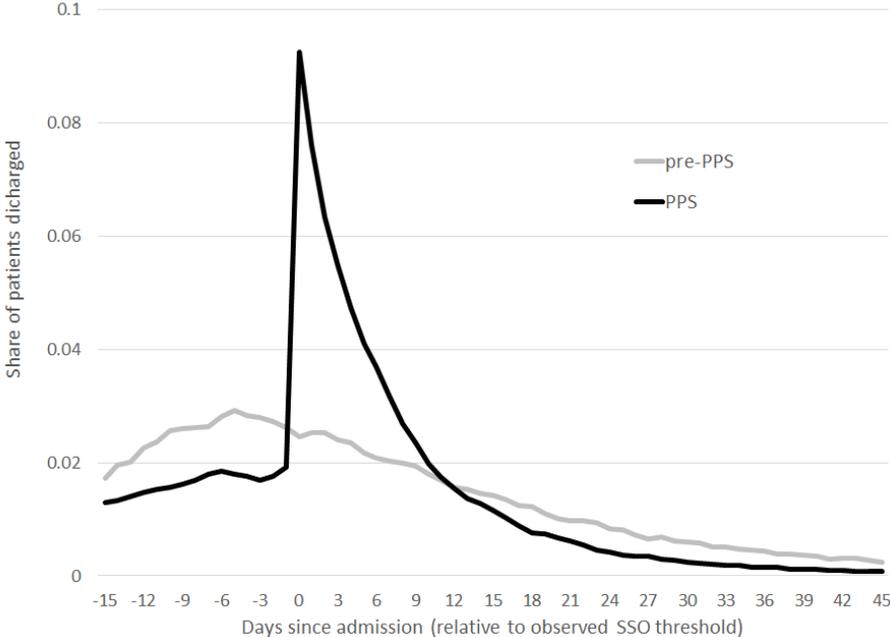


Figure presents the distribution of the time of discharge relative to the SSO threshold. That is, each number graphed represents the number of discharges at a given (relative) day divided by the total number of LTCH admissions. Sample pools admissions that are associated with different SSO thresholds, and x-axis is normalized by counting days relative to the threshold.

Figure 4: Discharge patterns across discharge destinations

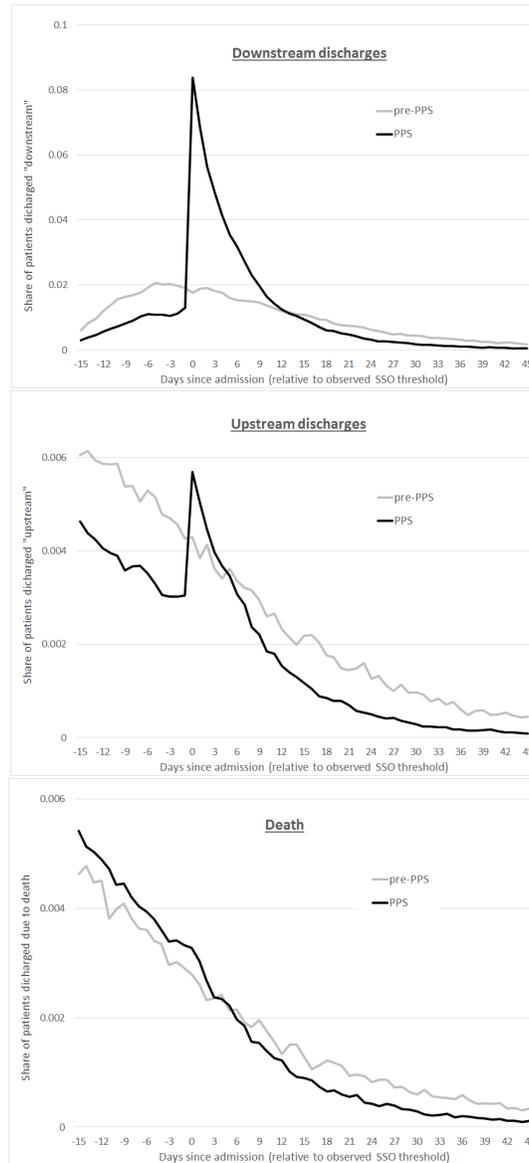


Figure is similar to Figure 3, but presents the distribution separately by discharge destination. Top panel presents discharges “downstream” (SNF, IRF, LTCH, home health, home, or other), middle panel presents discharges “upstream” (ACH or hospice), and the bottom panel presents discharges due to death.

Figure 5: Post-discharge payments

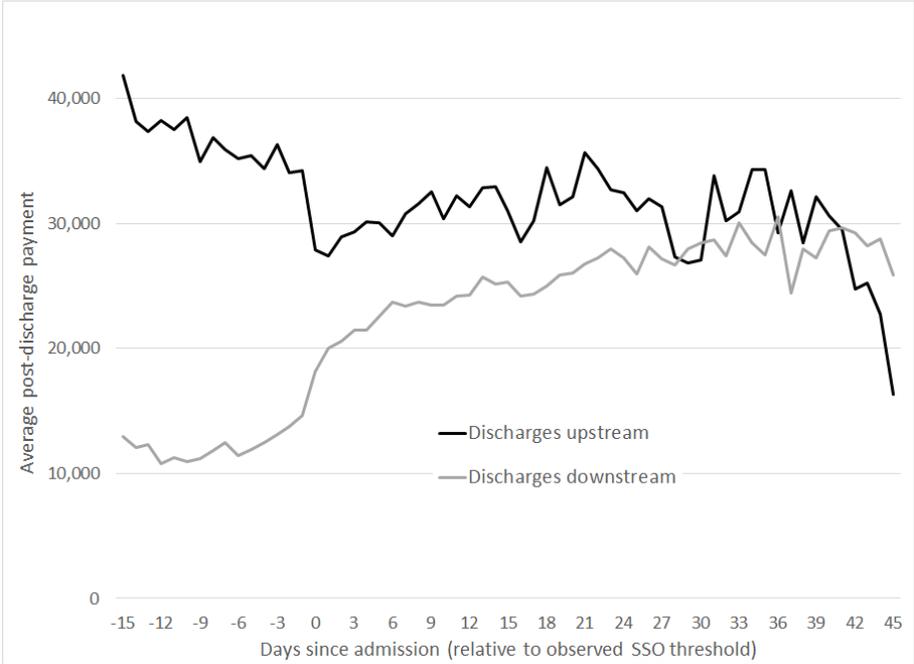


Figure presents the average post-discharge payments for patients discharged alive, by discharge day and discharge destination (upstream vs. downstream, as defined in Figure 4). We define a post-discharge episode as ongoing until there is a break of at least two days that does not involve a facility stay; see text for more details.

Figure 6: Mortality patterns by length of stay

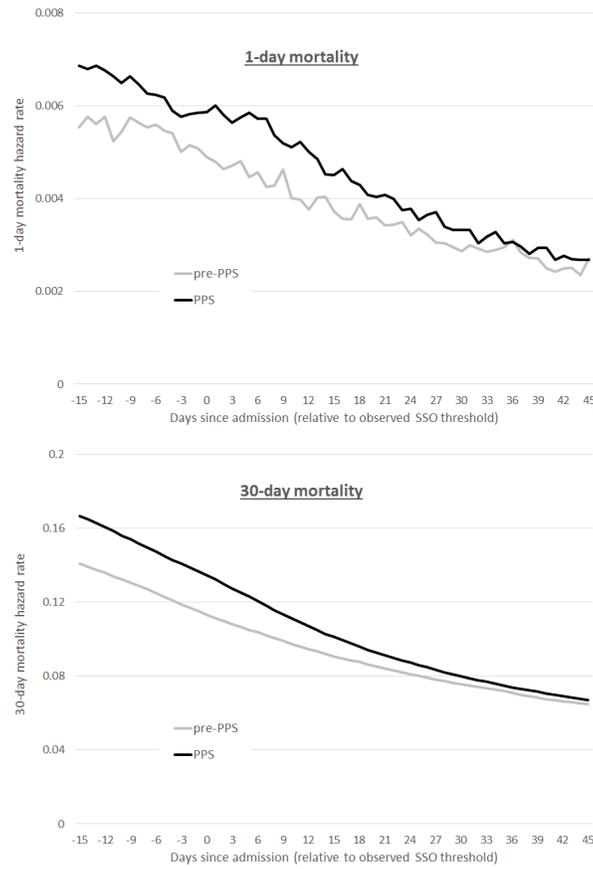


Figure presents post LTCH-admission mortality hazard rates by day. Mortality includes any mortality, whether it occurs within the LTCH or after discharge. Each panel presents hazard rates for different subsequent horizons: same day (top) and 30-day forward (bottom).

Figure 7: Differential patterns for for-profits and not-for-profit LTCHs

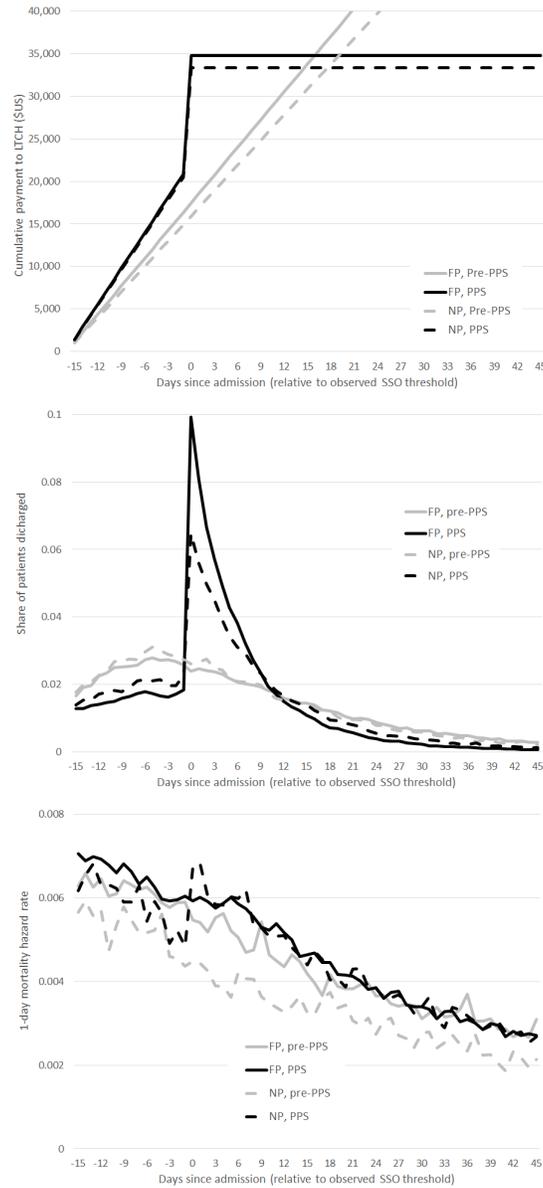


Figure replicates earlier figures, but separates the analysis for for-profit and non-profit LTCHs (the latter includes government-operated LTCHs). The top panel reports the payment schedule, replicating Figure 1. The middle panel reports discharge patterns, replicating Figure 3. The bottom panel reports mortality patterns, replicating the top panel of Figure 6.

Figure 8: Model fit: Discharge patterns

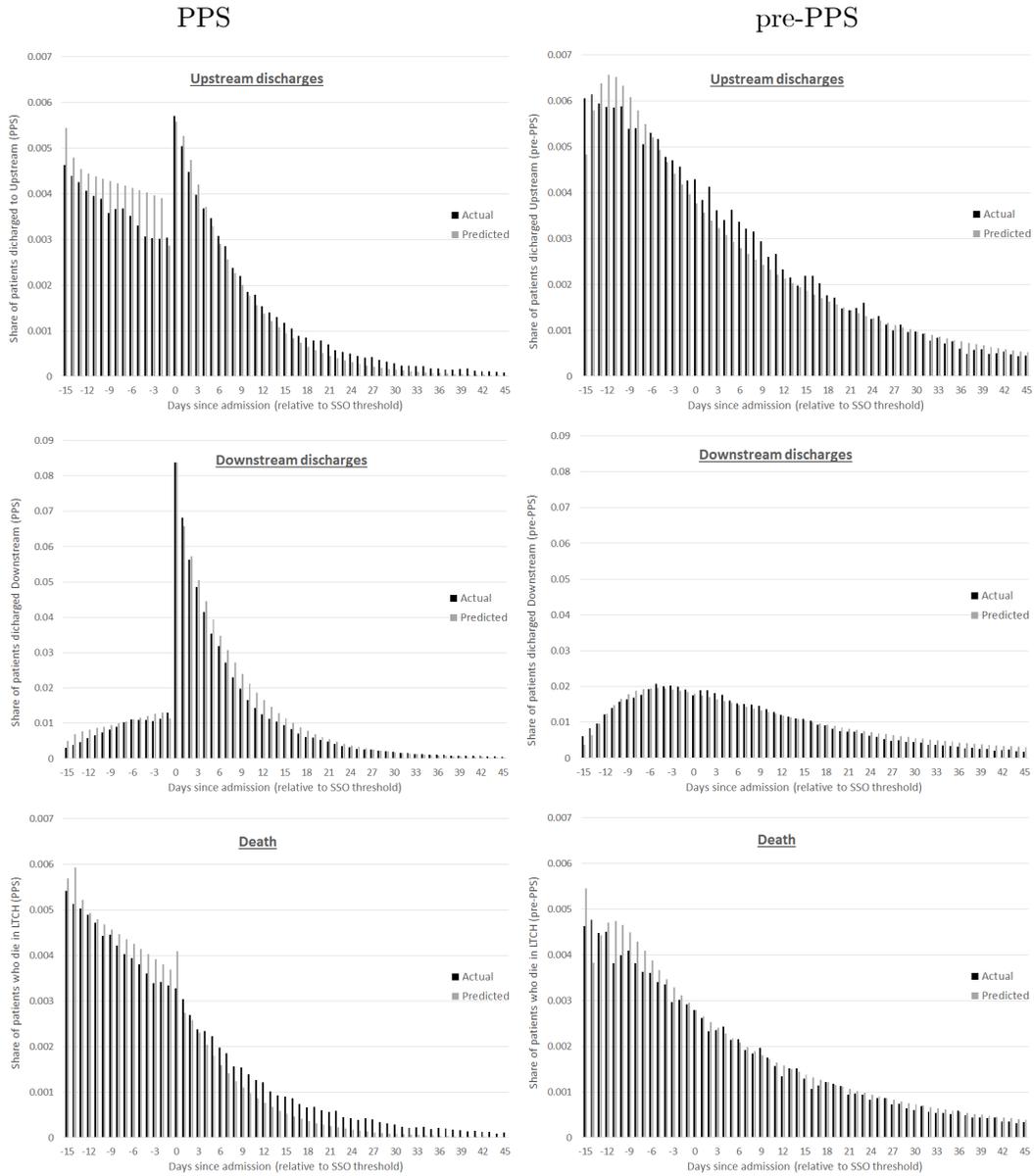


Figure shows the moments we use for estimation, and how the model is able to fit them. Black bars in each panel represent the actual moments from the data, and the gray bars represent the predicted moments from the model estimates. The left three panels represent the PPS period, and the right three panels represent the pre-PPS period. The top panels show discharge rates upstream, the middle panels show discharge rates downstream, and the bottom panels show mortality rates (within the LTCH).

Figure 9: Implied health processes and optimal discharge policy

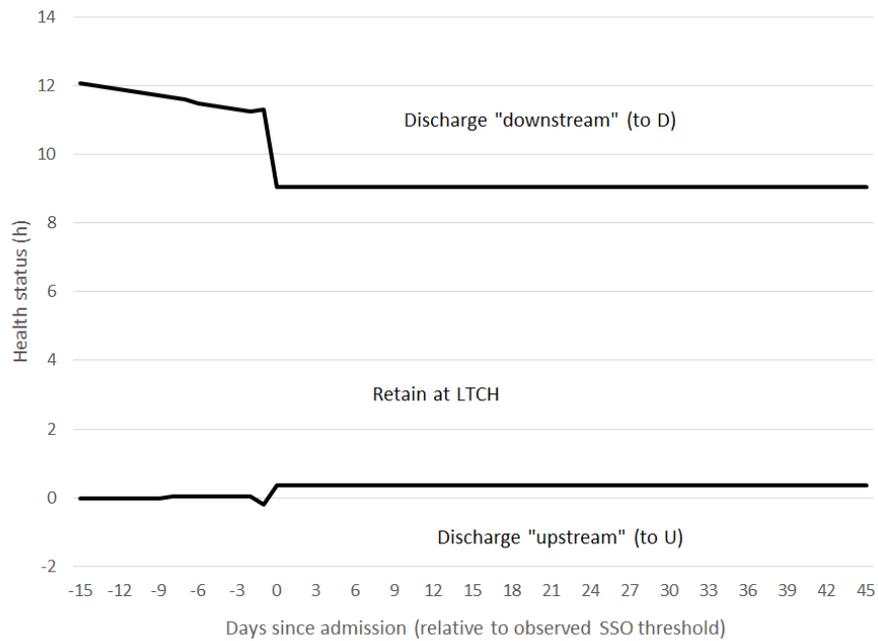


Figure shows the policy function implied by the estimated model. The top black line approximates the health level above which a patient is discharged to D , and the bottom black line approximates the health level below which a patient is discharged to U . Higher h denotes better health (lower mortality). Recall that the discharge decision is a function of the model parameters and an idiosyncratic error term; the lines are drawn to reflect the health level at which 50 percent of patients are discharged to D (top line) and U (bottom line).

Figure 10: Counterfactual payment schedules

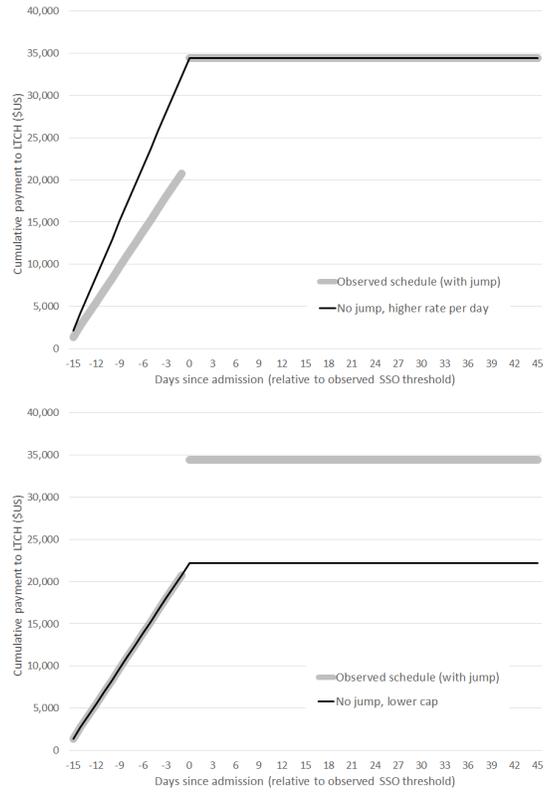


Figure shows the observed (PPS) payment schedule (thick gray line in both panels) and the first two counterfactual payment schedules we consider (black line in each panel). Both counterfactual schedules eliminate the jump in payments at the SSO threshold, but do this in different ways.

Figure 11: Counterfactual policy functions

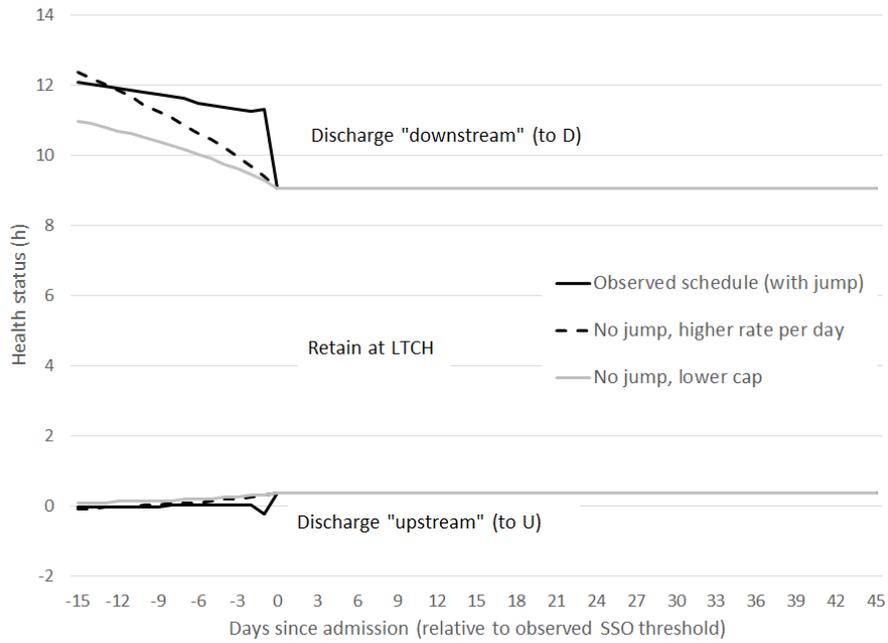


Figure shows the implied discharge policy function from the two “no jump” counterfactual payment schedules described in the main text and illustrated in Figure 10. The discharge policy function associated with the observed contract design is shown in gray and is the same as the one reported in Figure 9.

Figure 12: Counterfactual discharge patterns

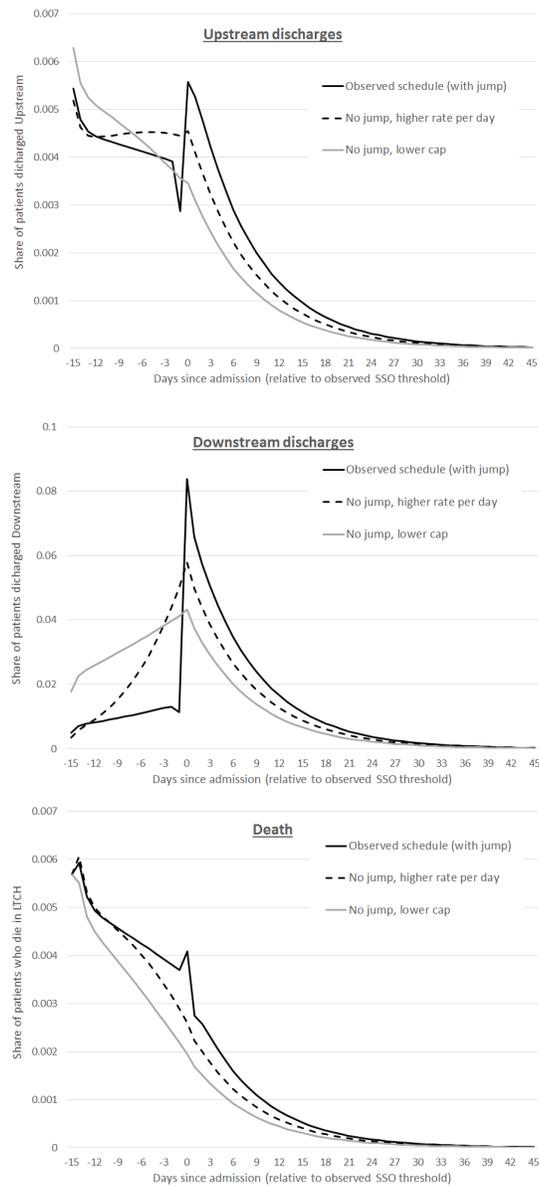


Figure shows discharges (upstream, downstream, and to death) under two counterfactual payment schedules described in the main text and illustrated in Figure 10. The solid black line reports results that are based on our parameter estimates (reported in Table 3) and the observed payment schedule, and each other line reports the results under a different counterfactual payment schedule.

Figure 13: Potential “Win-win” payment schedules

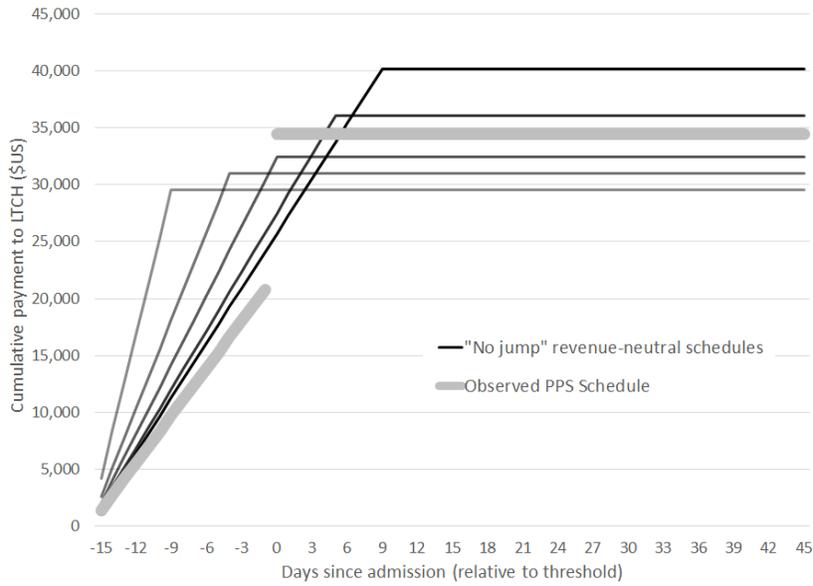


Figure shows some examples of the 21 potential “win-win” contracts we consider. All contracts pay a constant amount up to a threshold length of stay, where they are capped (so that per diem rate drops to zero) with no jump at the threshold. We consider threshold days ranging from +/- 10 days of the current threshold, with the unique payment schedule defined for each threshold day as the one that would hold payments to the LTCH (i.e., LTCH revenue) fixed if they did not change their discharge behavior under the observed contracts.

Figure 14: LTCH payments, LTCH profits, and total Medicare payments from potential “win-win” schedules

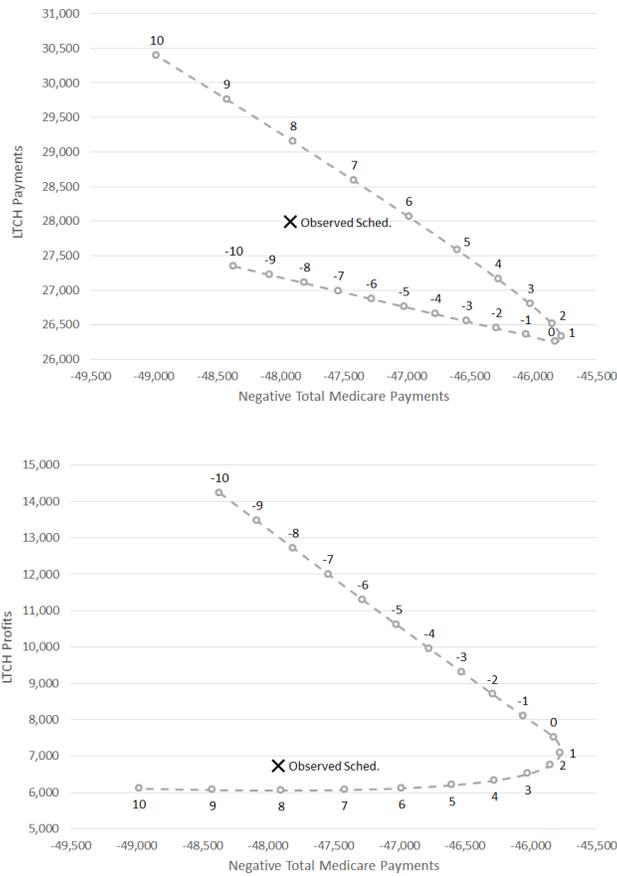


Figure shows outcomes (given the LTCH’s counterfactual behavior) under the various potential “win-win” payment schedules (see Figure 13). For each schedule (represented by a dot which is labeled with the day the payment schedule switches from a per-day rate to a cap) the top panel shows LTCH payments per admission against (the negative of) total Medicare payments (including estimated post-discharge payments) for the episode of care; the bottom panel shows LTCH profits per admission against total Medicare payments.

Table 1: Summary statistics

	Pre-PPS (Jan 2000 - Sep 2002)			PPS (Oct 2007 - Jul 2012)		
	ACH	LTCH	SNF/IRF	ACH	LTCH	SNF/IRF
Number of stays (000s)	29,223	218	5,058	47,559	583	11,172
Panel A. Patient attributes						
Average age	74.5	73.9	80.2	73.5	71.7	79.1
Fraction male	0.43	0.44	0.35	0.44	0.48	0.37
Fraction white	0.84	0.74	0.87	0.82	0.73	0.85
Fraction black	0.11	0.20	0.10	0.12	0.20	0.11
Fraction aged 65+	0.86	0.83	0.94	0.81	0.76	0.91
Fraction dual eligible	0.24	0.31	0.27	0.27	0.38	0.27
Panel B. Patient health indicators						
Number of Chronic Conditions ^a	5.6	6.4	4.6	4.9	7.8	5.6
30 Day Mortality Since Admission	0.082	0.142	0.114	0.079	0.159	0.088
90 Day Mortality Since Admission	0.142	0.275	0.219	0.139	0.308	0.186
Fraction home within 90 days ^b	0.810	0.560	0.501	0.797	0.463	0.556
Three most common DRGs: ^c				Joint Repl. (3.9%)	Ventilator (10.7%)	Rehab w/ CC (17.1%)
				Septicemia (2.8%)	Resp. Failure (8.3%)	Rehab w/o CC (10.2%)
				Dig. Disorders (2.1%)	Septicemia (5.7%)	Ungroupable (3%)
Panel C. Procedures during stay						
Length of stay ^d	5.6	26.6	23.5	5.2	25.4	25.6
Fraction with no procedures	0.43	0.61	0.95	0.40	0.28	0.98
Number of procedures (cond. on any)	2.5	2.4	2.0	2.7	2.9	2.1
Three most common procedures:	Transfusion (6.2%)	Cath (7.5%)	Phys. Therapy (2.6%)	Transfusion (10.2%)	Cath (19.9%)	Occ. Therapy (1.3%)
	Arteriography (5.5%)	Transfusion (5.5%)	Occ. Therapy (2.4%)	Cath. (6.6%)	Transfusion (18.3%)	Phys. Therapy (1.2%)
	Cardiac cath. (5.2%)	Occ. Therapy (5.0%)	Transfusion (0.3%)	Dialysis (4.6%)	Ventilation (14.4%)	Transfusion (0.2%)
Panel D. Payments and cost (2012 \$)						
Total Medicare payments per stay	9,425	28,442	9,727	10,845	35,365	12,983
Medicare payments per day	1,671	1,068	414	2,079	1,392	507
Out-of-pocket payments	772	2,344	1,568	837	1,915	1,954
Out-of-pocket payments per day	137	88	67	161	75	76
Total reported costs	--	28,442	--	--	36,236	--
Reported cost per day	--	1,068	--	--	1,426	--

^a Number of chronic conditions is measured in the calendar year prior to the stay.

^b Reports fraction home at least once during the 90 days after admission, where “home” means alive and not in a facility (ACH, LTCH, SNF/IRF, or hospice).

^c DRG groupings changed between the pre-period and post-period, so for simplicity we report this only for the post-period.

^d Length of stay is censored at 100 days for SNFs, since after that Medicare does not pay and therefore further days are not observed. This applies to about 2 percent of stays.

Table 2: Post-discharge outcomes

	Pre-PPS (Jan 2000 - Sep 2002)			PPS (Oct 2007 - July 2012)		
	Overall	Upstream	Downstream	Overall	Upstream	Downstream
Number of discharges (000s)	187.4	41.7	145.7	504.9	80.3	424.7
Post-discharge 30-day mortality	11.3	24.9	7.4	14.3	47.6	8.0
Post-discharge 90-day mortality	20.3	37.5	15.3	24.4	60.0	17.6
Post-discharge payments ^a	13,370	31,832	8,079	22,975	35,799	20,552
Post-discharge facility days ^a	17.5	33.2	13.0	26.3	33.0	25.0

Table presents summary statistics on post-discharge costs and facility days using the baseline sample of LTCH stays described in Table 1, excluding discharges due to death.

^a Post-discharge payments and post-discharge days refer to the entire post-discharge episode of care, which we define to begin at the day of discharge and end when the patient spends two consecutive days at home.

Table 3: Parameter estimates

Health process during pre-PPS:			Preferences:		
Parameter		Std. error	Parameter		Std. error
μ_0	9.72	0.026	γ	0.78	0.013
σ_0	3.68	0.016	v_1^U (000s)	-30.23	1.854
μ	0.37	0.002	v_0^D (000s)	-74.21	2.341
ρ	0.99	< 0.001	v_1^D (000s)	6.33	0.182
σ	2.13	0.003	σ_ε (00s)	5.50	0.169
Health process during PPS:			Post-discharge payments:		
μ_0	5.40	0.143	ζ_0^U	10.01	0.043
σ_0	1.89	0.065	ζ_1^U	-0.79	0.052
μ	4.45	0.106	ζ_0^D	12.67	0.039
ρ	0.28	0.007	ζ_1^D	-0.27	0.009
σ	2.31	0.067			

Table presents parameter estimates of the parameters in our baseline specification. Standard errors are computed using the asymptotic GMM formula, where the variance-covariance matrix is computed using the bootstrap method, sampling with replacement from the set of LTCH admissions.

Table 4: Discharges and payments from counterfactual payment schedules

	Observed schedule	Higher rate per day	Lower cap	Lowest Medicare Payment within "LTCH preferred" schedules	Linear schedule at estimated cost	Linear schedule at "opportunity" (SNF) cost
	(1)	(2)	(3)	(4)	(5)	(6)
LTCH payments:						
Total payments	27,987	28,357	16,033	26,328	34,998	7,259
Total profits	6,738	9,220	-217	7,082	-907	-8,664
Average LOS ^a	19.2	17.3	14.7	17.4	32.4	14.4
Payment per day	1,457	1,640	1,092	1,514	1,079	505
Discharges Upstream:						
Total payments	3,895	3,610	3,253	3,642	6,089	3,247
Share of discharges	0.12	0.11	0.10	0.11	0.17	0.10
Payment per discharge	33,802	33,441	32,636	33,424	35,481	32,466
Discharges Downstream:						
Total payments	16,038	16,024	16,167	15,805	10,364	15,325
Share of discharges	0.79	0.80	0.83	0.80	0.66	0.83
Payment per discharge	20,377	19,918	19,538	19,675	15,611	18,477
Total Medicare payments	47,921	47,991	35,453	45,776	51,451	25,831
Counterfactual Payment Schedule						
Pre-SSO per diem	1,384	2,151	1,384	1,936	1,107	507
Cap amount	34,419	34,419	22,144	32,907	66,397	30,420

Table presents results from the counterfactual payment schedules. Column (1) reports results that are based on our parameter estimates (reported in Table 3) and the observed payment schedule, and each other column reports the results from predicted discharge patterns under a different counterfactual payment schedule. The counterfactual payment schedules we consider are described in the main text.

^a Length of stay is measured from day -15. To make it comparable to the summary statistics reported in Table 1, all numbers should be increased by 7.5 days (because the average SSO threshold across admissions in our sample is 22.5 days).

Table 5: Discharges and payments from counterfactual organizational forms

	Observed (PPS)	All For-Profit	All Non-Profit
	(1)	(2)	(3)
LTCH payments:			
Total payments	27,673	27,578	28,036
Total profits	5,157	5,681	2,590
Average LOS ^a	19.1	18.7	21.7
Payment per day	1,449	1,477	1,292
Discharges to ACH:			
Total payments	3,701	3,612	4,268
Share of discharges	0.11	0.11	0.13
Payment per discharge	33,687	33,681	33,700
Discharges to SNF:			
Total payments	15,959	16,335	13,844
Share of discharges	0.79	0.79	0.76
Payment per discharge	20,287	20,652	18,227
Total Medicare payments	47,333	47,524	46,148
Counterfactual Payment Schedule			
Pre-SSO per diem	1,385	1,385	1,385
Cap amount	34,438	34,438	34,438

Table reports the discharge and payment patterns under the observed payment schedule. Column 1 reports results under the observed organizational forms, while columns 2 and 3 show two counterfactual scenarios, where we apply the model estimates from for-profit hospitals only and from non-profit hospitals only (see Appendix Table A4) to all LTCHs. Sample excludes approximately 1 percent of patients at hospitals whose organizational form we could not categorize.

^a Length of stay is measured from day -15. To make it comparable to the summary statistics reported in Table 1, all numbers should be increased by 7.5 days (because the average SSO threshold across admissions in our sample is 22.5 days).

Table 6: Robustness to alternative specifications

	Observed schedule	Higher rate per day	Lower cap	Lowest Medicare Payment within "LTCH preferred" schedules
A. Baseline Specification				
LTCH Total Payments	27,987	28,357	16,033	26,328
LTCH Total Profits	6,738	9,220	-217	7,082
LTCH Average LOS	19.2	17.3	14.7	17.4
Total Medicare Payments	47,921	47,991	35,453	45,776
B. Alternative Specification #1: Time-Varying Health Process				
LTCH Total Payments	28,037	28,384	16,006	26,365
LTCH Total Profits	6,137	8,664	-689	6,535
LTCH Average LOS	19.3	17.4	14.7	17.5
Total Medicare Payments	47,893	47,929	35,316	45,725
C. Alternative Specification #2: Post-PPS Moments Only				
LTCH Total Payments	28,009	28,381	16,058	26,367
LTCH Total Profits	7,766	10,128	545	8,010
LTCH Average LOS	19.3	17.4	14.8	17.5
Total Medicare Payments	47,878	47,963	35,445	45,765

Table reports the main results from Table 4 under two alternative specifications of the model. Panel A reports results from the baseline specification, which corresponds to the numbers that are already reported in Table 4. Panel B repeats the analysis, but we allow the AR(1) health process to vary over time by allowing the serial correlation parameter ρ to change linearly with the natural logarithm of days since LTCH admission. Panel C re-estimates the model using only data from the post-PPS period.

Appendix Figure A1: Empirical vs. Approximated payment schedules

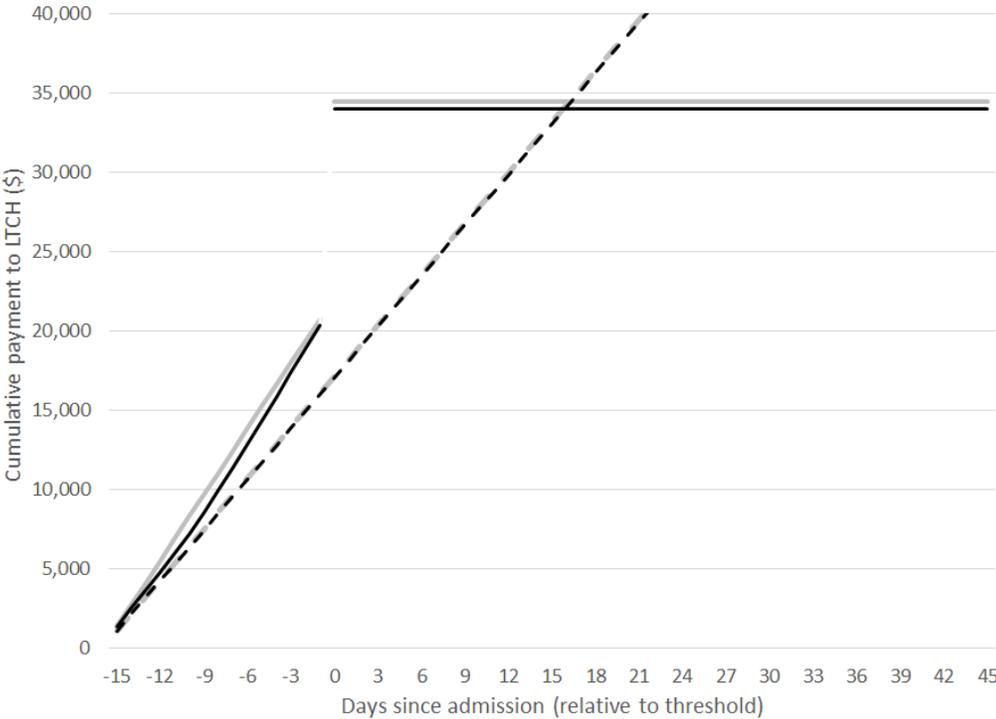


Figure presents the payment schedules used in the paper (gray lines, which are the same as Figure 1 in the main text) against the observed payments (black lines). Appendix B provides more detail about the (slight) differences.

Appendix Figure A2: Discharge patterns, re-weighted

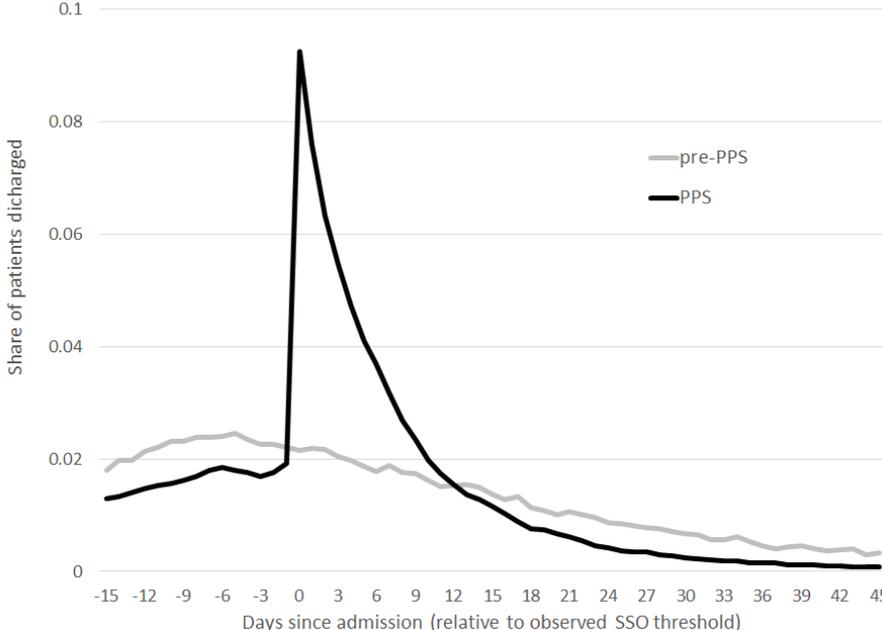


Figure is the same as Figure 3 in the main text, except that pre-PPS line is re-weighted to reflect the same DRG mix as in the PPS period.

Appendix Figure A3: Post-discharge mortality rates

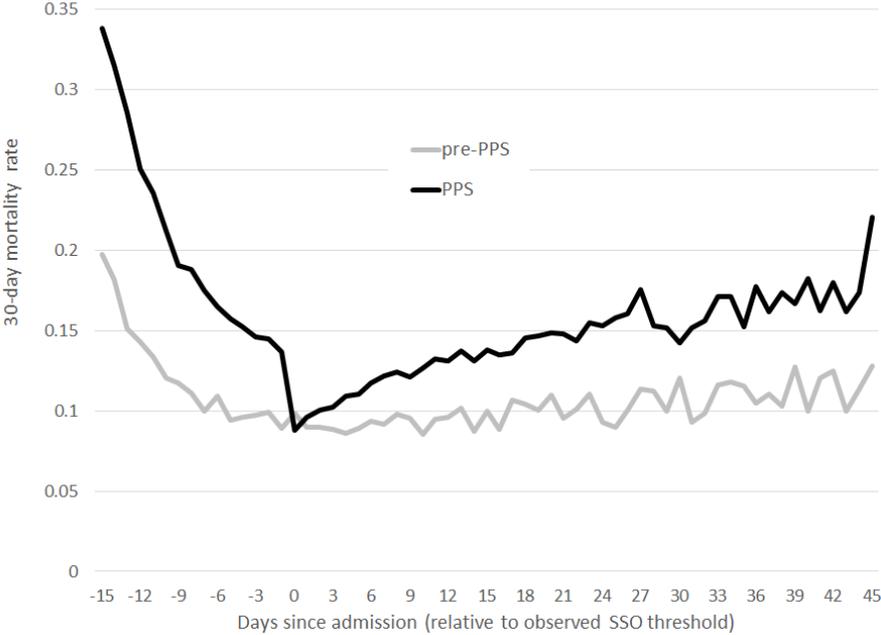


Figure presents the (forward looking) 30-day mortality rate after a (live) discharge, as a function of the day of discharge.

Appendix Figure A4: Perturbation tests for the estimated mortality effect

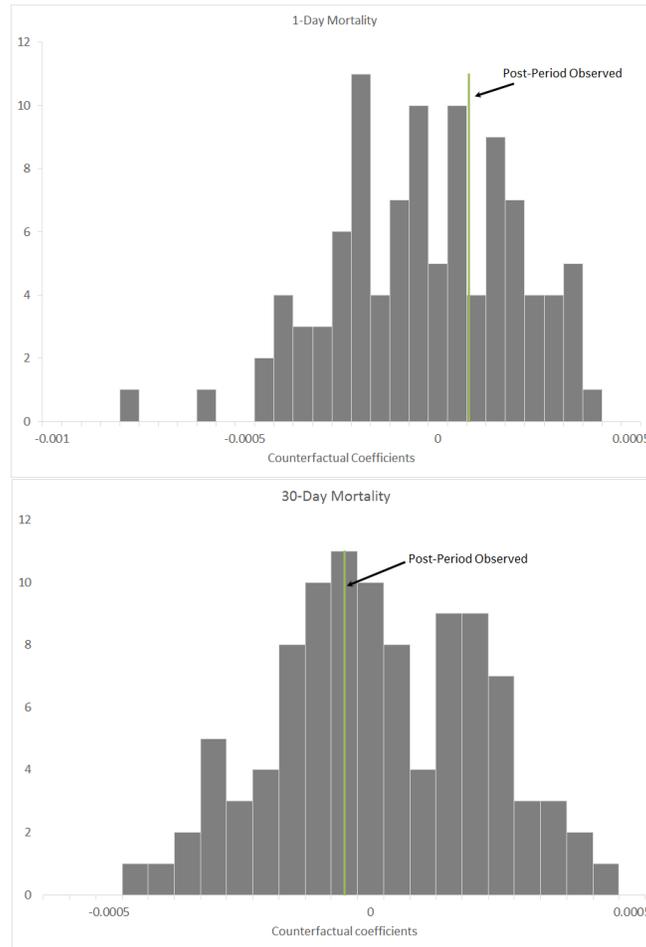


Figure shows perturbation tests for the mortality effect described in Appendix C. The top panel reports the estimated 1-day and the bottom panel reports estimated 30-day mortality effect from estimating equation (13) with a bandwidth of 3, but replacing the dummy variable for being to the right of the SSO threshold with a dummy variable for being to the right of a placebo threshold; see Appendix for more details. The figure also shows where the actual estimated effect falls within this range of placebo estimates.

Appendix Table A1: Discharge destinations

	Pre-PPS	PPS
Death	0.138	0.133
Upstream	0.192	0.138
Inpatient	0.981	0.761
Hospice	0.019	0.239
Downstream	0.670	0.729
LTCH	0.001	0.006
SNF	0.242	0.438
IRF	0.006	0.067
Home Health	0.216	0.305
Home	0.457	0.140
Other	0.077	0.043
Total	1.00	1.00
Number of Obs.	217,562	582,552

Table shows the percentage of discharges to death, “upstream,” and “downstream” in pre-PPS and PPS periods. The discharges upstream and downstream are further decomposed.

Appendix Table A2: Post-discharge outcomes

	Pre-PPS (Jan 2000 - Sep 2002)			PPS (Oct 2007 - July 2012)		
	Overall	Upstream	Downstream	Overall	Upstream	Downstream
Number of discharges (000s)	187.4	41.7	145.7	504.9	80.3	424.7
Post-discharge payments (upper bound)	13,370	31,832	8,079	22,975	35,799	20,552
Post-discharge payments (lower bound)	12,369	31,137	6,991	20,291	33,469	17,801
Post-discharge facility days (upper bound)	17.5	33.2	13.0	26.3	33.0	25.0
Post-discharge facility days (lower bound)	17.3	32.5	12.9	24.9	26.0	24.6

Table presents upper and lower bounds for our imputation of post-discharge payments and days using the baseline sample of LTCH stays described in Table 1, excluding discharges due to death. Appendix A provides more detail. The upper bound is used for our empirical analysis.

Appendix Table A3: Regression discontinuity effect on mortality

	(1)	Linear		(4)	Quadratic	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Effect on 1-Day Mortality Hazard						
Post Threshold Indicator	0.00007 (0.00023) [0.754]	0.00022 (0.00017) [0.204]	0.00039 (0.00013) [0.003]	0.00003 (0.00048) [0.950]	-0.00013 (0.00030) [0.665]	0.00019 (0.00019) [0.317]
Bandwidth	3	5	10	3	5	10
No. of Obs.	4,077,864	6,408,072	12,233,592	4,077,864	6,408,072	12,233,592
Panel B. Effect on 30-Day Mortality Hazard						
Post Threshold Indicator	-0.00005 (0.00017) [0.785]	-0.00022 (0.00016) [0.186]	0.00014 (0.00022) [0.514]	0.00012 (0.00026) [0.651]	-0.00008 (0.00021) [0.719]	-0.00015 (0.00021) [0.475]
Bandwidth	3	5	10	3	5	10
No. of Obs.	4,077,864	6,408,072	12,233,592	4,077,864	6,408,072	12,233,592

Table shows results from the regression discontinuity mortality analysis described in Appendix C. Columns (1)-(3) use a linear functional form (see equation (13)) before and after the SSO threshold, while columns (4)-(6) use a quadratic functional form (see equation (14)). The table reports the estimate of the β_o coefficient, which captures the jump in mortality rate at the SSO threshold in the PPS-period. Each column restricts the analysis to a different bandwidth number of days before and after the SSO threshold. The 1-day mortality hazard is defined as the share of individuals alive at a given day who die by the next day; the 30-day mortality hazard is similarly defined as the share of individuals alive at a given day who die in the next 30 days. Standard errors, clustered at the DRG level, are in regular brackets and p-values in squared brackets.

Appendix Table A4: Parameter estimates, by for-profit status

	For-profit LTCHs		Non-profit and gov. owned LTCHs	
	Parameter	Std. error	Parameter	Std. error
Health process during pre-PPS:				
μ_0	10.66	0.299	9.22	0.163
σ_0	4.07	0.111	3.25	0.073
μ	0.54	0.042	0.36	0.023
ρ	0.98	0.002	0.99	0.003
σ	2.56	0.097	2.05	0.012
Health process during PPS:				
μ_0	4.97	0.162	5.70	0.185
σ_0	1.67	0.075	2.00	0.085
μ	3.88	0.169	4.05	0.088
ρ	0.35	0.018	0.34	0.008
σ	2.16	0.089	2.20	0.078
Preferences:				
γ	0.81	0.024	0.91	0.238
v_1^U (000s)	-30.1	4.1	-123.5	218.6
v_0^D (000s)	-71.7	3.9	-119.8	49.2
v_1^D (000s)	6.50	0.260	8.37	0.401
σ_ε (00s)	8.13	0.503	4.20	5.486
Post-discharge payments during post-PPS:				
ζ_0^U	10.06	0.044	---	same as For-profit LTCHs ---
ζ_1^U	-0.74	0.081	---	same as For-profit LTCHs ---
ζ_0^D	12.67	0.048	---	same as For-profit LTCHs ---
ζ_1^D	-0.28	0.012	---	same as For-profit LTCHs ---

Table presents estimates of the parameters from a specification where we allow all of the parameters, except the post-discharge payment parameters, to vary by for-profit vs. non-profit status of the LTCH. Standard errors are computed using the asymptotic GMM formula, where the variance-covariance matrix is computed using the bootstrap method, sampling with replacement from the set of LTCH admissions.