Do Provider Incentives Always Affect Health Care Costs? New Evidence From Germany*

Jakob Schlockermann†

November 2017

For the latest version, please go to

https://scholar.princeton.edu/sites/default/files/schlockermann/files/JMP.pdf

Abstract

Existing evidence demonstrates that financial incentives for U.S. hospitals affect treatment choices, but the effectiveness of incentives in other cultural and institutional settings is an open question. This paper studies the causal impact of hospital incentives in Germany, which reformed hospital reimbursement in the hope of reducing health care costs. For identification, I make use of the fact that the German payoff schedule for hospitals is increasing in length of stay at first and then kinks and turns flat. I investigate the degree to which hospital discharges are bunched on the kink day. I show

*I am very grateful to my advisors Alex Mas, Ilyana Kuziemko and David Silver for their advice and guidance on this project. I also benefited from comments by David Arnold, Janet Currie, Felipe Goncalves, Daniel Herbst, Bo Honoré, Amanda Kowalski, Andrew Langan, Steve Mello and David Zhang as well as many participants of the Princeton Labor Lunch, the Princeton CHW lunch and the Princeton Labor Seminar. Special thanks to Melanie Scheller from the German Federal Statistical Agency for her patient help with accessing the data. The Princeton University Industrial Relations Section provided generous financial support. Any errors are my own.

†Industrial Relations Section, Princeton University, Princeton, NJ 08544. jakobs@princeton.edu

theoretically that a bunching design identifies the impact of the marginal reimbursement for keeping a patient another day on the patient's length of stay. My estimates are precise and I can reject a more than 0.05 day reduction in length of stay when cutting marginal reimbursement by $1,000$ 2013-Euro per day in the hospital, an order of magnitude smaller than U.S. estimates. A qualitative analysis suggests that the key institutional difference to the U.S. is the role of administrators. My results advise caution when extrapolating evidence from the U.S. to different cultural and institutional settings and point towards new directions for policy in Germany.

1 Introduction

In the face of rising health care costs, countries around the world have reformed the way they reimburse health care providers, since traditional payment systems were thought to give a financial incentive to overprovide. An early and very famous example of such a reform took place in 1983 when Medicare changed hospital reimbursement for inpatient care from a fee-for-service system, which pays the hospital for each individual service provided, to a so-called prospective payment system. Prospective payment systems make the hospital's reimbursement less dependent on the actually provided services and length of stay, but more strongly tied to the expected costs based on case-characteristics, hence giving less financial incentive to increase the number of services and the patient's length of stay.

The 1983 Medicare reform is widely perceived as having reduced the number of services and, in particular, average length of stay (Coulam and Gaumer 1992) and more recent quasi-experimental evidence from the U.S. confirms that financial incentives for U.S. hospitals affect treatment and discharge decisions (Einav et al. 2017 and Eliason et al. 2016). Since the financial incentives apply to the hospitals and not the treating doctors, this evidence implies that U.S. hospital are able to influence their doctors' treatment decisions in a way that makes them take hospital profits into account.[1] Therefore, the U.S. evidence might be very specific to its institutional and cultural context. For instance, German hospital doctors are, in contrast to their U.S. colleagues, unionized and salaried. While Germany in 2004 followed Medicare in moving to a prospective payment system for hospitals in the hope of reducing health care costs, it is an open question whether the U.S. experience of shorter hospital stays carries over to the German institutional context.

This paper provides quasi-experimental evidence on the causal impact of hospital financial incentives on length of stay in Germany. I make use of a unique feature in Germany's reimbursement schedule for hospitals. Like their Medicare counterparts, German patients are grouped into Diagnostic Related Groups (DRGs) based on diagnoses, major procedures and patient demographics. Within a DRG, the reimbursement increases linearly in length of stay up to a certain number of days at which it kinks and becomes flat. Hence, the

[1] There is also a large literature documenting that financial incentives that directly apply to the doctors affect treatment choices, e.g. Clemens and Gottlieb 2014

hospital's marginal reimbursement for keeping a patient in the hospital for another day drops discontinuously at the kink. Thus, discharging a patient becomes discontinuously more attractive the moment the kink is reached. If the hospital's financial incentives affect the decision about its patients' day of discharge, more patients will be discharged on the kink day than what would be expected under a smooth payment schedule without kinks, i.e. there will be bunching of discharges on the kink day. I demonstrate theoretically from estimating the amount of bunching one can infer the causal impact of changing the marginal reimbursement to the hospital for keeping a patient another day on average length of stay.

The empirical analysis is conducted using administrative data covering the universe of in-patient hospitalizations in Germany from 2005-2013, amounting to more than 130 million cases. I start by presenting suggestive evidence on the effect of the marginal reimbursement for another day in the hospital on length of stay in Germany. First, Germany's major 2004 reform that introduced a prospective payment system based on DRGs and reduced the marginal reimbursement for most patients to zero produced no notable break in average length of stay in the time series, suggesting little causal impact. Second, examining the hazard rates around the kink day reveals no notable excess mass of patients being released, again pointing towards very small effects.

This simple static bunching analysis cannot be used to tightly bound the causal effect though, since due to the discrete nature of the assignment variable —days in the hospital— one would need to make strong functional form assumptions regarding the shape of the hazard function. For my main analysis, I therefore make use of the fact that the exact day at which the payment schedule kinks is not only DRG-specific, but can also change from one year to the next. Hence, one can directly evaluate how a DRG's patients' hazard rates compare from one year to the next as the DRG's kink location changes, allowing for a compelling visual assessment to which degree the discharge decisions respond to the financial incentives. Econometrically, the changing kink locations make it possible to estimate the amount of bunching purely from changes in hazard rates from one year to the next without the need for any functional form or smoothness assumptions regarding the shape of the hazard.

Consistent with the suggestive evidence, the visual assessment shows no indications that the hazard rates respond to the changing kink location. Moreover, I can tightly bound the effect of changing the marginal reimbursement hospitals receive for keeping patients another day. Specifically, I can reject that length of stay would fall by more than 0.05 days if the marginal reimbursement for another day in the hospital was reduced by $1,000€$.[2]

My results stand in sharp contrast to recent research by Einav et al. (2017), Eliason et al. (2016) and Kim et al. (2015). Their papers study the effects of hospital incentives in the U.S. by exploiting the fact that the Medicare reimbursement for post-acute care hospitals jumps at a certain length of stay. This notch causes a pronounced excess mass of discharges just after the payment amount jumps. In order to gain directly comparable estimates, I use data from Einav et al. (2017) to implement a bunching analysis similar to my main analysis for the Medicare notch and find an implied reduction in length of stay by 0.34 days when cutting marginal reimbursement for another day in the hospital by $1,000€$, an estimate seven times as big as the upper bound of my confidence interval.

My results also contrast with the experiences from the 1983 Medicare reform. While the magnitudes cannot be compared directly, I argue later in the paper that my estimates would suggest much smaller effects on length of stay than what is normally attributed to the 1983 Medicare reform.

In the discussion section of this paper, I consider why the effects of marginal incentives for hospitals are an order of magnitude smaller in the German setting than in the United States. Building on facts about the institutional structure —German doctors are unionized and under collective bargaining, while most U.S. doctors contract with the hospital regarding the use of the facilities—, anecdotal evidence and interviews with doctors, I argue that U.S. hospital administrators are more active and better able to influence doctors to take hospital profits into account than German doctors who appear to decide very independently of their hospitals' wishes. Because of such institutional and cultural differences, politicians and researchers should be very cautious when extrapolating reduced-form effects from one country to another.

What do my results imply for the effectiveness of the 2004 German reform? Germany introduced its prospective payment system based on DRGs coming from a per diem system, that is a system that pays a fixed amount per hospitalization day.[3] German politicians

[2]Euro in this paper always refers to 2013-Euro.

[3]The system was introduced in 2003 on a voluntary basis. In 2004, it became mandatory for hospitals to participate. Before 2004, the per diem system applied to the majority of cases, but some cases were already reimbursed prospectively.

reformed the hospitals' incentives aiming for a reduction in length of stay. Ulla Schmidt, who was the federal minister of health when the German prospective payment system was introduced, expresses her belief that the system was successful in achieving this goal in a 2009 interview as follows:[4]

> The parliament has passed many reforms in the last years that have proven to make a positive difference in the health care sector. Some things that caused protests initially are now universally accepted as successful. For instance, the DRG system was portrayed like the end of the hospital as we know it. Today we know: Length of stay has decreased and the system has become more efficient.

Since the reimbursement per day under the pre-2004 per-diem system was, on average, less than $1,000$€, the 0.05 provides an upper bound for the reform-induced reduction in length of stay as well. The politicians claiming success apparently confused a secular declining trend in length of stay with the causal effect of their reform.

For future policy, my results suggest that politicians —in institutional settings in which doctors act very independently— could be more effective in their efforts to cut health care costs if they shifted their focus on incentivizing these doctors directly instead of incentivizing the hospitals. Moreover, future research should investigate whether —in countries such as Germany with very independently acting doctors— tying hospital reimbursement more closely to length of stay again might improve welfare, because it would make the reimbursement more closely connected to the hospitals' actually incurred costs. This closer connection might, firstly, reduce the hospitals' financial risk of drawing particularly sick patients and, secondly, might help eliminate the incentive for hospitals to discriminate against particularly costly patients along the admissions margin.

The rest of the paper is organized as follows. Section 2 gives institutional background. In Section 3, I demonstrate theoretically that the bunching design identifies the effect of marginal reimbursement on length of stay. Section 4 discusses the data and the sample selection. Section 5 presents the results. Section 6 provides a discussion and concludes.

[4]Dtsch Arztebl 2009; 106(26) - translation by this paper's author

2 Institutional Background

Germany's health care system is one of the most expensive among OECD countries. In 2013, Germany spent 11.0% of its GDP on health care (OECD average 8.9%, U.S. 16.4%) putting it on fifth position in the OECD. With 8.3 hospital beds per $1,000$ people in 2013 (OECD average 4.8 beds, U.S. 2.9 beds) the hospital sector has a high level of utilization in international comparison, reflecting a high average length of stay (9.1 days in 2013, OECD average 8.1 days, U.S. 6.1 days) as well as a large total number of hospitalizations (252 per $1,000$ people in 2013, OECD average 155, U.S. 125).

Hospital reimbursement in Germany is determined at the federal level and is the same nationwide (except for hospital-specific proportional shift factors as discussed below), irrespective of the patient's health insurer.[5]

Until 2004[6], Germany reimbursed hospitals using a cost-based per diem system in about 80% of cases (the remaining 20% were already reimbursed using a fixed prospective payment, see Theilen 2004). That is, the fee payable to the hospital increased linearly (with a hospital- and department-specific slope depending on the hospital's historical costs) in the number of days a patient stayed hospitalized. In the face of rising health care costs, the German government decided to transition to a prospective payment scheme based on DRGs. The vast majority of cases (more than 94% in 2013) are now reimbursed according to the DRG system (the most prominent exception are the psychiatric cases which only in recent years started to transition to a separate prospective payment system). Based on diagnoses, major procedures and the patient's age, each case is grouped into one out of more than 1,000 DRGs. Due to the complexity of the grouping, more than 75% of hospitals in 2011 employed clinical coders whose main duty is to correctly code diagnoses, procedures and, ultimately, DRGs (Franz et al 2011).

In some cases, the DRG classification can also depend on further variables like birth weight, the discharge reason (e.g., whether the person died) or length of stay. In particular, there are many one-day-DRGs which determine reimbursement in the special case of a patient having a certain diagnosis and staying just one day. These one-day-DRGs do not pose a

[5]For privately insured patients the reimbursement differs - see details in the "supply side institutions" subsection.

[6]Technically, the system already switched in 2003, but it only became compulsory in 2004

problem for my design, however, since for my main research design I only use year-to-year changes in hazard rates for DRGs for which the patient composition is the same from one year to the next according to the official DRG migration tables.[7] That is, DRGs for which the patient composition changes mechanically because, e.g., a new one-day-DRG is introduced are not part of the sample.

The DRG definitions are updated every year and designed to maximize cost homogeneity within DRGs while keeping the number of different DRGs within reasonable limits. The definitions for year t are based on cost data that are collected from a sample of hospital in $t-2$.

Payment Scheme

Within a DRG, the fee payable to the hospital (if the patient is not transferred to or from the hospital - transferred patients are subject to a different payment schedule and analyzed later in a separate section) is a function of the hospital stay length as depicted in Figure 1 (Figure 2 shows a specific example of a DRG payment schedule. As it is apparent from the graph, this DRG has its kink point at five days.). The parameters of the payment scheme are —as the DRG definitions— based on the hospital cost data from two years before. The payment increases linearly until a third of the average length of stay (rounded and measured two years prior) of all patients in this DRG is reached (but at least until day 2 is reached). The slope is determined by dividing average variable costs (that is, total costs excluding costs of major procedures, e.g., bypass surgery) of all patients with this DRG by the number of days at which the kink occurs (again, costs measured two years prior). After the kink, the payment schedule remains flat until the average plus two times the standard deviation of the length of stay of all patients with this DRG two years prior is reached.[8] From then on it increases again linearly. Any out-patient treatments - prior to admission or post discharge - by the hospital are included in the DRG payment (but do not count towards the number

[7] The DRG migration table from $t-1$ to t considers all patients from $t-2$ and groups them into the appropriate DRG according to the system in $t-1$ and according to the system in t. The table then shows how DRGs from $t-1$ map into DRGs from t. For the analysis, I restrict the attention to DRGs that have a one-to-one mapping from $t-1$ to t, that is DRGs with an unchanged patient composition.

[8] To be precise, the upper kink point is the average length of stay plus the maximum of two times the standard deviation or a maximum difference that is determined every year (e.g., in 2005 the upper kink point could at most be the average length of stay plus 17 days)

of days in the hospital).[9]

Interestingly, while the overall goal of the DRG reform was to reduce length of stay, the reduced reimbursement to the left of the lower kink point was introduced in order to discourage hospitals from discharging patients extremely early.

Doctors can easily get information on the DRG's kink location - either because they code the diagnoses and procedures themselves (in which case the software tells them all the information about the patient's DRG) or because they can ask their clinical coder with whom they work closely.

Figure 3 demonstrates how strongly marginal reimbursement changes at the kink. For each DRG, I calculate the slope in € to the left of the lower kink, that is how much revenue the hospital loses if the patient is discharged the day before the kink instead of the kink day. The graph is a histogram of this measure across all years and DRGs. The distribution is centered around about $1,000$€, but with a heavy right tail.

In Figure 4, I also show a histogram of the percentage loss in hospital revenue if the patient is discharged the day before the kink instead of the kink day. I.e. for each DRG I calculate the ratio of the slope to the left of the lower kink and the total amount that the hospital gets paid in the flat part of the schedule and Figure 4 is a histogram of this measure across all DRGs and years. In general, the financial loss of discharging a patient a day before her kink day is quite substantial, although there is a lot of heterogeneity across DRGs.

One concern for identification is that a patient's DRG is not necessarily fixed throughout her hospital stay, but can change if, for instance, the patient gets an infection and therefore a new major diagnosis. In the interviews I conducted, doctors working in German hospitals confirmed that in typical cases the patient's DRG is very predictable from day one.

Demand Side Institutions

Nearly 90% of Germans are in the public health insurance system. There are more than 100 different public health insurers (all public corporations) which compete with each other for patients. The rules for the reimbursement of providers as well as copayments are, however, highly regulated and very similar across insurers. Public health insurance covers all costs of

[9] This is unless the total number of days (in-patient days plus treatment days pre-admission and post-discharge) exceeds the upper kink point (average length of stay plus two times the standard deviation)

hospital stays except for a copay of 10 Euro per day. The copay is only payable for up to 28 days a year, so there is a minor kink from the patient's perspective at 28 days, but she is not affected by the kink in the hospital payment schedule. Civil servants as well as people who earn above a certain threshold can opt to be privately insured. About 10% of people are covered by private health insurers. Private health insurance contracts do typically include a deductible. Hence, these patients have an incentive to contain costs. Unfortunately, I cannot distinguish between publicly and privately insured patients in my data. So only about 10% of patients have to pay a deductible and hence have financial incentives inverse to those of the hospital. These patient incentives contrast with Medicare which features a deductible for all patients, i.e. in Medicare patient incentives are a stronger force countering the hospital incentives than in Germany.[10]

The health insurers in Germany can —as in Medicare— audit bills and appeal. In 2013, 4.4% of cases were successfully audited concerning the length of stay. If an audit is unsuccessful, the health insurer has to pay 300 Euros to the hospital in compensation for the wrong accusation and the resulting work load for the hospital. If the audit is successful, on the other hand, the bill is simply adjusted, but there is no fine for the hospital. According to the health insurers, there therefore is little incentive for the hospitals to adjust their bills in anticipation of the audits.[11] As demonstrated in the generalized model in appendix B, the presence of audits would not confound the research design even with anticipatory behavior, since the causal effect of interest is the effect given the presence of demand side constraints. The audits can, however, introduce measurement error into the length of stay variable. I discuss the issue of measurement error in detail in Section 4.

Supply Side Institutions

In 2013, there were 1,995 hospitals in Germany, with 596 being public, 706 non-profit and 693 for-profit.[12] The payment schedule shown in Figure 1 is identical across hospitals except for a proportional shift factor. The shift factor was different for each hospital when the new

[10] Medicare patients do not pay a deductible if they are also on Medicaid or pay for supplemental Medigap.

[11] see, e.g., Faktenblatt Thema: Abrechnungsprüfung in Krankenhäusern from 06/06/2014

[12] This contrasts with a market for long-term care hospitals studied by Einav et al. (2017), Eliason et al. (2016) and Kim et al. (2015). The long term care hospital market in the U.S. is dominated by for-profit hospitals (Eliason et al. 2016).

system was introduced (the hospital-specific shift factors were introduced in order to have a smooth transition and no sudden jump in hospital revenue relative to the old system which paid hospitals a hospital-specific amount for each day a patient stayed in the hospital), but has since converged to a single factor within each state. Since 2010 the statewide shift factors have been converging towards a factor that is common nationwide. Every year, each hospital and the health insurers agree on a hospital budget, the expected total revenue of the hospital. Deviations from this expected revenue are only partially compensated in order to insure the hospital against random fluctuations in revenue - i.e. changes in treatment decisions that affect hospital reimbursement only partially translate into actual revenue initially until the expected revenue is adjusted.

Hospital have further revenue streams besides the DRG reimbursement. Hospitals can bill separately for a specified list of rare and highly expensive procedures that are not tied to one specific DRG (e.g., implementation of a vagus nerve stimulator). Moreover, hospitals receive additional funds depending on, for example, the amount of investments, whether the hospital provides an emergency room or the degree to which the hospital participates in the training of new doctors. However, these additional funds do not affect the discontinuous break in marginal reimbursement at the kink and are therefore no threat to identification.

German doctors working in hospitals are salaried and unionized, except for the head physicians whose pay is individually contracted and does often depend on economic outcomes in her department (e.g., contracts can depend on the number of times a specific procedure like hip replacement takes place in the head physician's department). In the case of privately insured patients (or publicly insured patients who are willing to pay extra money in order to be treated by the head physician) the head physician can charge additionally per service. Typically, these additional charges are then shared with the other doctors in her department. The discharge decisions is usually made by the patient's responsible doctor. While patients can choose to leave the hospital against their doctors advice, such cases are coded in the data and very rare events. Discharges typically take place midday after the doctor's ward round. The employer-employee relationship between hospitals and doctors with salaries and union protection contrasts with the United States. While there is an increasing fraction of employed hospital doctors in the U.S. —albeit without union protection and collective bargaining—, most are still reimbursed by Medicare directly using a fee for service system and contracting with the hospital (often times as a group of physicians) regarding the use of the facilities.

Medical liability risk in Germany is generally perceived to be small relative to that in the U.S. due to comparatively small awards against physicians.[13] Richard A. Epstein, director of the law and economics program at the University of Chicago Law School, can be quoted with "Nobody is as hospitable to potential liability as we are in this country. The unmistakable drift is we do much more liability than anybody else, and the evidence on improved care is vanishingly thin".[14]

3 Theory

Summary

This section demonstrates that a bunching design identifies the causal impact of marginal reimbursement for another day in the hospital on average length of stay. The model features heterogeneous patients, allows for generic health production and costs functions as well as agency frictions between the hospital and the medical personnel. Under the one additional regularity assumption relative to the standard bunching design as in Kleven (2016) —the assumption is needed due to the discreteness of the assignment variable 'days in the hospital'—, the causal impact of marginal reimbursement can be calculated from the estimated amount of bunching and the estimated mass of patients that would have been discharged at the days above the kink day under a counterfactual smooth payment schedule.

Setup

I present a simplified model here - in appendix B I show that the result holds in a more general model with risk-averse hospitals (the identified effect then applies to a change in marginal reimbursement that also adjusts the fixed payment component such that hospital profits remain constant in equilibrium) as well as the possibility of audits by the health insurers.

The hospital admits a continuum of patients of type θ_i who stay d_i days and enjoy health

[13]see, e.g., Law Library Of Congress - Medical Malpractice Liability Systems In Selected Countries
[14]American Medical News, May 3, 2010

benefit $h\left(d_i, \theta_i\right)$ which is concave in the number of days in the hospital. Patients with higher θ_i are sicker and benefit more from staying in the hospital for longer. That is, $\frac{\partial^2 h(d_i,\theta_i)}{\partial d_i^2} < 0$ and $\frac{\partial^2 h(d_i,\theta_i)}{\partial d_i \partial \theta_i} > 0$. Since there are no functional form assumptions on how θ_i affects $h\left(d_i, \theta_i\right)$, one can assume a uniform distribution $\theta_i \sim U\left[0,1\right]$ without loss of generality.

The hospital receives payment $P\left(d_i\right)$ and incurs costs $C\left(d_i\right)$. The hospital gains utility from profits and from its patients' health (either because of an intrinsic concern for their patients' health or because they fear lawsuits or reputational costs if patients are mistreated). The hospital values profits relative to patient health benefits according to preference parameter $\lambda_h \geq 0$. Since in practice, the medical personnel and not the hospital shareholders make the discharge decision, the hospital faces an agency problem in implementing its objective function. I model this agency problem in the form of parameter $0 \leq \lambda_d \leq 1$ which dampens the degree to which profits are taken into account. That is, the patients' length of stay is determined by the solution to

$$\max_{\{d_i(\theta_i)\epsilon\mathbb{N}\}} \lambda_d \lambda_h \int_i \left[P\left(d_i\right) - C\left(d_i\right)\right] + \int_i h\left(d_i, \theta_i\right)$$

If the agency problem were modeled in a different way, the bunching design would not necessarily identify the causal effect of interest anymore, because the kinks might have effects on hazard rates away from the kink point. I discuss this point in more detail later in the results section and provide evidence against such effects of the kink on hazard rates away from the kink.

Note also that changes in admission and coding behavior - while interesting subjects to study in their own right - do not threaten the validity of this paper's findings. If anything, adjustments in coding and admission behavior would lead to an overestimate of the bunching mass in my setting. This is because the incentive to deny admission or to upcode to a different diagnosis with a different kink location is smallest for those patients who would otherwise be discharged on the profit maximizing kink day.

Optimal Hospital Behavior

I assume that $P\left(d_i\right)$, $C\left(d_i\right)$ and $h\left(d_i, \theta_i\right)$ are shaped such that the objective function is globally concave. At baseline, consider a linear payment schedule $P^{baseline}\left(d_i\right) = \bar{p} + p \cdot d_i$.

Optimization than amounts to choosing cutoff values for θ_i determining which patient types are kept for how many days. $\overline{\theta}_d^{baseline}$ denotes the highest θ_i for which the patient stays d days under the baseline schedule. A patient with a θ_i just above $\overline{\theta}_d^{baseline}$ would stay $d+1$ while a patient with a θ_i just beneath $\overline{\theta}_d^{baseline}$ would stay d days. The cutoff values defining the range of patients who are discharged on day d^* are implicitly defined by equations

$$\lambda_d \lambda_h \left[C\left(d^*+1\right)-C\left(d^*\right)-p\right]=h\left(d^*+1, \overline{\theta}_{d^*}^{baseline}\right)-h\left(d^*, \overline{\theta}_{d^*}^{baseline}\right)$$

$$\lambda_d \lambda_h \left[C\left(d^*\right)-C\left(d^*-1\right)-p\right]=h\left(d^*, \overline{\theta}_{d^*-1}^{baseline}\right)-h\left(d^*-1, \overline{\theta}_{d^*-1}^{baseline}\right)$$

That is, for patient type $\overline{\theta}_{d^*}^{baseline}$ the hospital is just indifferent between the net profit valued with $\lambda_d\lambda_h$ of keeping her d^*+1 instead of d^* days and the net health benefit it would bring to the patient. A patient with θ_i a little bigger than $\overline{\theta}_{d^*}^{baseline}$ would be kept d^*+1 days, since her health benefit of staying another day is higher than for the $\overline{\theta}_{d^*}^{baseline}$ patient. Similarly, for patient type $\overline{\theta}_{d^*-1}^{baseline}$ the hospital is indifferent between the marginal health benefit of keeping her d^* instead of d^*-1 days and the profit impact.

Now consider the policy experiment of interest, reducing marginal reimbursement by $\triangle p > 0$ throughout the schedule, i.e. $P^{reform}\left(d_i\right)=\bar{p}+\left(p-\triangle p\right)\cdot d_i$. with $\triangle p > 0$. Note that with risk-neutral hospitals, the fixed payment amount $\bar{p}$ does not affect hospital behavior - in appendix B, I consider risk-averse hospitals. The cutoff values are now defined by

$$\lambda_d \lambda_h \left[C\left(d^*+1\right)-C\left(d^*\right)-\left(p-\triangle p\right)\right]=h\left(d^*+1, \overline{\theta}_{d^*}^{reform}\right)-h\left(d^*, \overline{\theta}_{d^*}^{reform}\right)$$
$$\lambda_d \lambda_h \left[C\left(d^*\right)-C\left(d^*-1\right)-\left(p-\triangle p\right)\right]=h\left(d^*, \overline{\theta}_{d^*-1}^{reform}\right)-h\left(d^*-1, \overline{\theta}_{d^*-1}^{reform}\right)$$

which implies that the cutoff values increase, i.e. $\overline{\theta}_d^{reform} > \overline{\theta}_{d^*}^{baseline}\ \forall d$. That is, the patients stay on average for a shorter time. Due to the discreteness of the assignment variable —length of stay— I need to make an additional regularity assumption relative to the standard bunching setting. Specifically, I assume that patients who share the same length of stay d under the old schedule $P^{baseline}\left(d_i\right)$ move towards at most two different length of stay values under the new schedule $P^{reform}\left(d_i\right)$. That is, those patients who stay, e.g., 5 days under the old schedule, stay for $3-4$ days or for $4-5$ days under the new schedule, but never for $2-4$ or $3-5$ days.

Now consider the introduction of a convex kink at d^*. That is, the payment schedule becomes

$$P^{kink}\left(d_i\right)=\begin{cases}\bar{p}+p\cdot d_i & d_i\leq d^* \\ \\ \bar{p}+\left(p-\triangle p\right)\cdot d_i & d_i>d^*\end{cases}$$

Under the new kinked payment schedule, the new cutoff values defining who is discharged at d^* —$\overline{\theta}_{d^*}^{kink}$ and $\overline{\theta}_{d^*-1}^{kink}$— are defined by

$$\lambda_d\lambda_h\left[C\left(d^*+1\right)-C\left(d^*\right)-\left(p-\triangle p\right)\right]=h\left(d^*+1,\overline{\theta}_{d^*}^{kink}\right)-h\left(d^*,\overline{\theta}_{d^*}^{kink}\right)$$

$$\lambda_d\lambda_h\left[C\left(d^*\right)-C\left(d^*-1\right)-p\right]=h\left(d^*,\overline{\theta}_{d^*-1}^{kink}\right)-h\left(d^*-1,\overline{\theta}_{d^*-1}^{kink}\right)$$

Hence, $\overline{\theta}_{d^*-1}^{kink}=\overline{\theta}_{d^*-1}^{baseline}$ and $\overline{\theta}_{d^*}^{kink}=\overline{\theta}_{d^*}^{reform}>\overline{\theta}_{d^*}^{baseline}$ and $\overline{\theta}_{d^*}^{kink}-\overline{\theta}_{d^*}^{baseline}$ is the excess mass or bunching at d^* under the kinked schedule. Hence, $\overline{\theta}_{d^*}^{kink}$ is the marginal buncher who responds to the introduction of $P^{kink}\left(d_i\right)$ the same way as to the introduction of $P^{reform}\left(d_i\right)$.

What Does a Bunching Design Identify?

We established that the marginal buncher responds to the introduction of the kink the same way as to the policy experiment of interest (that is, changing marginal pay by $\triangle p$ throughout the schedule). Let $\tilde{d}$ denote the length of stay that the marginal buncher $\overline{\theta}_{d^*}^{kink}$ would have enjoyed under the baseline linear schedule. For this marginal buncher, the causal effect of interest —the effect of changing marginal reimbursement per day by $\triangle p$ on length of stay— is $\frac{d(d_i)}{d(p)}\triangle p=\tilde{d}-d^*$. Using the assumption discussed above, the causal effect $\frac{d(d_i)}{d(p)}\triangle p$ is equal to $\tilde{d}-d^*$ for all patients who would have stayed $\tilde{d}$ under the baseline linear schedule and for whom $\theta_i<\overline{\theta}_{d^*}^{kink}$, but the causal effect $\frac{d(d_i)}{d(p)}\triangle p$ is $\tilde{d}-\left(d^*+1\right)$ for all patients who would have stayed $\tilde{d}$ under the baseline linear schedule and for whom $\theta_i>\overline{\theta}_{d^*}^{kink}$. Therefore, the total causal effect on patients staying $\tilde{d}$ under the old baseline schedule is

$$
\begin{aligned}
E\left[\frac{d\left(d_i\right)}{d\left(p\right)}\triangle p\left|\bar{\theta}_{\tilde{d}}^{baseline} > \theta_i > \bar{\theta}_{\tilde{d}-1}^{baseline}\right.\right] &= \left(\tilde{d}-d^*\right)\frac{\bar{\theta}_{d^*}^{kink}-\bar{\theta}_{\tilde{d}-1}^{baseline}}{\bar{\theta}_{\tilde{d}}^{baseline}-\bar{\theta}_{\tilde{d}-1}^{baseline}} \\
&\quad + \left(\tilde{d}-\left(d^*+1\right)\right)\frac{\bar{\theta}_{\tilde{d}}^{baseline}-\bar{\theta}_{d^*}^{kink}}{\bar{\theta}_{\tilde{d}}^{baseline}-\bar{\theta}_{\tilde{d}-1}^{baseline}} \\
&= \tilde{d}-d^*-1+\frac{\bar{\theta}_{d^*}^{kink}-\bar{\theta}_{\tilde{d}-1}^{baseline}}{\bar{\theta}_{\tilde{d}}^{baseline}-\bar{\theta}_{\tilde{d}-1}^{baseline}}
\end{aligned}
$$

A simple example makes the formula intuitive: If the observed bunching mass is only a small fraction of the observed mass at d^*+1 —say, 10%— $\tilde{d}=d^*+1$, because the marginal buncher is coming from d^*+1, and $\frac{\bar{\theta}_{d^*}^{kink}-\bar{\theta}_{\tilde{d}-1}^{baseline}}{\bar{\theta}_{\tilde{d}}^{baseline}-\bar{\theta}_{\tilde{d}-1}^{baseline}} \approx 0.1$, since only patients who are at d^*+1 under the counterfactual linear schedule and whose $\theta_i < \bar{\theta}_{d^*}^{kink}$ bunch at d^* together with the marginal buncher. In the example, the formula tells us that the average causal effect on the patients staying d^*+1 days under the counterfactual linear schedule is 0.1 days, since that is the fraction of patients who move from d^*+1 to d^* due to the kink.

Since d^* is known, we need to estimate $\tilde{d}$ and $\frac{\bar{\theta}_{d^*}^{kink}-\bar{\theta}_{\tilde{d}-1}^{baseline}}{\bar{\theta}_{\tilde{d}}^{baseline}-\bar{\theta}_{\tilde{d}-1}^{baseline}}$ in order to get the causal effect of interest. Let B denote the bunching mass estimated from the data and $f\left(d\right)$ the estimated expected mass of patients at d under the contrafactual linear schedule. Then $\tilde{d}$ can be inferred from the data by finding the value for $\tilde{d}$ for which

$$
f\left(d^*+1\right)+...+f\left(\tilde{d}\right) \geq B
$$

$$
f\left(d^*+1\right)+...+f\left(\tilde{d}-1\right) \leq B,
$$

since the bunching mass is equal to the mass at the days from d^*+1 up to $\tilde{d}-1$ plus the fraction of the mass at $\tilde{d}$ that bunches. This fraction is the bunching mass that is not explained by the mass coming from d^*+1 up to $\tilde{d}-1$, i.e.

$$
\frac{\bar{\theta}_{d^*}^{kink}-\bar{\theta}_{\tilde{d}-1}^{baseline}}{\bar{\theta}_{\tilde{d}}^{baseline}-\bar{\theta}_{\tilde{d}-1}^{baseline}} = B - f\left(d^*+1\right)+...+f\left(\tilde{d}-1\right)
$$

Hence, the remaining challenge is to estimate B and $f(d^* + 1)$, etc. in order to estimate the parameter of interest $E\left[\frac{d(d_i)}{d(p)} \triangle p \middle| \bar{\theta}_{\tilde{d}}^{baseline} > \theta_i > \bar{\theta}_{\tilde{d}-1}^{baseline}\right]$.

4 Data and Sample Selection

Data

I use administrative data from the Federal Statistical Agency in Germany. It covers the universe of in-patient hospitalizations covered by the DRG system[15] from 2005 - 2013, more than 10,000,000 cases each year with variables including baseline patient-characteristics like sex, age and region as well as case-characteristics like diagnoses, procedures, length of stay, hospital identifier and admission and discharge date. All hospitals are required by law to report all of the previous year's hospitalizations until March 31st to the Federal Statistical Agency. The data the hospitals send to the agency is based on the data generated for billing purposes and hence of very high quality.

Measurement Error in Length of Stay

Length of stay may be mismeasured for two reasons: First, if a patient is readmitted within 30 days or before her DRG's upper kink point is reached —counting from the first admission date— and if the patient obeys certain criteria, the two hospitalizations are merged into one case and length of stay is summed up. That is, only one case would show up in my data and the hospital is reimbursed as if it were one case. Second, health insurer's audits might introduce measurement error into my length of stay variable. If a bill is successfully audited the actual length of stay and the billed one (which is the one that shows in the data) can deviate. If patients discharged on the kink day are especially likely to be audited or to be readmitted, this measurement error could lead to an underestimate of bunching.

There are two reasons why my results appear to be robust with respect to the measurement error problem: First, I have two different measures for length of stay which are differentially affected by the two types of measurement error. The billed number of days —which yields the correct length of stay in the case of a readmission, but misreports length

[15] As mentioned before, the vast majority - more than 94% in 2013 - are reimbursed according to the DRG system. The only major exception are psychiatric patients.

of stay in case of an audit, since the billed number of days is adjusted after an audit— and the difference between discharge and admission date —which produces some measurement error in the case of readmissions (since days after the first discharge and before the readmission would wrongly be counted towards total length of stay), but is not affected by audits, since admission and discharge date are unadjusted after a successful audit. I conducted my analysis using both measures and the results are very similar, suggesting that the measurement error is of little importance. The reported results in this paper are for the difference between discharge and admission date.

Second, while the billed number of days mismeasures length of stay in the case of a successful audit, this problem is much larger for cases admitted early in the calendar year than for cases admitted closer to the end of the calendar year due to the point in time at which the data is collected - for a detailed discussion see appendix A. Using the billed number of days, I find no evidence that there is more bunching in the data for cases from later in the year, again suggesting that the measurement error is not the driving force behind the results.

Sample

Throughout the analysis, I exclude those with missing data on DRG, length of stay, discharge reason or admittance reason. Also I focus on discharges the timing of which actually are under the doctor's control - i.e. I drop deaths or discharges against the doctor's advice from the sample. Furthermore, I analyze transfers to or from other hospitals in a separate section, since those are subject to special reimbursement rules. After applying these restrictions I am left with more than 85% of the overall number of cases. The remaining 15% are mostly due to deaths and transfers.

For data privacy reasons I cannot make use of observations for which there are less than 3 patients that are discharged with a certain DRG, in a certain year and after a certain number of days in the hospital. This measurement error will only affect very uncommon DRGs and is unlikely to significantly affect any of my results, especially for the weighted regressions.

5 Results

Time Series Evidence

Figure 5 shows the development of average length of stay in Germany over time around the 2004 introduction of the DRG system.[16] Average length of stay is on a secular declining trend (the trend is also present before and after the time window shown in Figure 5). This trend is not specific to Germany, but can be observed in many developed countries. Figure 17 in appendix C shows how length of stay evolved over the 2000s in a sample of OECD countries.

The secular decline in length of stay appears to not only be common across countries, but also across hospitalization reasons and hospitals within Germany. First, technical progress making surgeries less taxing is not the sole driver of the decline. Figure 18 in appendix C shows how length of stay evolves for DRGs that involve a major procedure compared to DRGs that do not. While the absolute decline in length of stay is larger for DRGs involving procedures, the trend for the other DRGs is similar. Second, the trend is not specific to the hospitals that started out with particularly long durations. I split up the hospitals into quartiles based on their average residualized length of stay in 2005. Figure 19 in appendix C shows how average length of stay evolves for those four groups of hospitals. The hospitals in the fourth quarter that had unusually long durations in 2005 do indeed see the largest fall in length of stay afterwards, but a declining trend is also visible for the remaining hospitals.

As discussed in more detail in Section 6, the German 2004 reform changed hospital reimbursement for most patients from being linearly increasing towards being flat in length of stay, i.e. the marginal reimbursement was reduced to zero for most patients. Had the marginal reimbursement for another day in the hospital any meaningful effect on length of stay in Germany, we would expect to see a downward jump in length of stay around the time of introduction. Yet there is no apparent break in the time series around 2004 suggesting little causal impact. However, the pre-existing trend towards shorter stays and the missing control group as well as the possibility of changes in coding and admission behavior make it difficult to draw confident conclusions just from the time series.

[16]Note that the source is the official German hospital statistic here which is why the numbers are different than the OECD numbers for length of stay for Germany. I use the OECD numbers only when comparing Germany to other countries.

Static Bunching Analysis

Therefore, I next turn to the static bunching analysis making use of the kink in the payment schedule. To provide some sense for the distribution of the kink locations as well as length of stay in the cross section, Figure 6 shows the distribution of kink locations across DRGs, Figure 7 provides a histogram for length of stay and Figure 8 shows how often patients stay shorter respectively longer than their DRGs' kink locations.

If marginal reimbursement had an influence on the discharge decision, the hazard rates on the kink day should be unusually large. Figure 9 presents the hazard rates around the kink for the example DRG discussed earlier (the payoff schedule is shown in Figure 2). There is no bunching apparent at the kink at five days.

To graphically analyze the degree of bunching for the pooled data, I restrict the sample to DRGs with a kink location at six days or higher and center all observations around their respective kink location and pool them. Figure 10 shows the resulting hazard rates plotted against the number of days in the hospital relative to the kink location of the patient's DRG. If the marginal reimbursement for the hospital had a meaningful impact on discharge decisions, we would see an unusually large hazard rate at 0. Yet, there is no apparent excess hazard at the kink point, again pointing towards no major effects of marginal reimbursement on length of stay decisions. The graph looks similar for other restrictions on the data like selecting only DRGs with a kink location of at least 5 or 7 days.

Dynamic Bunching Analysis

The coarse nature of the running variable 'days in the hospital' makes it difficult to implement the standard static bunching design econometrically, since the smooth counterfactual hazard rate in the absence of the kink cannot be pinned down precisely. Therefore, I make use of DRGs with changing kink locations over time to get precise counterfactual hazards without functional form assumptions.

Since the DRG definitions are updated every year, there are mechanical changes in the patient composition for some DRGs from one year to the next. As discussed previously, the official DRG migration table from $t-1$ to t considers all actually observed patients in $t-2$ and groups them into the appropriate DRG according to the system in $t-1$ and according to the system in t. The table then shows how DRGs from $t-1$ map into DRGs from t. For the analysis, I restrict the attention to DRGs with an unchanged patient composition (that is, DRGs for which the migration table shows a one-to-one mapping from $t-1$ to t) as well as a change in kink location from $t-1$ to t.

Table 1 shows descriptive statistics for the analysis sample in comparison with the remaining patient population. The analysis sample is comparable to the remaining cases in terms of covariates, but patients in the analysis sample have a longer average length of stay by construction, since in order for a DRG to have a changing kink location it must necessarily have had an average duration of at least 7.5 days at some point (the analysis of transfers later in the paper will provide some sense for whether the results carry over to less serious diseases). The stability of covariates from one year to the next within a DRG for which the kink location changes is discussed and shown in the form of regression analyses at the end of this section.

To provide some sense of how often kink locations change, Table 2 shows for how many DRGs the kink location changes from one year to the next in a certain way as well as how many patients these DRGs cover. It is apparent that the kink locations decrease more often than rise due to the secular trend towards shorter stays.

Graphical Analysis

I start with an example. Figure 11 shows the payoff schedule for DRG H62A in years 2005 and 2006. The kink moves from four to three days. In Figure 12, I present this DRG's hazard rates for the two years. There is no noticeable increase in the hazard at three or decrease in the hazard at four days in 2006.

In order to conduct the graphical analysis for a pooled sample of DRGs, I focus on the most common change in kink location, that is on DRGs for which the kink location decreases by one day from year t to year $t+1$. I center all observations around their respective kink location in t and pool them. Figure 13 shows the hazard rates in t and in $t+1$ for this pooled sample plotted against days relative to the kink location in t. If there were meaningful bunching behavior, we would see relatively higher hazard rates at 0 for year t (since 0 is that year's kink) and relatively higher hazard rates at -1 (which is the kink location in $t+1$) for year $t+1$. The shape of the hazard rates, however, looks very similar in t and $t+1$ except for a tendency towards higher hazards in $t+1$ in general. This tendency is reflective of the trend towards shorter stays. The fact that there is notable move of discharges away from 0 and towards -1 provides strong evidence that the hospitals do not adjust treatment length in response to marginal reimbursement.[17]

Figure 14 shows the same graph for a tighter time window, specifically with the data restricted to October until December for year t and January to March for year $t+1$. Figure 14 supports the conclusions from Figure 13, albeit being a little bit less precise due to the smaller underlying mass of data.[18]

One possible identification concern is that doctors might need more than a year to adjust to a new kink location. In order to test for this possibility, Figure 15 shows the evolution of hazard rates from one year to the next for DRGs for which the kink location did not change. If doctors learned over time to optimize their discharge behavior with regards to the kink, we would expect the hazard rate at the kink day to rise (relative to the hazard rates for the other days) in year $t+1$ compared to year t. The graph shows no indication of such learning behavior.

Regression Analysis

For the econometric analysis, I calculate the hazard rate $hazard_{drg,t,d}$ for each DRG drg for each year t for each possible length of stay d. I then estimate the following type of specifications

$$\ln hazard_{drg,t,d} = \delta \cdot \triangle p_{d,drg,t} + \alpha_{drg,d} + \gamma_{drg,t} + \epsilon_{drg,d,t}$$

$\triangle p_{d,drg,t}$ is zero if d is not the kink day for DRG drg in year t and denotes the decrease in marginal reimbursement at the kink (measured in 1,000€) otherwise - for instance, $\triangle p_{d,drg,t} = 3$ for the kink day if the payoff schedule features slope $3,000$€ per day to the left of the kink and becomes flat afterwards. Hence, δ is the percentage change in the hazard

[17]Note that the hazard rates at -3 drop relative to the hazard rates at -2 by construction, because if, e.g., a DRG has its kink location at 3 days in t, it will necessarily have a hazard of zero at -3. But this effect is identical for year t and year $t+1$ and the hazard rates can be compared directly.

[18]In contrast to Figure 13, Figure 14 features hazard rates in t that are generally higher than in $t+1$. This most likely due to month effects. Patients admitted in December typically feature higher hazard rates - possibly, because major surgeries with a long expected duration in the hospital are postponed until after Christmas and new year.

rate if at d there is a kink at which marginal reimbursement is decreased by $1,000€$. The $\alpha_{drg,d}$ are indicators for each DRG-day-combination allowing for an arbitrary hazard rate pattern across days for each DRG. Note that this implies that there are no functional form assumptions regarding the shape of the hazards across days within a DRG. Instead, δ is identified purely via how strongly the hazard rate for the same $drg-d$-pair changes from one year to the next when $\triangle p_{d,drg,t}$ changes. That is, the identifying variation comes from changing kink locations for the same DRG from year to year as well as from changes in the size of the jump of marginal reimbursement at the kink for the same DRG from year to year.

The indicators for each DRG-year-pair $\gamma_{drg,t}$ allow for a DRG-specific proportional shift in hazard rates each year. These indicators capture general time effects for each DRG, e.g., a trend towards shorter stays/higher hazards for some DRGs. Note that for a bunching design with a kink (in contrast to one with a notch) there is no hole in the distribution to be expected to the right of the kink which is why no indicator for the day above the kink day is needed.

All standard errors are bootstrapped with $N = 400$ and clustered at the DRG-level. I convert the estimated parameter δ to the parameter of interest —the causal effect of cutting marginal reimbursement by 1,000 2013-Euro on length of stay $E\left[\frac{d(d_i)}{dp}\middle|\bar{\theta}_{\tilde{d}}^{baseline} > \theta_i > \bar{\theta}_{\tilde{d}-1}^{baseline}\right]$— as discussed in the theory section. The conversion takes place inside the bootstrap so that the standard errors are accurately adjusted. The conversion procedure requires the expected mass of patients at the days above the kink day $f(d^*+1)$, etc. under the hypothetical linear schedule without a kink. I use the observed mass at d^*+1, etc., since if the bunching mass is small the difference is negligible (Kleven 2016). In order to limit the influence of days d that are far away from the kink on the estimation of $\gamma_{drg,t}$, I restrict the range of d over which I estimate the equation to the lowest observed kink location of DRG drg minus ten days and the largest one plus ten days. The results are barely changed by varying this range or not restricting the range at all.

Table 3 presents the regression results. Columns one and two present the causal effect of interest for the unweighted and weighted regression equation without the $\gamma_{drg,t}$ fixed effects. Columns three and four repeat these specifications including $\gamma_{drg,t}$. The estimates are similar and close to zero for all four specifications, yet a little more precise if weighted and if one includes the $\gamma_{drg,t}$ fixed effects. The specification reported in column four —weighted and including the $\gamma_{drg,t}$— is the most precise and implies an increase of 0.019 days in average length of stay when increasing marginal reimbursement by $1,000$€. Across the four specifications one can reject an implied causal effect of more than a 0.05 days increase.

Next, I investigate the possibility of downward rigidity in treatment decisions. Specifically, it is imaginable that it is easier for doctors to keep patients a day longer in response to an increase in kink location than it is to keep them shorter when the kink location goes down. Columns five and six restrict the sample to DRGs that feature a decreasing respectively increasing kink location over time. Note that the samples are not entirely distinct, because some DRGs have increases as well as decreases in kink location over time. Only using the DRGs with kink location decreases, one finds results very consistent with the prior results. Restricting the sample to DRGs with increases shows a somewhat larger effect albeit with reduced precision. But even this specification still allows me to bound the causal effect below 0.12 days.

Lastly, I test whether hospitals learn how to optimize profits in the presence of the DRG system over time. I repeat the specification from column four, but restrict the sample to data from 2010 or later. There is no clear upward jump in the coefficient, suggesting that even after some years of learning the new system the hospitals still do not discharge patients in a profit maximizing manner.

Heterogeneity

I investigated the possibility of heterogeneous treatment effects by hospital size, by discharge reason (regular discharge vs death vs discharge into post-care such as hospice or rehabilitation), by type of DRG (medical vs surgical - defined by whether a major procedure is part of the DRG's definitions) and by hospital ownership (public and non-profit vs private). None of these regressions showed any relevant effect sizes. Hence, the result is appears to be very robust even across cuts of the data. The results are available in appendix C.

The absence of an effect of hospital ownership is particularly remarkable, since hospital ownership is known to be a predictor of the degree to which U.S. hospitals respond to incentives (see, e.g., Duggan 2000 or Einav et al. 2017). This suggests that the reason for my findings is not in German hospitals being less profit oriented, but that an agency friction prevents them from reacting to the incentives.

Bunching Estimates for Less Serious Diseases and Transferred Patients

A possible concern with this paper's analysis is that the causal effect is identified only for fairly serious diseases, since the average duration for a DRG must necessarily be above 7.5 days at some point in order for the DRG to possibly have a changing kink location. Transferred patients - who have been excluded from the analysis so far - can shed some light on whether the results carry over to less seriously sick patients, because transfers from or to the hospital obey special reimbursement rules. For transfers, the reimbursement for the hospital is linearly increasing until the rounded average duration from two years prior is reached. Afterwards, the payment schedule becomes flat and identical to the one for non-transferred patients. Since the kink for transfers is located at the rounded average duration measured two years prior, transfers allow me to implement the dynamic bunching design for all DRGs that are comparable from one year to the next and for which the average duration —rounded and measured two years prior— changes. Importantly, this includes DRGs with fairly small average durations such as two or three days. Table 4 presents the results when estimating the main analysis equation $\ln hazard_{drg,t,d} = \delta \cdot \triangle p_{d,drg,t} + \alpha_{drg,d} + \gamma_{drg,t} + \epsilon_{drg,d,t}$ using transferred patients and the changing kink locations of their respective payoff schedules. The results are again very close to zero and even more precise than those of the main analysis, suggesting that this paper's conclusions are not limited to severely sick patients.

Robustness to Alternative Forms of Agency Frictions

The bunching design identifies the causal effect of marginal reimbursement on length of stay if the agency friction takes the form modeled in Section 3, i.e. if the hospitals can make the medical decision makers take hospital profits into account, albeit to a dampened degree. It is imaginable, however, that the hospitals evaluate their doctors based on some rule of thumb - e.g., it could be that the hospital responds to a decrease in the kink location by asking the doctors to reduce the patients' average length of stay. In this case, the payment schedule would have an effect on length of stay, yet there would not be any bunching.

In order to investigate whether the kink has effects on hazard rates of patients away from the kink, I again make use of the changing kink locations over time. Specifically, I estimate the following fixed-effect regressions to test whether the kink location has any impact on *average* length of stay:

$$days_{i,drg,t} = \beta \cdot kinklocation_{drg,t} + \alpha_{month} + \alpha_{drg} + \sum_{j=1}^{J} \delta_j \cdot (avgduration_{drg,t-2})^j + \gamma X_{it} + \epsilon_{i,drg,t}$$

$days_{i,drg,t}$ denotes the patient's (who is in DRG drg in year t) length of stay and $kinklocation_{drg,t}$ is the location of the kink of DRG drg in year t (for instance, $kinklocation_{drg,t} = 5$ in the case of the example DRG in Figure 2). α_{month} and α_{drg} are fixed effects for each month and each DRG.

Recall that the kink location of a DRG is determined by the average location measured two years prior, specifically $kinklocation_{drg,t} = max\left(2, round\left[\frac{avgduration_{drg,t-2}}{3}\right]\right)$. Hence, DRGs for which the kink locations decrease are on a downward trend in length of stay. In order to account for that, I control for polynomials in $avgduration_{drg,t-2}$ in most specifications. After controlling for $avgduration_{drg,t-2}$ there is still variation in $kinklocation_{drg,t}$ left due to the rounding. Essentially, controlling for $avgduration_{drg,t-2}$ allows to compare the development of length of stay for a DRG for which the kink location changes with another DRG for which the average duration evolved similarly in the past, but for which the kink location does not change because of the rounding. X_{it} are further controls that I employ in some specifications, specifically hospital-month fixed effects and DRG-specific age effects and indicators for gender.

The results are presented in Table 5. The first column shows the result without controlling for the average duration from two years prior. As expected, this specification shows a positive effect of the kink location on length of stay, since any time trend in length of stay for a DRG will also affect the kink location. Controlling for the average duration two years prior, however, moves the coefficient very close to zero and adding further controls does not change the estimates meaningfully. Hence, the kink does neither have an effect on the hazards at the kink nor away from the kink, implying that the results are robust to alternative models of the agency friction.

Recall also that the time series evidence from Figure 5 shows no effect of going from a schedule that is increasing in length of stay for most patients to a mostly flat schedule, which —irrespective of how one models the agency friction— further supports the case that the marginal reimbursement for the hospital has no impact on length of stay in Germany.

Kink Location and Covariates

The dynamic bunching design analyzing the hazard rates at the kink locations when the kink locations change did not require the observables and unobservables of a DRG's patient population to be orthogonal to the DRG's kink location. Instead, the dynamic bunching design only required that any other determinants of length of stay that are correlated with the kink location do not *differentially* affect the hazard rates depending on whether the day is a kink day or not.

The fixed effect regressions in the last subsection, however, do require other factors to be orthogonal to the kink location changes. While the robustness of the results to including individual controls already demonstrated this stability, Pischke and Schwandt (2015) show that cofounders should be analyzed as the dependent variable also.

I investigate whether changes in observables are correlated with changes in the kink location by running specifications of the form

$$covariate_{i,drg,t} = \beta \cdot kinklocation_{drg,t} + \alpha_{drg} + \alpha_{month} + \sum_{j=1}^{J} \delta_j \cdot (avgduration_{drg,t-2})^j + \epsilon_{i,drg,t}$$

with $covariate_{i,drg,t}$ denoting the the observable of interest for patient i with DRG drg in year t. The other variables are defined as in the regressions in the last subsection.

Tables 6 to 9 show the results for age, gender, the number of diagnoses and the number of procedures. None of these regressions show a significant association between the covariate and the kink location after controlling for the fixed effects and the polynomials in the average duration measured two years prior.

6 Discussion

Contrast to the recent U.S. evidence

The results presented in the last section stand in a sharp contrast to recent evidence from the United States. Einav et al. (2017), Eliason et al. (2016) and Kim et al. (2015) study the effects of a jump in the Medicare reimbursement for post-acute hospitals after patients pass a certain length of stay threshold. These papers find a pronounced excess mass of discharges right after the payment jumps.

I used data from Einav et al. (2017) to construct an estimate that can be compared quantitatively to my paper's numbers, i.e. an estimate of how length of stay responds to a cut in the marginal reimbursement per day by 1,000€.[19] The main estimate is a reduction in length of stay by 0.34 days, seven times more than the upper bound of my paper's confidence interval. The contrast is especially remarkable, since hospital care —as measured by the number of hospital beds per capita, the number of hospitalizations and average length of stay— is generally much more extensive in Germany than in the U.S. (the U.S. has had a comparatively —by OECD standards— low utilization of hospital services throughout the last decades, for the recent OECD numbers see Section 2). The larger baseline amount of hospital services suggests lower medical returns to care at the margin in Germany than in the U.S. and, hence, if anything, more ability for the German health care providers to adjust care in response to financial incentives.

I interviewed two doctors who work or recently worked in German hospitals.[20] Both of them were aware of the kink in the schedule, both of them realized that a discharge on the kink day is particularly financially attractive for the hospitals and both of them confirmed that they were aware of the kink locations for often occurring DRGs. Moreover, getting to know a DRG's kink location is easy by using the coding software or —if they have one working with them— by asking the medical coder. While one doctor recalled a general information session that informed about DRGs and the payoff structure for the hospitals, none of them reported any pressure from the hospital in their day-to-day activities. One of the interviewed doctors denied taking the kink location into account in her decision making at all and the other doctor claimed that the kink location could only affect his decision if he were otherwise completely indifferent between two different discharges dates. Moreover, he

[19] I use the probability mass distribution of discharges across days reported in Einav et al. (2017) for before and for after the jump in the Medicare reimbursement was introduced. I compare the probability mass distribution before and after the introduction in order to estimate the effect of the jump in reimbursement on length of stay. For details, see appendix D.

[20] The interviews were conducted August 25th and October 19th 2017 via phone

stated to "not associate with his hospital's motives".[21]

This anecdotal evidence stands in a sharp contrast to stories from the U.S. (originally reported in an article by Weaver et al. 2015 in the Wall Street Journal (WSJ) and cited in Eliason et al. 2016) that suggest a strong pressure to discharge patients after the threshold at which the hospital's Medicare payment jumps is reached. The WSJ article describes that during meetings, hospital staffers would discuss treatment plans, "armed with printouts from a computer tracking system that included, for each patient, the date at which reimbursement would shift to a higher, lump-sum payout." Reports also allege that long-term care hospital administrators "sometimes ordered extra care or services intended in part to retain patients until they reached their thresholds, or discharged those who were costing the hospitals money regardless of whether their medical conditions had improved," while "bonuses depended in part on maintaining a high share of patients discharged at or near the threshold dates to meet earnings goals." Further anecdotal evidence supporting the case of hospital managers actively trying to influence treatment decisions comes from a 2014 blog post by Richard Gunderman, professor of radiology at the Indiana University School of Medicine, who reports about a document —literally called "How to Discourage a Doctor"— outlining strategies for hospitals to gain more control over the doctor's decisions.

Apparently, hospital administrators play a much larger role in the U.S. in shaping treatment decisions. Moreover, German doctors are salaried and unionized employees under collective bargaining, making them relatively independent of their respective hospital. In contrast, U.S. doctors are typically not hospital employees —and if they are, there is no unionization and collective bargaining—, but they are billing their services to Medicare separately from the hospital.[22] That is, the hospital is billing Medicare for the hospital side of services (using DRGs as in Germany), the doctors are billing Medicare per service for the physician side and the hospitals and physicians need to contract regarding the use of the facilities.

Since other institutional features —discussed in Section 2— such as patient incentives and liability risk cannot explain the observed differences between the U.S. and Germany, I

[21] Translation by this paper's author.

[22] In 2013, "roughly 25 percent of all specialty physicians who see patients at hospitals are employed" ("7 Trends in Hospital-Employed Physician Compensation" in Becker's Hospital Review 01-25-2013)

interpret my findings as the agency friction λ_d from the model being smaller in Germany than in the U.S., i.e. as German hospitals being less able to make the medical decision makers take hospital profits into account. The importance of agency frictions in hospitals is also a key theme in recent work by Sacarny (2016) who finds that U.S. hospitals differ drastically in their ability to make doctors code diagnoses in a way that benefits the hospital financially.

Implications for the German reform and contrast to the Medicare reform

This paper bounds the causal effect of reducing marginal reimbursement for another day by $1,000€$ below 0.05 days. As discussed in Section 2, prior to 2004 Germany used a cost-based per diem system in about 80% of cases, that is hospitals were reimbursed using a linear schedule in length of stay with hospital- and department-specific slopes that depended on the hospital's historical costs.[23]

While I do not have data on these hospital- and department-specific slope parameters, the average slope must have been smaller than $1,000€$ per day, because —given the average length of stay before 2004— the inflation-adjusted pre-2004 total yearly hospital revenue would otherwise have been bigger than the current level while in reality it was smaller. Hence, in 2004 Germany switched from a system that paid on average less than $1,000€$ per day to the current system which is mostly flat in length of stay. Therefore, the estimated 0.05 days provides an upper bound for the effect of the 2004 reform on length of stay. Thus, German politicians —like the minister of health quoted in the introduction— took false credit for the fall in length of stay after 2004, since it apparently was just a continuation of the previous trend and not a causal effect of the reform.

The 1983 Medicare reform is widely perceived as having reduced length of stay. While the exact size of the reform's impact is contested, the discussion evolves around magnitudes that are an order of magnitude larger than what my paper attributes to the German reform. For instance, Russell (1989) states:

> Historically, length of stay for the elderly had declined steadily, drifting slowly downward from 13.8 days in 1968 to 10.1 days in 1982. The declines in the two

[23]The remaining 20% were reimbursed using a fixed prospective payment, see Theilen 2004

> years before prospective payment were unusually steep by historical standards, but the decline between 1983 and 1984, when the average dropped by nearly a day, was unprecedented, ample reason to suspect that prospective payment was the cause.

The Medicare reform moved to prospective pay coming from a fee for service system, i.e. a system that pays for each individual provided service. A fee for service system does provide the hospitals with a financial incentive to increase length of stay, since additional services can be provided (including the service of providing a bed, etc. for another night). Since pre-1983 Medicare did pay per service and not per day, my estimates do not directly apply. But a rough back-of-envelope calculation suggests that my estimates from Germany would imply a smaller effect for the 1983 reform than what has been observed:

I approximate the 'reimbursement per day' for the pre-1983 Medicare system by dividing Medicare's total expenditures for hospitals by the total number of Medicare hospital days. I describe the procedure in appendix E. The calculation provides an estimate of the average reimbursement per day, but an overestimate of the marginal reimbursement for keeping a patient another day at the end of her spell, since the costs-weighted total amount of services provided normally increases less than proportionally with the patient's length of stay (Ishak et al. 2012).

I find that Medicare paid approximately 750€ per hospitalization day in 1984, implying that the 1983 reform —which reduced the marginal reimbursement to zero for most patients— cut the marginal reimbursement per day by less than $1,000$€. Hence, my 0.05 days upper bound for the effects of a cut of marginal reimbursement by $1,000$€ stands in sharp contrast to what conventional wisdom attributes to the 1983 reform, again supporting the case that incentives for hospitals are more effective in the U.S. than in Germany.

Implications for policy and future research

My results suggest that future research should investigate whether countries with institutions and cultural norms as in Germany could improve welfare by making hospital reimbursement depend more strongly on length of stay again. Paying hospitals in a way that is more closely tied to the actual incurred costs would better financially compensate hospitals that draw a lot of patients who are comparatively sick conditional on their DRG. The closer tie between costs and reimbursement could have two advantages.

First, compensating hospitals better for patients who are relatively expensive conditional on their DRG would reduce the incentive for the hospitals to discriminate against such patients along the admission margin. Such discrimination is a concern despite this paper's results, since the hospital management might have more ability to manipulate admissions than actual treatment once admitted.[24]

Second, paying hospitals in a way that corresponds more closely to the actual incurred costs might raise welfare, because —given this paper's results— such a policy change would barely affect total treatment volume, but it would reduce the hospitals' financial risk of drawing patients that are comparatively expensive conditional on their DRG. The financial risk would be reduced across hospitals (i.e. the risk of a hospital drawing comparatively sick patients permanently due to, e.g., its location) and within hospitals (i.e. the risk of a hospital drawing many comparatively sick patients within a particular year).

The within-hospital financial risk is non-negligible for smaller hospitals, since fluctuations in length of stay —conditional on the DRG composition— cancel imperfectly over the course of a year.[25] Reducing the financial risk for hospitals would —if hospital owners are risk-averse—allow to reduce equilibrium hospital profits without inducing hospital exit.

For future policy in Germany and countries with similar institutions, my results also suggest that politicians who want to reduce health care costs should consider changing the financial incentives for the doctors directly. In the U.S., various pilot programs test the effect of paying the doctors working in hospitals in a way that encourages costs saving behavior — see, for instance, the evaluation of such a program in Alexander (2016).

The experiences from the 1983 Medicare reform have shaped the way researchers and politicians around the globe think about designing incentives for health care providers. For future research, this paper advises caution when extrapolating reduced form effects from one cultural and institutional setting to another.

[24] Alexander (2016) provides evidence of such discrimination at the admission margin in the U.S.

[25] For instance, the average length of stay at the hospital-year level, residualized for hospital and DRG-year fixed effects, has a standard standard deviation of 0.78 days for the smallest quartile of hospitals (for the largest quartile it is only 0.24), nearly a tenth of the average length of stay.

References

[1] Diane Alexander. How do doctors respond to incentives? unintended consequences of paying doctors to reduce costs. *Working Paper*, 2016.

[2] Jeffrey Clemens and Joshua D Gottlieb. Do physicians' financial incentives affect medical treatment and patient health? *The American Economic Review*, 104(4):1320–1349, 2014.

[3] Robert F Coulam and Gary L Gaumer. Medicare's prospective payment system: a critical appraisal. *Health Care Financing Review*, page 45, 1992.

[4] Mark G Duggan. Hospital ownership and public medical spending. *The Quarterly Journal of Economics*, 115(4):1343–1373, 2000.

[5] Liran Einav, Amy Finkelstein, and Neale Mahoney. Provider incentives and health care costs: Evidence from long term care hospitals. *Working Paper*, 2017.

[6] Paul J Eliason, Paul LE Grieco, Ryan C McDevitt, James W Roberts, et al. Strategic patient discharge: The case of long-term care hospitals. *Working Paper*, 2016.

[7] D Franz, HE Helling, H Bunzemeier, M Heumann, and N Roeder. Wer kodiert im Krankenhaus? Ergebnisse einer Umfrage zur Durchführung der DRG-Kodierung in Krankenhäusern Nordrhein-Westfalens. *Gesundheitsökonomie & Qualitätsmanagement*, 16(06):363–368, 2011.

[8] Richard Gunderman. How hospitals discourage doctors: A step by step guide. *KevinMD.com (09/26/2014)*.

[9] Stuart Guterman and Allen Dobson. Impact of the medicare prospective payment system for hospitals. *Health Care Financing Review*, 7(3):97, 1986.

[10] K Jack Ishak, Marilyn Stolar, Ming-yi Hu, Piedad Alvarez, Yamei Wang, Denis Getsios, and Gregory C Williams. Accounting for the relationship between per diem cost and los when estimating hospitalization costs. *BMC health services research*, 12(1):439, 2012.

[11] Yan S Kim, Eric C Kleerup, Patricia A Ganz, Ninez A Ponce, Karl A Lorenz, and Jack Needleman. Medicare payment policy creates incentives for long-term care hospitals to time discharges for maximum reimbursement. *Health Affairs*, 34(6):907–915, 2015.

[12] Henrik Jacobsen Kleven. Bunching. *Annual Review of Economics*, 8:435–464, 2016.

[13] Gerald F Kominski and Christina Witsberger. Trends in length of stay for medicare patients: 1979-87. *Health care financing review*, 15(2):121, 1993.

[14] Katharine R Levit, Helen Lazenby, Daniel R Waldo, and Lawrence M Davidoff. National health expenditures, 1984. *Health Care Financing Review*, 7(1):1, 1985.

[15] Jörn-Steffen Pischke and Hannes Schwandt. Poorly measured confounders are more useful on the left than on the right. *Working Paper*, 2015.

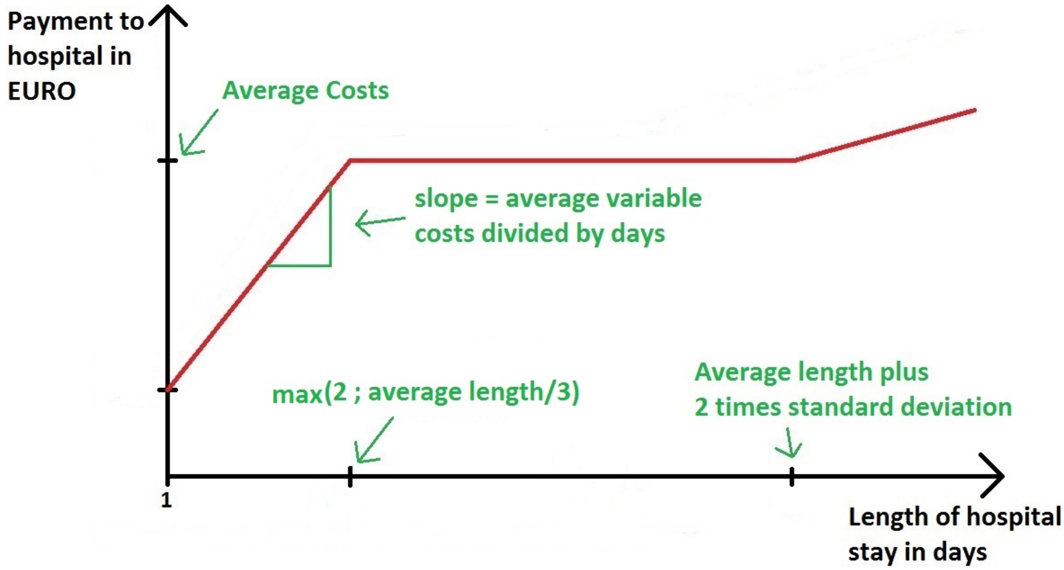
[16] Louise B Russell and Carrie Lynn Manning. The effect of prospective payment on medicare expenditures. *New England Journal of Medicine*, 320(7):439–444, 1989.

[17] Adam Sacarny. Technological diffusion across hospitals: The case of a revenue-generating practice. *Working Paper*, 2016.

[18] Thilo Theilen. Der preis fuer krankenhausleistungen - entwicklung der regulierung in der bundesrepublik deutschland. *Manuscript*, 2004.

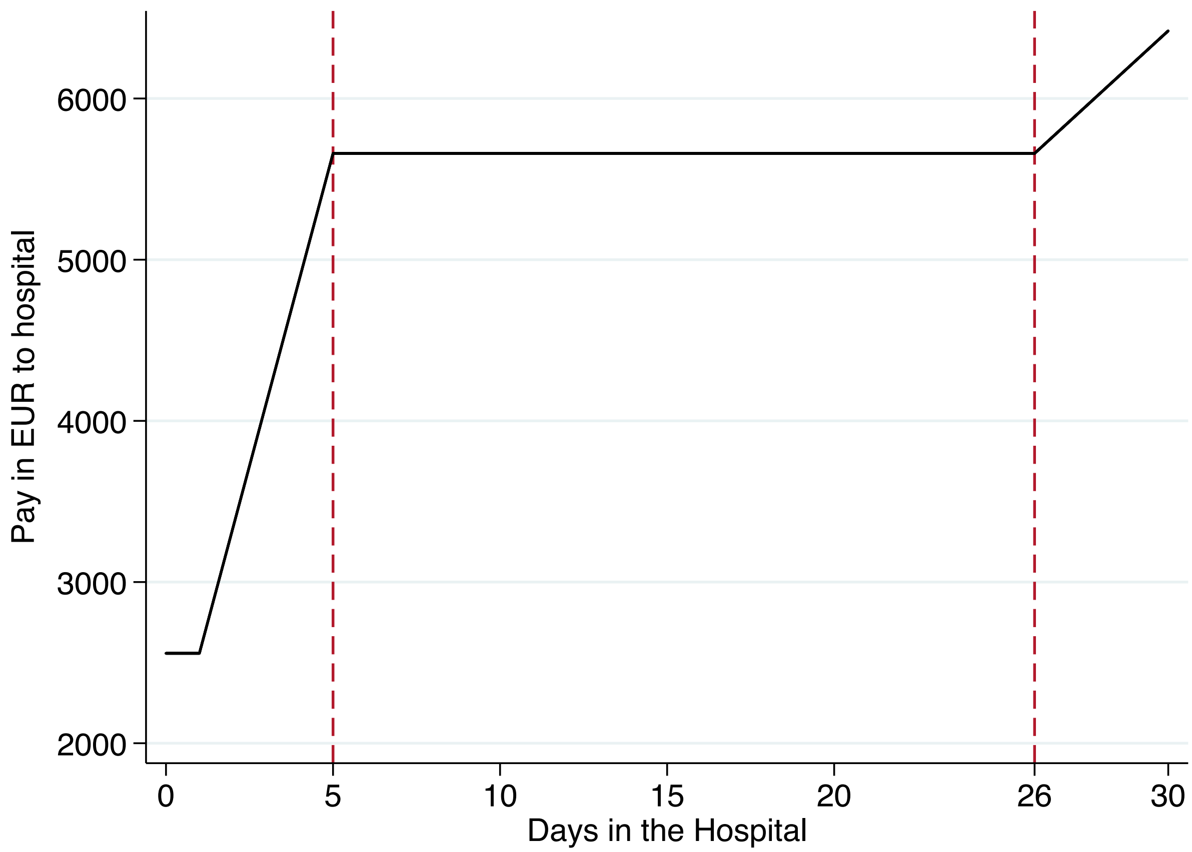
[19] Christopher Weaver, Anne Wilde Matthews, and Tom McGinty. Hospital discharges rise at lucrative times. *Wall Street Journal (2/17/2015).*

Figure 1: Within-DRG Payment Schedule as a Function of Length of Stay - Generic



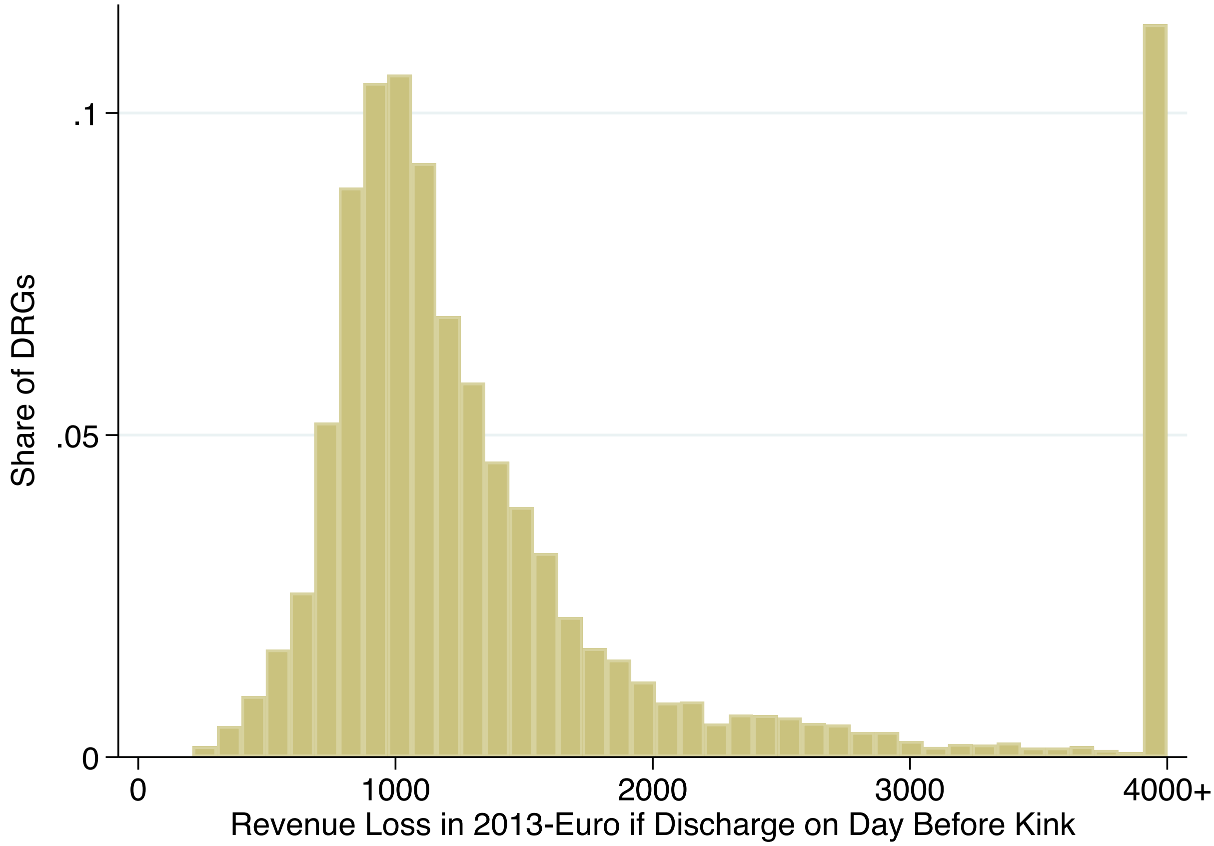
The parameters of the payment scheme are —as the DRG definitions— based on the hospital cost data from two years before. The payment increases linearly until a third of the average length of stay (rounded and measured two years prior) of all patients in this DRG is reached (but at least until day 2 is reached). The slope is determined by dividing average variable costs (that is, total costs excluding costs of major procedures, e.g., bypass surgery) of all patients with this DRG by the number of days at which the kink occurs (again, costs measured two years prior). After the kink, the payment schedule remains flat until the average plus two times the standard deviation of the length of stay of all patients with this DRG two years prior is reached. From then on it increases again linearly.

Figure 2: Within-DRG Payment Schedule as a Function of Length of Stay - Example DRG I51Z



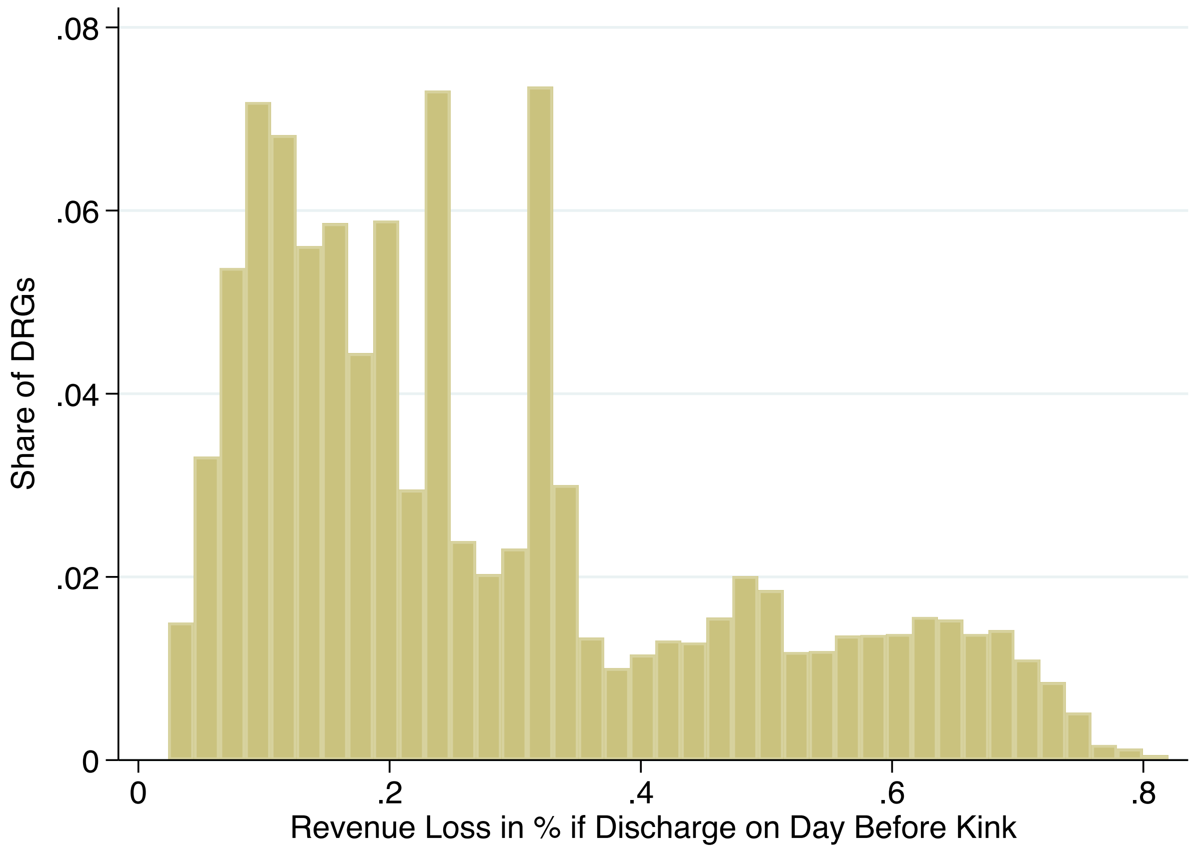
Example of a payment schedule. The graph shows the pay for a patient with DRG I51Z in 2005 as a function of number of days in the hospital. DRG I51Z is for 'other procedures at the hip joint or femur, without major complications'. Pay does shift proportionally vertically across locations - the graph corresponds to the average proportional shift factor for the state of Hamburg.

Figure 3: Histogram of the Revenue Loss when Discharging the Day Before the Kink Instead of the Kink Day (in €)



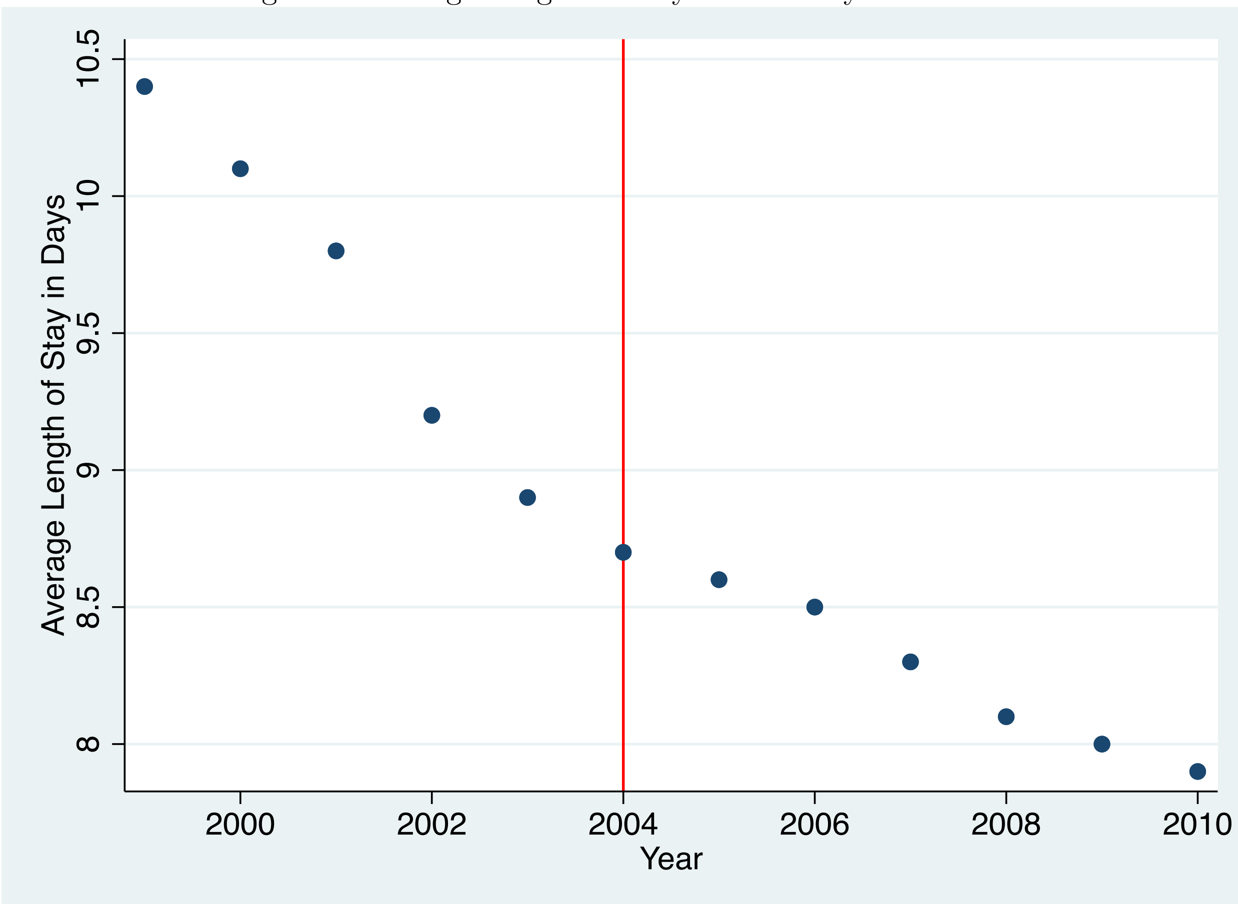
For each DRG, I calculate the change in slope at the lower kink in 2013-Euro, i.e. how much money the hospital loses by discharging the day before the kink day instead of on the kink day. The plot is a histogram of this measure across all DRGs and all years 2005-2013.

Figure 4: Histogram of the Revenue Loss when Discharging the Day Before the Kink Instead of the Kink Day (in Percent)



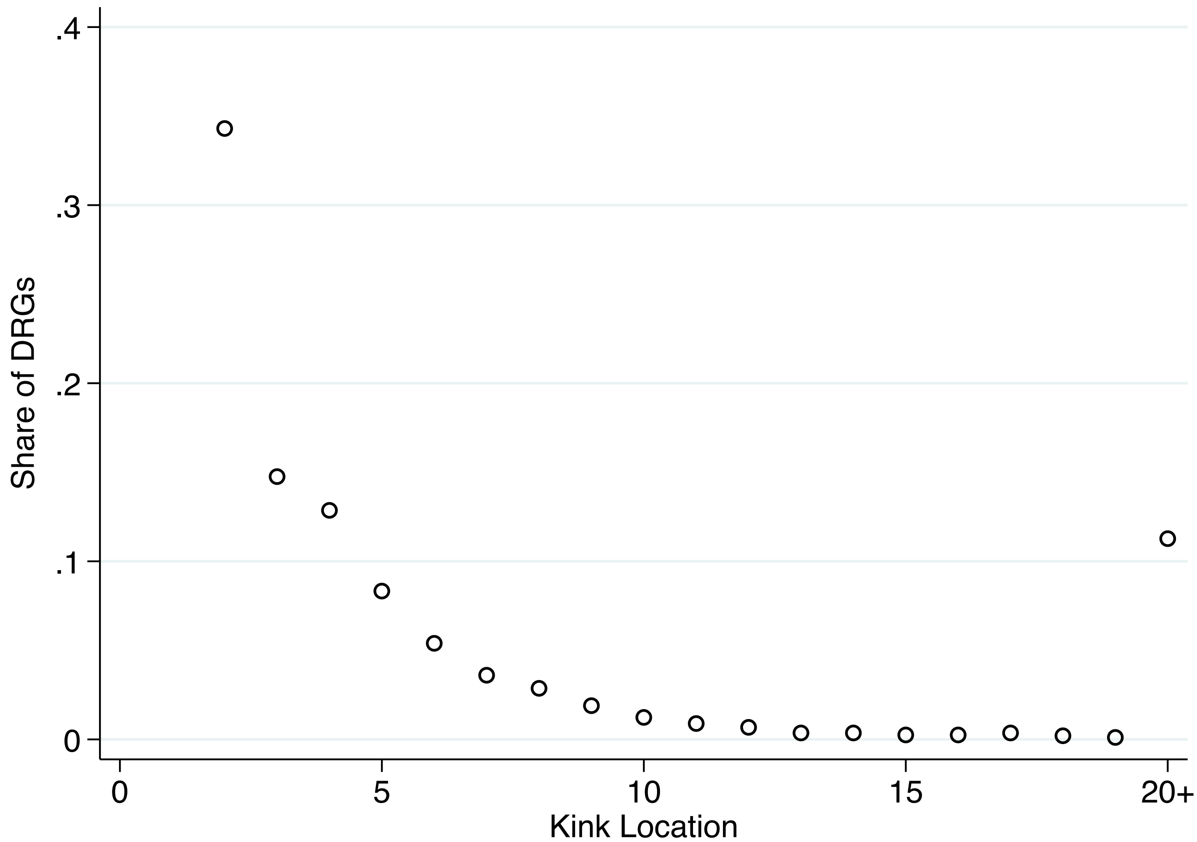
For each DRG I calculate the ratio of the slope to the left of the lower kink and the amount that is paid to the hospital in the flat part of the schedule - i.e. the share of revenue that is lost by discharging the patient a day earlier than the kink day instead of on the kink day. The graph shows a histogram of this measure across all DRGs and all years 2005-2013.

Figure 5: Average Length of Stay in Germany Over Time



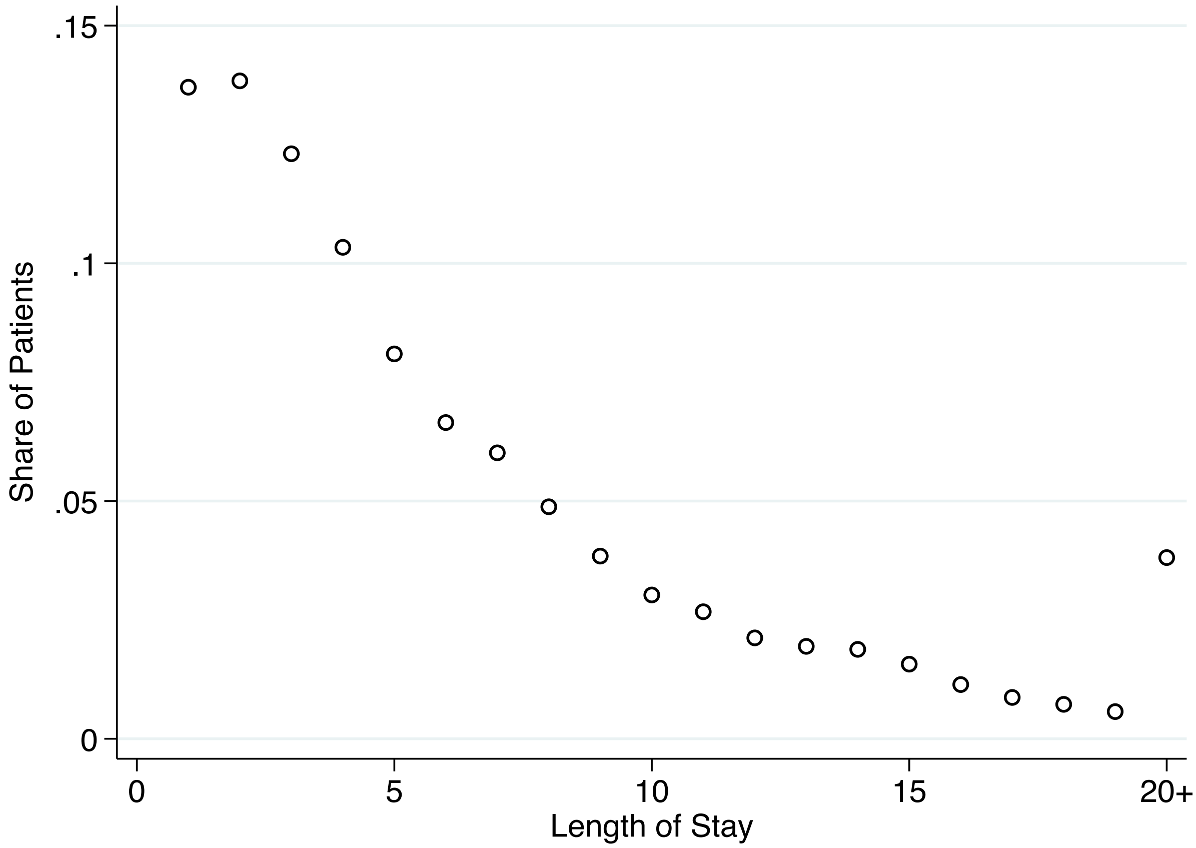
Source: Krankenhausstatistik (Hospital Statistic)
The averages include cases not covered in the DRG system such as psychiatric cases.
The source is not the OECD and uses different definitions, which is why the numbers differ from the OECD numbers I use in the international comparisons in the paper.

Figure 6: Histogram Kink Locations



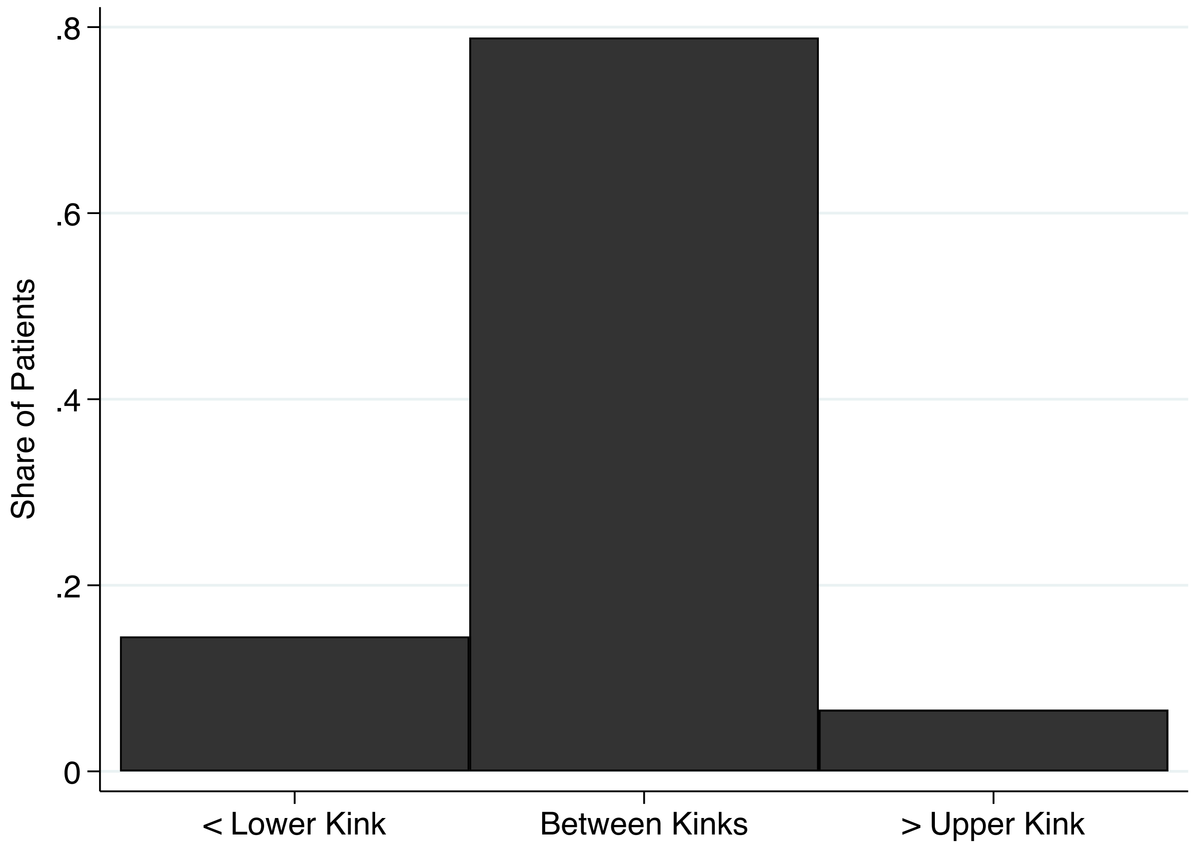
Shows the share of DRGs with the respective kink location (for the lower kink) for years 2005-2013. E.g. the example DRG in Figure 2 has the kink location 5 days.

Figure 7: Histogram Length of Stay



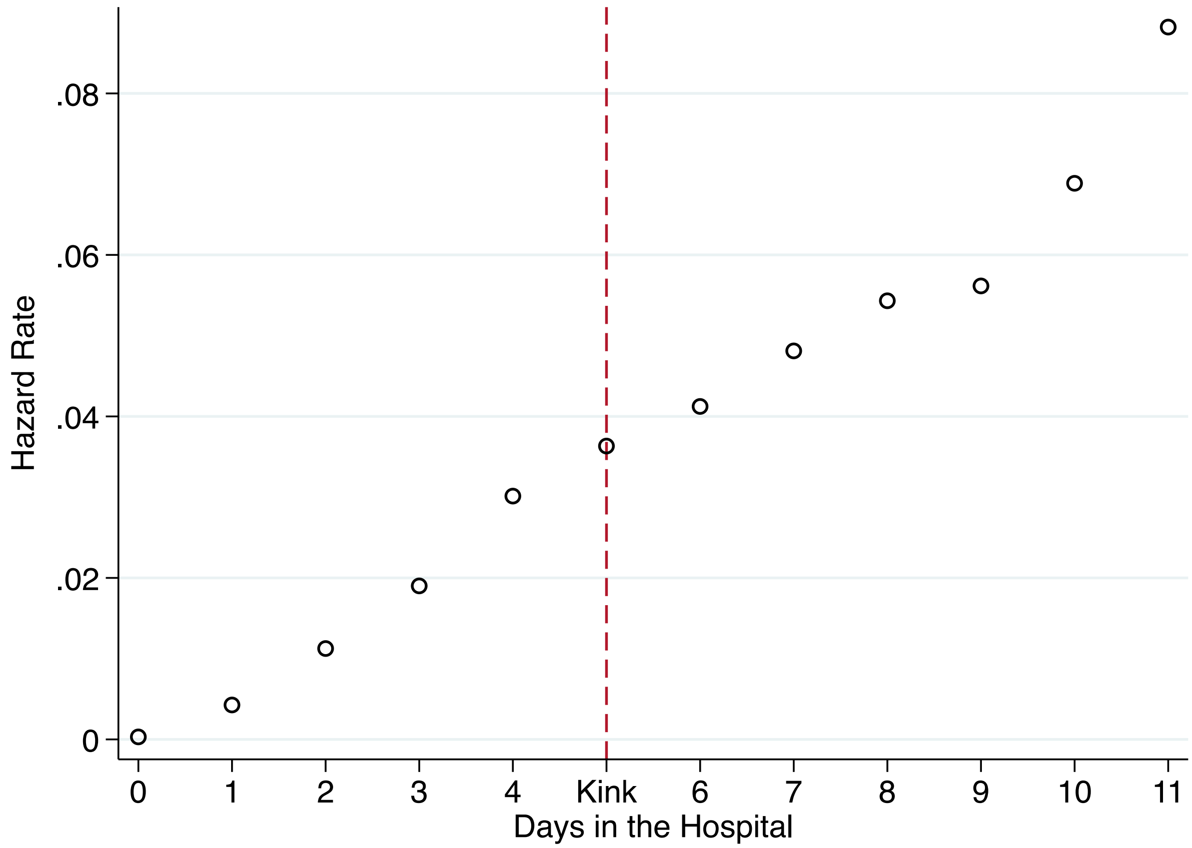
Shows the share of patients staying 1, 2, 3 etc days in the hospitals across all years 2005-2013.

Figure 8: Histogram of Length of Stay Relative to the Patient's Kink Location



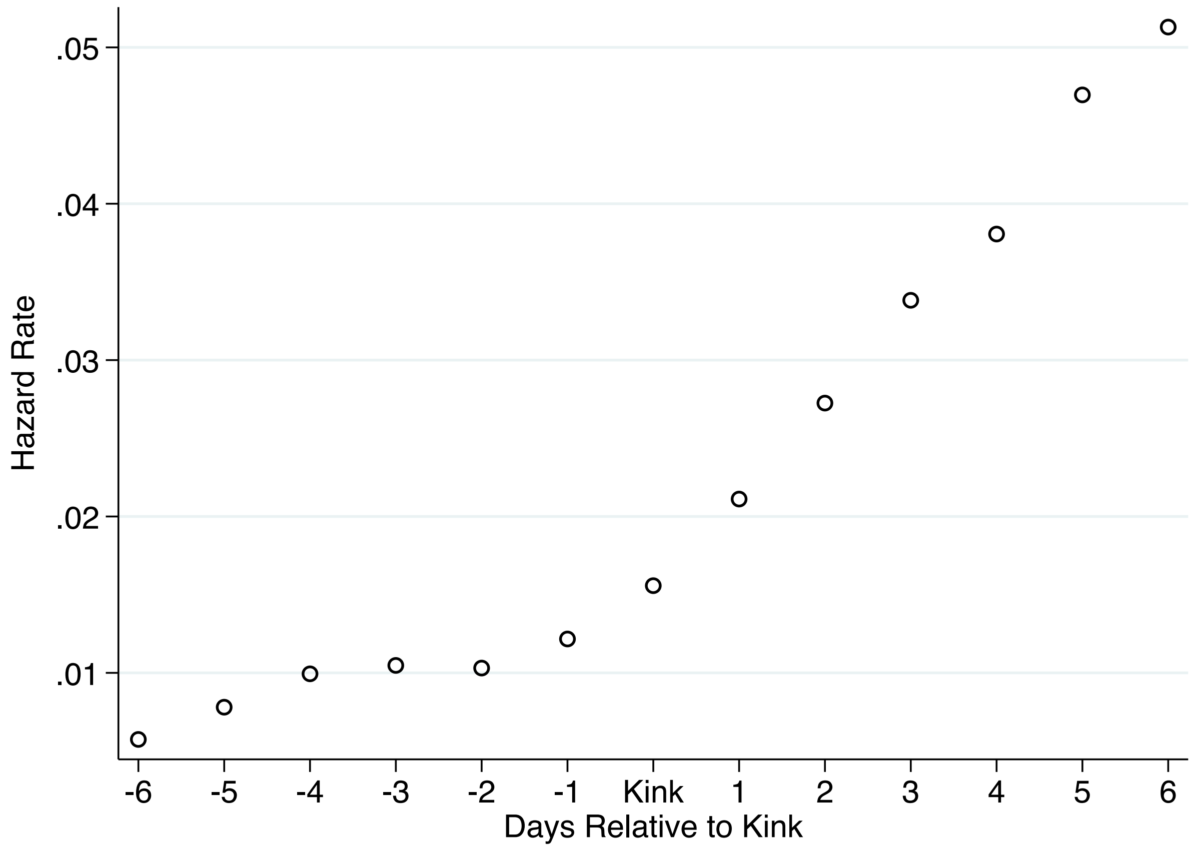
Shows the share of patients getting discharged earlier than the lower kink of their respective DRG, between the two kinks or after the right kink of their respective DRG. Graph is a histogram for all patients in the data, i.e. 2005-2013

Figure 9: Hazard Rates Around Kink - Example DRG I51Z



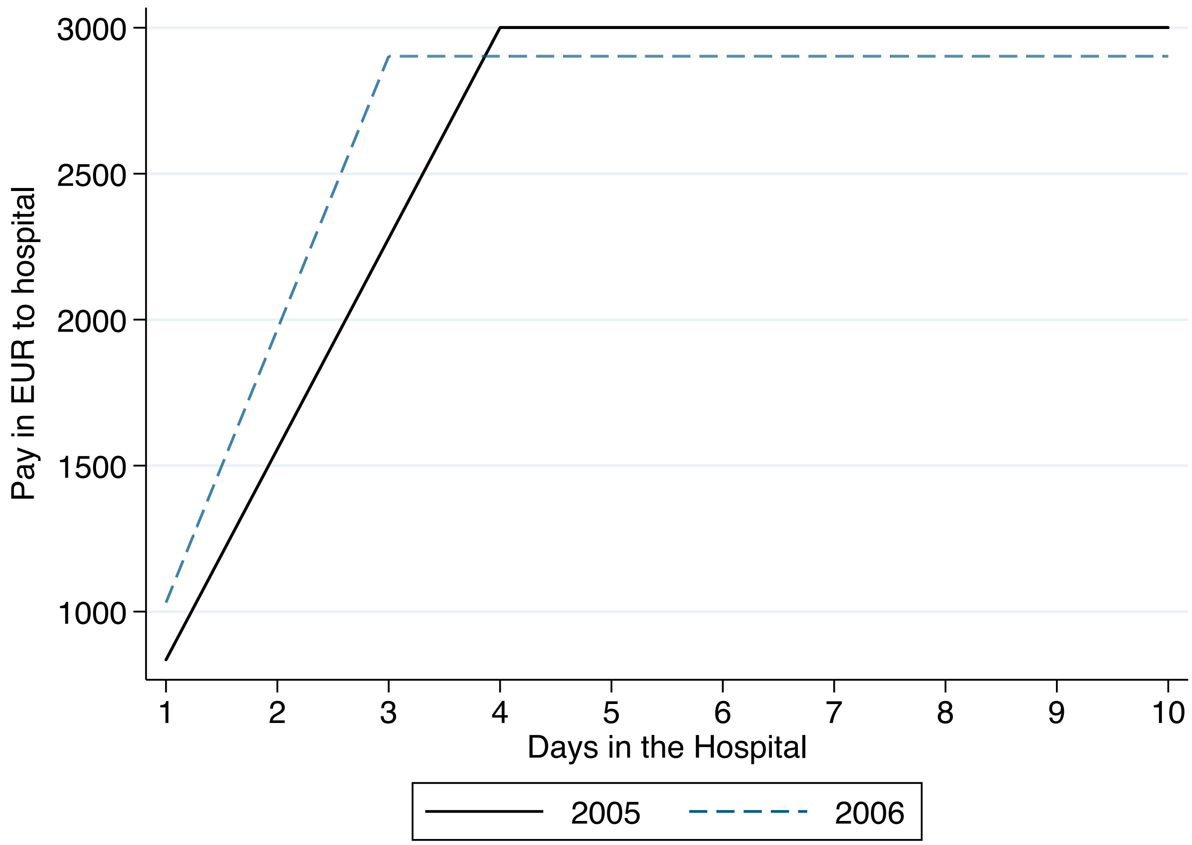
Hazard rates for the example DRG I51Z with payment schedule shown in Figure 2. Graphs show the hazard rates for hospital discharge in year 2005 for DRG I51Z after 0, 1, 2, etc days in the hospital.
DRG I51Z is for 'other procedures at the hip joint or femur, without major complications'

Figure 10: Hazard Rates Around Kink - Pooled Sample



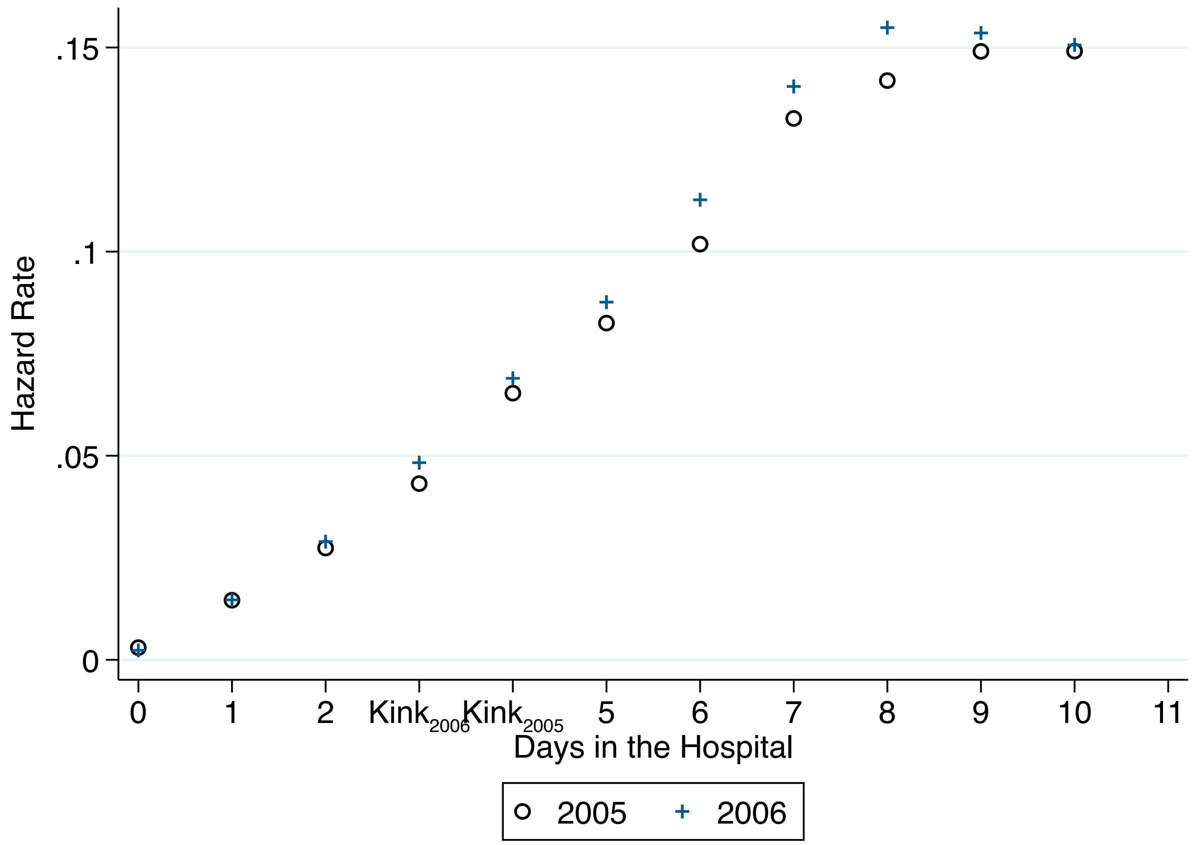
I normalize the length of stay for each patient with respect to her DRG's lower kink location. E.g., if a patient stays 3 days and her kink is at 4 days, the normalized length of stay is -1. I then pool all cases and calculate the hazard rates w.r.t. the normalized length of stay. I then plot these hazard rates against the normalized length of stay. I restrict the sample to DRGs with kink location of at least 6 days. Graph looks similar for alternative restrictions.

Figure 11: Payment Schedule Change - Example DRG H62A for Which the Kink Location Goes Down by One Day



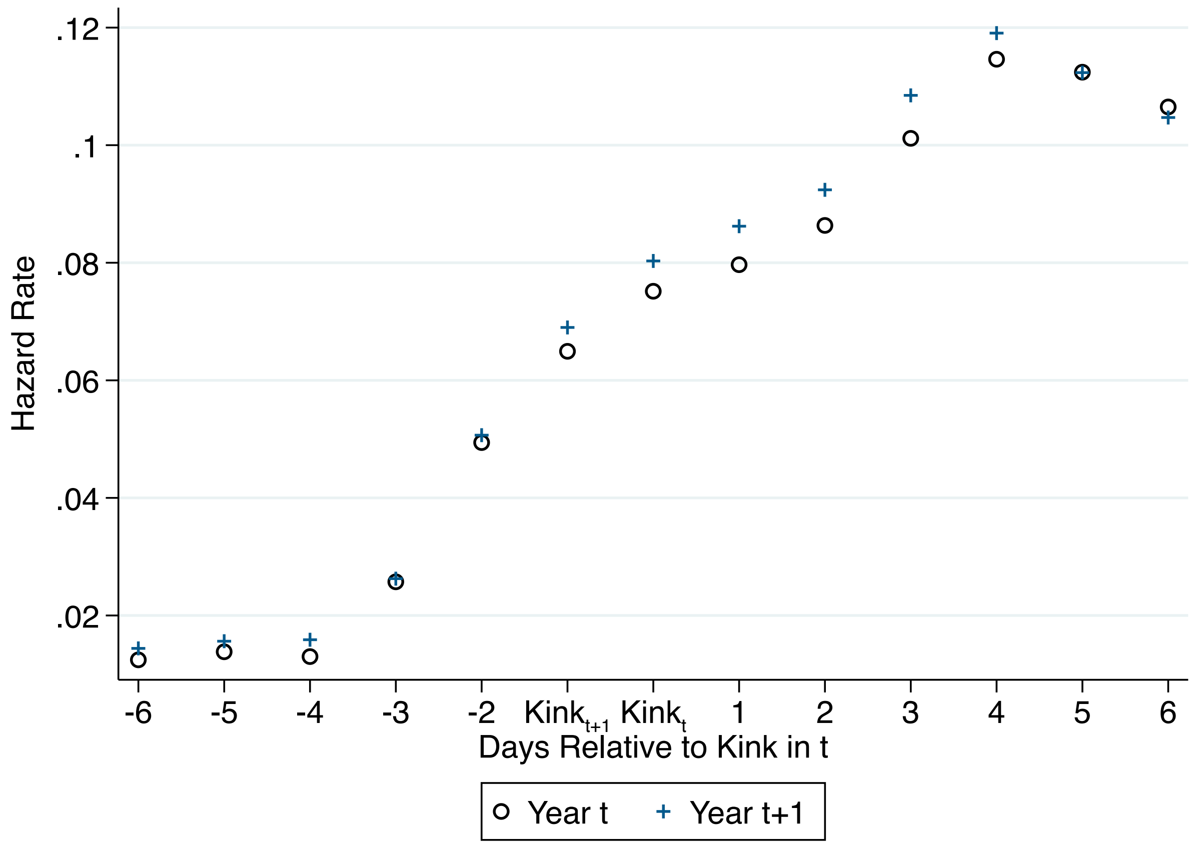
Example of how payment schedules change from one year to the next for the same DRG. The graph shows the pay to the hospital for a patient with DRG H62A in 2005 and 2006 as a function of the number of days in the hospital. The kink location changes from 4 to 3 days. DRG H62A is for 'Diseases of the pancreas, except for malignant neoformation with acute pancreatitis'. Pay does shift proportionally vertically across locations - the graph corresponds to the average proportional shift factor for the state of Hamburg.

Figure 12: Hazard Rates Around Kink from One Year to the Next - Example DRG H62A for Which the Kink Location Goes Down by One Day



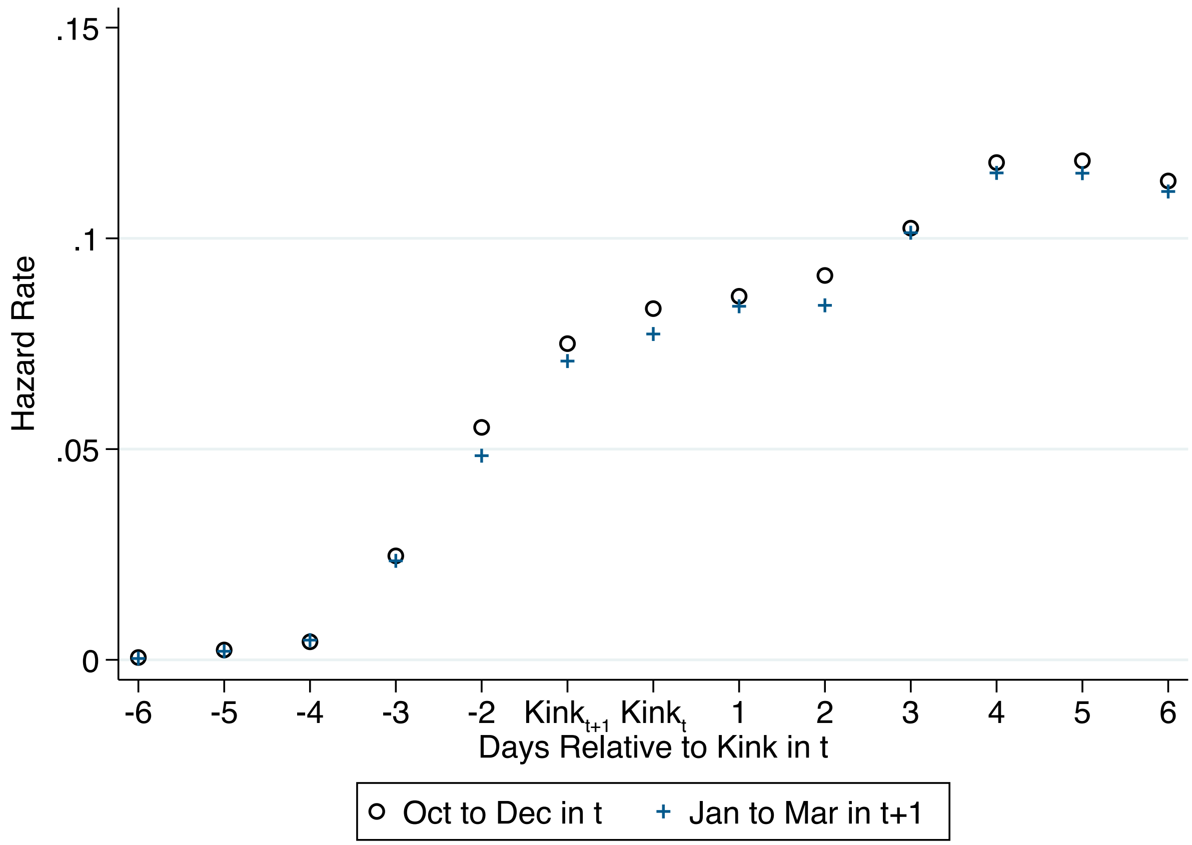
Shows the hazard rates for the example DRG H62A in 2005 and 2006 - see Figure 11 for this DRG's payment schedule in 2005 and 2006 (the kink location changes from 4 days in 2005 to 3 days in 2006). The graphs plots the hazard rates for discharge from hospital in year 2005 and year 2006 after 0, 1, 2, etc days in the hospital.
DRG H62A is for 'Diseases of the pancreas, except for malignant neoformation with acute pancreatitis'

Figure 13: Hazard Rates Around Kink from One Year to the Next if the Kink Location Goes Down by One Day - Pooled Sample



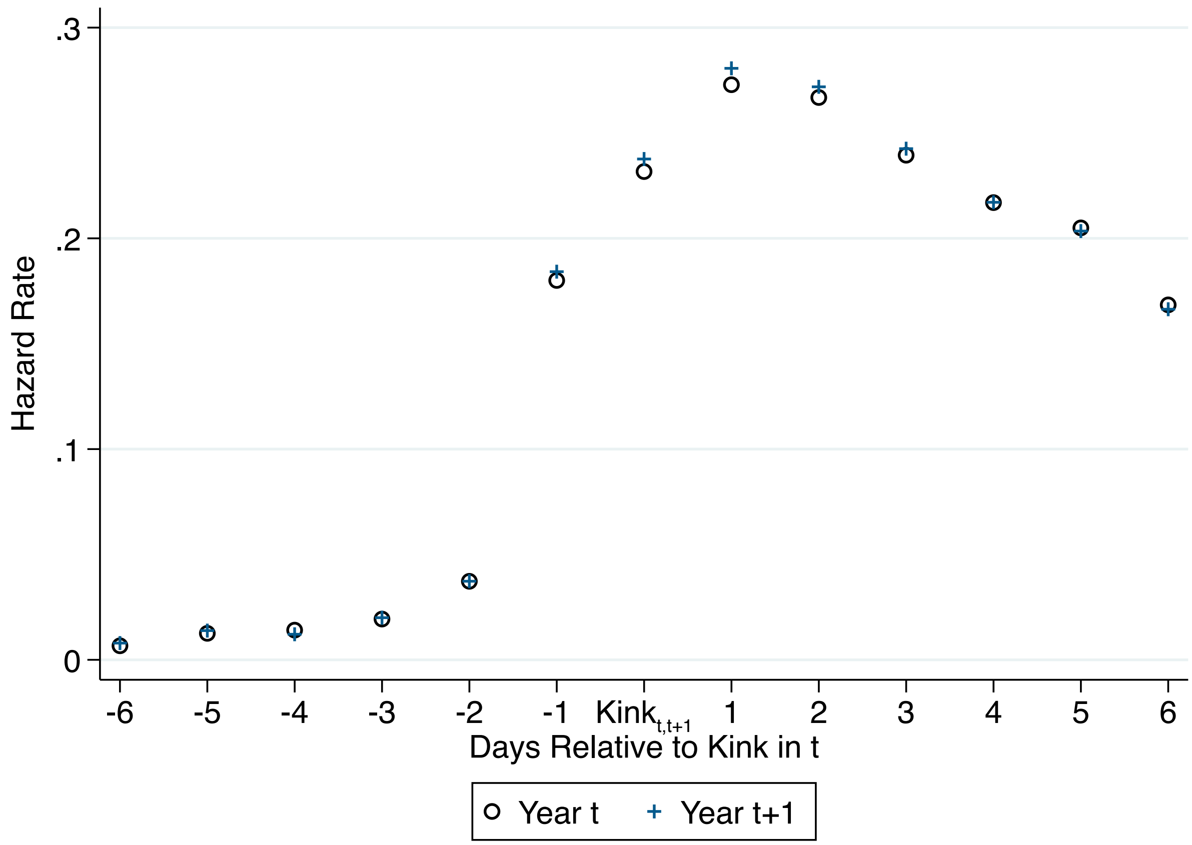
I restrict the sample to DRGs that are comparable from t to t+1 and for which the kink point went down one day from t to t+1. For each patient, I normalize length of stay by the patient's DRG's kink location in t (e.g., normalized length of stay is -1 if she stayed 4 days, but her DRG's kink in t is at 5 days). I then pool the sample and calculate hazard rates w.r.t. the normalized length of stay. That is, the hazard rate at 0 is the hazard rate at the year t's kink and at -1 is the hazard rate at year t+1's kink.
Note that the hazard rates to the left of the kink do not necessarily behave smoothly, because e.g. a DRG with kink in t at 3 days has a hazard of 0 at -3 by construction. But the composition effects are identical for t and t+1, so the comparison is still valid.

Figure 14: Hazard Rates Around Kink from One Year to the Next if the Kink Location Goes Down by One Day- Pooled Sample - October until March



(Identical to Figure 13, except restricted to the narrower time window from October of year t to March of year t+1)
I restrict the sample to DRGs that are comparable from t to t+1 and for which the kink point went down one day from t to t+1. For each patient, I normalize length of stay by the patient's DRG's kink location in t (e.g., normalized length of stay is -1 if she stayed 4 days, but her DRG's kink in t is at 5 days). I then pool the sample and calculate hazard rates w.r.t. the normalized length of stay. That is, the hazard rate at 0 is the hazard rate at the year t's kink and at -1 is the hazard rate at year t+1's kink.
Note that the hazard rates to the left of the kink do not necessarily behave smoothly, because e.g. a DRG with kink in t at 3 days has a hazard of 0 at -3 by construction. But the composition effects are identical for t and t+1, so the comparison is still valid.

Figure 15: Hazard Rates Around Kink from One Year to the Next if the Kink Location Does Not Change - Pooled Sample



(Identical to Figure 13, except restricted to DRGs for which the kink location stays the same from year t to year t+1)
I restrict the sample to DRGs that are comparable from t to t+1 and for which the kink point does not change t to t+1. For each patient, I normalize length of stay by the patient's DRG's kink location (which is the same in t and t+1). E.g., normalized length of stay is -1 if she stayed 4 days, but her DRG's kink is at 5 days). I then pool the sample and calculate hazard rates w.r.t. the normalized length of stay.
Note that the hazard rates to the left of the kink do not necessarily behave smoothly, because e.g. a DRG with kink at 3 days has a hazard of 0 at -3 by construction. But the composition effects are identical for t and t+1, so the comparison is still valid.

Table 1: Summary Statistics Analysis Sample vs Remaining Sample

	Analysis Sample	Remaining Cases	Difference
Length of Stay	14.05 (16)	6.9 (7.88)	7.15 (0.0091)
Age	52.2 (26.76)	53.58 (25.47)	-1.38 (0.0293)
Share Female	0.5 (0.5)	0.54 (0.5)	-0.04 (0.0006)
Year of Admission	2009 (2.45)	2009.11 (2.58)	-0.11 (0.003)
Month of Admission	6.47 (3.41)	6.43 (3.46)	0.04 (0.004)
N	761505	130 million	

The table presents summary statistics for the analysis sample of patients (i.e. restricted to DRGs that are comparable from one year to the next and for which the kink location changes) as well as for the remaining observations. For month of admission 1 corresponds to January and 12 to December.

Table 2: How Often Do Kink Locations Change?

Kink in t	Kink in t+1	DRGs	Patients
2	3	15	23 014
3	2	33	153 470
3	4	8	13 539
4	3	34	169 243
4	5	7	3 957
5	4	23	68 280
5	6	5	11 204
6	4	2	785
6	5	11	25 407
6	7	11	4 728

The sample is restricted to DRGs that are comparable from one year to the next and for which the kink location changes. For each combination of kink location in t and kink location in t+1 the table reports how many DRGs feature this change in kink location from one year to the next and how many patients are grouped into such a DRG. I restrict it to DRGs with a kink location of at most 6 in t. Patients as well as DRGs can appear several times - e.g. because a DRG has kink location 2 in t, then 3 in t+1 and then 2 again in t+2. In that case the DRG (and the patients grouped into this DRG) are counted twice in t+1: once they appear in the 2 to 3 row and once in the 3 to 2 row.

Table 3: Causal Effect of Increasing Marginal Reimbursement per Day in Hospital by 1,000€ on Length of Stay

Dep. Var.	Length of Stay						
Causal Effect	-0.004 (0.017)	0.026** (0.01)	0.015 (0.01)	0.019* (0.009)	0.066* (0.027)	0.011 (0.008)	0.028 (0.018)
Year-DRG-FE	no	no	yes	yes	yes	yes	yes
weighted	no	yes	no	yes	yes	yes	yes
Patients	6149229	6149229	6144604	6144604	791623	5762258	2623847
Clusters	107	107	100	100	27	87	48

'Causal Effect' refers to the effect of increasing marginal pay for another day in the hospital by 1,000 2013-Euro. Standard errors are clustered at the DRG level using 400 bootstrap draws. The sample is restricted to DRGs that are comparable from one year to the next and for which the kink location changes. Moreover, Column 5 is restricted to DRGs that feature a decrease in kink location over time, while column 6 is restricted to those featuring an increase. Column 7 only uses post 2010 data. Some DRGs experience both, decreases and increases, which is why the number of clusters in columns 5 and 6 does not add up to the number in columns 1 to 4. Weighting refers to weighting by the number of patients still present in the hospital that day. pvalues: * $<$ .05 ** $<$ .01

Table 4: Causal Effect of Increasing Marginal Reimbursement by $1,000€$ per Day in Hospital on Length of Stay - for Transfers

Dep. Var.		Transfers		
Causal Effect	<0.01	<0.01	<0.01	<0.01
Standard Error	<0.01	<0.01	<0.01	<0.01
Year-DRG-FE	no	no	yes	yes
weighted	no	yes	no	yes
Patients	1205114	1205114	1205036	1205036
Clusters	140	140	135	135

Causal Effect refers to the effect of increasing marginal pay for another day in the hospital by 1,000 2013-Euro. Standard errors are clustered at the DRG level. The sample is restricted to transfers and to DRGs that are comparable from one year to the next and for which the kink location - in the case of transfers, the average duration two years prior - changes. Weighting refers to weighting by the number of patients still present in the hospital that day. pvalues: * $<$ .05 ** $<$ .01

Table 5: Fixed Effect Regression for the Effect of the Kink Location on Length of Stay in Days

Dep. Var.	Length of Stay				
Kink location	0.203 (0.111)	-0.037 (0.136)	-0.003 (0.126)	0.01 (0.1)	-0.053 (0.092)
Month FE	yes	yes	yes	no	no
Month-Hospital FE	no	no	no	yes	yes
DRG FE	yes	yes	yes	yes	yes
avg duration linear	no	yes	yes	yes	yes
avg duration quadr.	no	no	yes	yes	yes
Control set 1	no	no	no	yes	yes
Control set 2	no	no	no	no	yes
Observations	929376	929376	929376	904527	904527
Cluster	147	147	147	147	147
Mean Dep.	14.76	14.76	14.76	14.76	14.76

Standard errors are clustered at the DRG level. Outcome variable is length of stay (winsorized at the 99th percentile). 'avg duration linear' and 'avg duration quadr.' refer to a linear and a quadratic term in the average duration from two years prior. Control set 1 covers DRG-specific indicators for gender and slopes for age. Control set 2 covers number of diagnoses and number of procedures. 'Mean Dep.' refers to the mean of the dependent variable. pvalues: * $<$ .05 ** $<$ .01

Table 6: Fixed Effect Regression for the Effect of the Kink Location on Covariates - Age

Dep. Var.		Age		
Kink location	-0.036 (0.126)	-0.042 (0.236)	-0.048 (0.252)	0.297 (0.19)
Month FE	no	no	no	yes
DRG FE	yes	yes	yes	yes
avg duration linear	no	yes	yes	yes
avg duration quadr.	no	no	yes	yes
Observations	929376	929376	929376	929376
Cluster	147	147	147	147
Mean Dep.	51.83	51.83	51.83	51.83

Standard errors are clustered at the DRG level. Outcome variable is age. 'avg duration linear' and 'avg duration quadr.' refer to a linear and a quadratic term in the average duration from two years prior. 'Mean Dep.' refers to the mean of the dependent variable. pvalues: * $<$ .05 ** $<$ .01

Table 7: Fixed Effect Regression for the Effect of the Kink Location on Covariates- Gender

Dep. Var.		Gender		
Kink location	0 (0.002)	0 (0.003)	0 (0.003)	0.001 (0.003)
Month FE	no	no	no	yes
DRG FE	yes	yes	yes	yes
avg duration linear	no	yes	yes	yes
avg duration quadr.	no	no	yes	yes
Observations	929376	929376	929376	929376
Cluster	147	147	147	147
Mean Dep.	0.49	0.49	0.49	0.49

Standard errors are clustered at the DRG level. Outcome variable is an indicator for female. 'avg duration linear' and 'avg duration quadr.' refer to a linear and a quadratic term in the average duration from two years prior. 'Mean Dep.' refers to the mean of the dependent variable. pvalues: * $<$ .05 ** $<$ .01

Table 8: Fixed Effect Regression for the Effect of the Kink Location on Covariates- Number Diagnoses

Dep. Var.		Diagnoses		
Kink location	-0.19** (0.033)	-0.192** (0.052)	-0.177** (0.065)	-0.021 (0.047)
Month FE	no	no	no	yes
DRG FE	yes	yes	yes	yes
avg duration linear	no	yes	yes	yes
avg duration quadr.	no	no	yes	yes
Observations	929376	929376	929376	929376
Cluster	147	147	147	147
Mean Dep.	6.72	6.72	6.72	6.72

Standard errors are clustered at the DRG level. Outcome variable is the number of diagnoses. 'avg duration linear' and 'avg duration quadr.' refer to a linear and a quadratic term in the average duration from two years prior. 'Mean Dep.' refers to the mean of the dependent variable. pvalues: * $<$ .05 ** $<$ .01

Table 9: Fixed Effect Regression for the Effect of the Kink Location on Covariates- Number Procedures

Dep. Var.	Procedures			
Kink location	0.054 (0.037)	0.022 (0.068)	0.05 (0.067)	-0.008 (0.06)
Month FE	no	no	no	yes
DRG FE	yes	yes	yes	yes
avg duration linear	no	yes	yes	yes
avg duration quadr.	no	no	yes	yes
Observations	929376	929376	929376	929376
Cluster	147	147	147	147
Mean Dep.	6.49	6.49	6.49	6.49

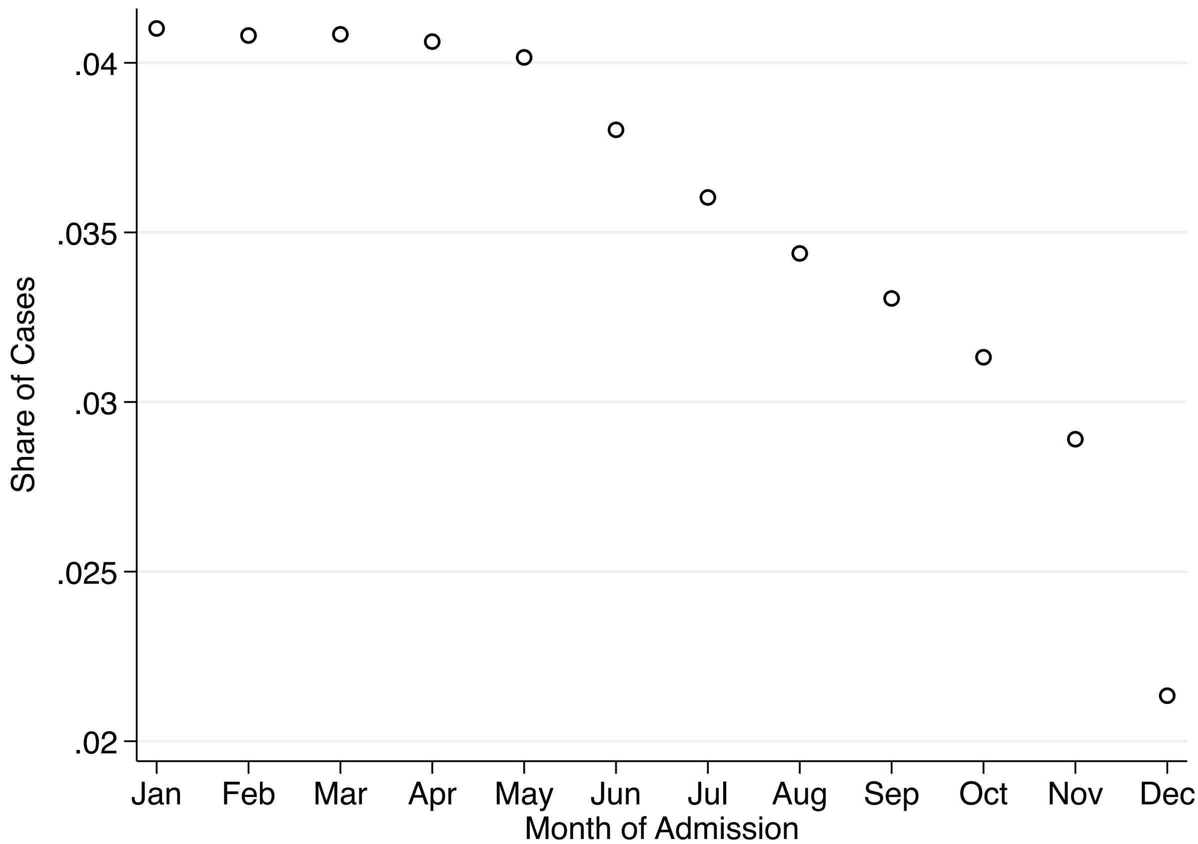
Standard errors are clustered at the DRG level. Outcome variable is the number of procedures. 'avg duration linear' and 'avg duration quadr.' refer to a linear and a quadratic term in the average duration from two years prior. 'Mean Dep.' refers to the mean of the dependent variable. pvalues: * $<$ .05 ** $<$ .01

A Measurement in Length of Stay

In the main text, I argue that the measurement error problem for the billed number of days that is introduced by successful audits is less severe for cases from the end of the calendar year. The reason for this seasonality is because the data for each year is collected on March 31 of the following year. Only if the audit has been completed at that point in time, it will change the billed number of days in the data. Since the median time until an audit completion is over 3 months, the share of completed audits is much smaller for December cases than for patients admitted in January. Therefore, the billed number of days for December cases is often times still pre-audit and identical to the difference between discharge and admission date.

Figure 16 shows the share of cases for which the billed number of days is smaller than the difference between discharge and admission date. The share of cases with diverging numbers is clearly higher for cases from early in the year than for cases from later in the year, supporting the case that audits are less of a problem later for data from later in the calendar year.

Figure 16: Deviating billed number of days and difference between discharge and admission date - by admission month



The graph shows (separately by admission month) the share of cases for which the billed number of days in the data deviates from the difference between discharge and admission date.

B Generalized Model

Summary

Here I extend the model from the main text with risk-averse hospitals as well as the possibility for audits by health insurers. The key difference in the results is that the causal effect of changing marginal pay under these generalized assumptions is identified for a reform that changes marginal pay per day while adjusting a fixed payment component to the hospital in a way that keeps hospital profits constant in equilibrium (i.e. taking into account that the hospitals will respond to the changed marginal pay as well as to the changed fixed payment component).

Setup

The hospital admits a continuum of patients of type θ_i who stay d_i days and enjoy health benefit $h\left(d_i, \theta_i\right)$ which is concave in the number of days in the hospital. Patients with higher θ_i are sicker and benefit more from staying in the hospital for longer. That is, $\frac{\partial^2 h(d_i,\theta_i)}{\partial d_i^2} < 0$ and $\frac{\partial^2 h(d_i,\theta_i)}{\partial d_i \partial \theta_i} > 0$. Since there are no functional form assumptions on how θ_i affects $h\left(d_i, \theta_i\right)$, one can assume a uniform distribution $\theta_i \sim U\left[0,1\right]$ without loss of generality.

The hospital receives payment $P\left(d_i\right)$ and incurs costs $C\left(d_i\right)$. The hospital gains utility from profits and from its patients' health (either because of an intrinsic concern for their patients' health or because they fear lawsuits or reputational costs if patients are mistreated). The hospital is potentially risk-averse and values profits relative to patient health benefits according to $\lambda_h \cdot U\left(profits\right)$ with $\lambda_h \geq 0$ and $U'' \leq 0$. Since in practice, the medical personnel and not the hospital shareholders make the discharge decision, the hospital faces an agency problem in implementing its objective function. I model this agency problem in the form of parameter $0 \leq \lambda_d \leq 1$ which dampens the degree to which profits are taken into account.

Moreover, the health insurers can audit bills with probability $\gamma \epsilon \left(0,1\right)$ and change the billed number of days to a patient-specific $\tilde{d}\left(\theta_i\right)$ with $\tilde{d}'\left(\theta_i\right) > 0$. Hence, expected audit costs for patient type θ_i are $\gamma\left[P\left(d_i\right) - P\left(\widetilde{d}\left(\theta_i\right)\right)\right]$.

That is, the patients' length of stay is determined by the solution to

$$\max_{\{d_i(\theta_i)\epsilon\mathbb{N}\}} \lambda_d\lambda_h \int_i U\left(P\left(d_i\right) - C\left(d_i\right) - \gamma\left[P\left(d_i\right) - P\left(\widetilde{d}\left(\theta_i\right)\right)\right]\right) + \int_i h\left(d_i, \theta_i\right)$$

If the agency problem were modeled in a different way, the bunching design would not necessarily identify the causal effect of interest anymore, because the kinks might have effects on hazard rates away from the kink point. I discuss this point in more detail in the results section and provide evidence against such effects of the kink on hazard rates away from the kink.

Note also that changes in admission and coding behavior - while interesting subjects to study in their own right - do not threaten the validity of this paper's findings. If anything, adjustments in coding and admission behavior would lead to an overestimate of the bunching mass in my setting. This is because the incentive to deny admission or to upcode to a different diagnosis with a different kink location is smallest for those patients who would otherwise be discharged on the profit maximizing kink day.

Optimal Hospital Behavior

I assume that $P\left(d_i\right)$, $C\left(d_i\right)$ and $h\left(d_i, \theta_i\right)$ are shaped such that the objective function is globally concave. At baseline, consider a linear payment schedule $P^{baseline}\left(d_i\right) = \bar{p} + p \cdot d_i$.

Optimization than amounts to choosing cutoff values for θ_i determining which patient types are kept for how many days. $\overline{\theta}_d^{baseline}$ denotes the highest θ_i for which the patient stays d days under the baseline schedule. A patient with a θ_i just above $\overline{\theta}_d^{baseline}$ would stay $d+1$ while a patient with a θ_i just beneath $\overline{\theta}_d^{baseline}$ would stay d days. The cutoff values defining the range of patients who are discharged on day d^* are implicitly defined by equations

$$\begin{aligned}&\lambda_d\lambda_h U'\left(\pi\right)\left[C\left(d^*+1\right) - C\left(d^*\right) - p\left(1-\gamma\right)\right]\\&\quad = h\left(d^*+1, \overline{\theta}_{d^*}^{baseline}\right) - h\left(d^*, \overline{\theta}_{d^*}^{baseline}\right)\end{aligned}$$

$$\lambda_d \lambda_h U'(\pi)\left[C(d^*) - C(d^*-1) - p(1-\gamma)\right]$$
$$= h\left(d^*, \overline{\theta}_{d^*-1}^{baseline}\right) - h\left(d^*-1, \overline{\theta}_{d^*-1}^{baseline}\right)$$

with π denoting hospital profits in equilibrium.

That is, for patient type $\overline{\theta}_{d^*}^{baseline}$ the hospital is just indifferent between the net profit valued with $\lambda_d \lambda_h$ of keeping her d^*+1 instead of d^* days and the net health benefit it would bring to the patient. A patient with θ_i a little bigger than $\overline{\theta}_{d^*}^{baseline}$ would be kept d^*+1 days, since her health benefit of staying another day is higher than for the $\overline{\theta}_{d^*}^{baseline}$ patient. Similarly, for patient type $\overline{\theta}_{d^*-1}^{baseline}$ the hospital is indifferent between the marginal health benefit of keeping her d^* instead of d^*-1 days and the profit impact. Note that the individual patient has no impact on hospital profits.

Now consider the policy experiment of interest, reducing marginal reimbursement by $\triangle p > 0$ throughout the schedule, i.e. $P^{reform}(d_i) = \tilde{p} + (p - \triangle p) \cdot d_i$ with $\triangle p > 0$, while adjusting the fixed payment component such that hospital profits remain at π in equilibrium. That is, $\tilde{p}$ is chosen such that the profits remain at π taking into account the fact that the patients' length of stay will change with this change in the payment schedule. The cutoff values are now defined by

$$\lambda_d \lambda_h U'(\pi)\left[C(d^*+1) - C(d^*) - (p - \triangle p)(1-\gamma)\right]$$
$$= h\left(d^*+1, \overline{\theta}_{d^*}^{reform}\right) - h\left(d^*, \overline{\theta}_{d^*}^{reform}\right)$$

$$\lambda_d \lambda_h U'(\pi)\left[C(d^*) - C(d^*-1) - (p - \triangle p)(1-\gamma)\right]$$
$$== h\left(d^*, \overline{\theta}_{d^*-1}^{reform}\right) - h\left(d^*-1, \overline{\theta}_{d^*-1}^{reform}\right)$$

which implies that the cutoff values increase, i.e. $\overline{\theta}_{d}^{reform} > \overline{\theta}_{d^*}^{baseline}\ \forall d$. That is, the patients stay on average for a shorter time. Due to the discreteness of the assignment variable —length of stay— I need to make an additional regularity assumption relative to the standard bunching setting. Specifically, I assume that patients who share the same length of stay d

under the old schedule $P^{baseline}(d_i)$ move towards at most two different length of stay values under the new schedule $P^{reform}(d_i)$. That is, those patients who stay, e.g., 5 days under the old schedule, stay for $3-4$ days or for $4-5$ days under the new schedule, but never for $2-4$ or $3-5$ days.

Now consider the introduction of a convex kink at d^*. That is, the payment schedule becomes

$$P^{kink}(d_i) = \begin{cases} \bar{p} + p \cdot d_i & d_i \leq d^* \\ \\ \bar{p} + (p - \triangle p) \cdot d_i & d_i > d^* \end{cases}$$

Under the new kinked payment schedule, the new cutoff values defining who is discharged at d^* —$\overline{\theta}_{d^*}^{kink}$ and $\overline{\theta}_{d^*-1}^{kink}$— are defined by

$$\begin{aligned} &\lambda_d \lambda_h U'(\pi)\left[C(d^*+1) - C(d^*) - (p - \triangle p)(1-\gamma)\right] \\ &\qquad = h\left(d^*+1, \overline{\theta}_{d^*}^{kink}\right) - h\left(d^*, \overline{\theta}_{d^*}^{kink}\right) \end{aligned}$$

$$\begin{aligned} &\lambda_d \lambda_h U'(\pi)\left[C(d^*) - C(d^*-1) - p(1-\gamma)\right] \\ &\qquad = h\left(d^*, \overline{\theta}_{d^*-1}^{kink}\right) - h\left(d^*-1, \overline{\theta}_{d^*-1}^{kink}\right) \end{aligned}$$

Hence, $\overline{\theta}_{d^*-1}^{kink} = \overline{\theta}_{d^*-1}^{baseline}$ and $\overline{\theta}_{d^*}^{kink} = \overline{\theta}_{d^*}^{reform} > \overline{\theta}_{d^*}^{baseline}$ and $\overline{\theta}_{d^*}^{kink} - \overline{\theta}_{d^*}^{baseline}$ is the excess mass or bunching at d^* under the kinked schedule. Hence, $\overline{\theta}_{d^*}^{kink}$ is the marginal buncher who responds to the introduction of $P^{kink}(d_i)$ the same way as to the introduction of $P^{reform}(d_i)$.

What Does a Bunching Design Identify?

We established that the marginal buncher responds to the introduction of the kink the same way as to the policy experiment of interest (that is, changing marginal pay by $\triangle p$ throughout the schedule). Let $\tilde{d}$ denote the length of stay that the marginal buncher $\overline{\theta}_{d^*}^{kink}$ would have enjoyed under the baseline linear schedule. For this marginal buncher, the causal effect of interest —the effect of changing marginal reimbursement per day by $\triangle p$ on length of stay— is $\frac{d(d_i)}{d(p)}\triangle p = \tilde{d} - d^*$. Using the assumption discussed above, the causal effect $\frac{d(d_i)}{d(p)}\triangle p$

is equal to $\tilde{d}-d^{*}$ for all patients who would have stayed $\tilde{d}$ under the baseline linear schedule and for whom $\theta_i < \bar{\theta}_{d^*}^{kink}$, but the causal effect $\frac{d(d_i)}{d(p)}\triangle p$ is $\tilde{d}-(d^{*}+1)$ for all patients who would have stayed $\tilde{d}$ under the baseline linear schedule and for whom $\theta_i > \bar{\theta}_{d^*}^{kink}$. Therefore, the total causal effect on patients staying $\tilde{d}$ under the old baseline schedule is

$$\begin{aligned}
E\left[\frac{d\left(d_i\right)}{d\left(p\right)}\triangle p \,\middle|\, \bar{\theta}_{\tilde{d}}^{baseline} > \theta_i > \bar{\theta}_{\tilde{d}-1}^{baseline}\right] &= \left(\tilde{d}-d^{*}\right)\frac{\bar{\theta}_{d^*}^{kink}-\bar{\theta}_{\tilde{d}-1}^{baseline}}{\bar{\theta}_{\tilde{d}}^{baseline}-\bar{\theta}_{\tilde{d}-1}^{baseline}} \\
&\quad + \left(\tilde{d}-(d^{*}+1)\right)\frac{\bar{\theta}_{\tilde{d}}^{baseline}-\bar{\theta}_{d^*}^{kink}}{\bar{\theta}_{\tilde{d}}^{baseline}-\bar{\theta}_{\tilde{d}-1}^{baseline}} \\
&= \tilde{d}-d^{*}-1+\frac{\bar{\theta}_{d^*}^{kink}-\bar{\theta}_{\tilde{d}-1}^{baseline}}{\bar{\theta}_{\tilde{d}}^{baseline}-\bar{\theta}_{\tilde{d}-1}^{baseline}}
\end{aligned}$$

A simple example makes the formula intuitive: If the observed bunching mass is only a small fraction of the observed mass at $d^{*}+1$ —say, 10%— $\tilde{d}=d^{*}+1$, because the marginal buncher is coming from $d^{*}+1$, and $\frac{\bar{\theta}_{d^*}^{kink}-\bar{\theta}_{\tilde{d}-1}^{baseline}}{\bar{\theta}_{\tilde{d}}^{baseline}-\bar{\theta}_{\tilde{d}-1}^{baseline}} \approx 0.1$, since only patients who are at $d^{*}+1$ under the counterfactual linear schedule and whose $\theta_i < \bar{\theta}_{d^*}^{kink}$ bunch at d^{*} together with the marginal buncher. In the example, the formula tells us that the average causal effect on the patients staying $d^{*}+1$ days under the counterfactual linear schedule is 0.1 days, since that is the fraction of patients who move from $d^{*}+1$ to d^{*} due to the kink.

Since d^{*} is known, we need to estimate $\tilde{d}$ and $\frac{\bar{\theta}_{d^*}^{kink}-\bar{\theta}_{\tilde{d}-1}^{baseline}}{\bar{\theta}_{\tilde{d}}^{baseline}-\bar{\theta}_{\tilde{d}-1}^{baseline}}$ in order to get the causal effect of interest. Let B denote the bunching mass estimated from the data and $f\left(d\right)$ the estimated expected mass of patients at d under the contrafactual linear schedule. Then $\tilde{d}$ can be inferred from the data by finding the value for $\tilde{d}$ for which

$$f\left(d^{*}+1\right)+...+f\left(\tilde{d}\right) \geq B$$

$$f\left(d^{*}+1\right)+...+f\left(\tilde{d}-1\right) \leq B,$$

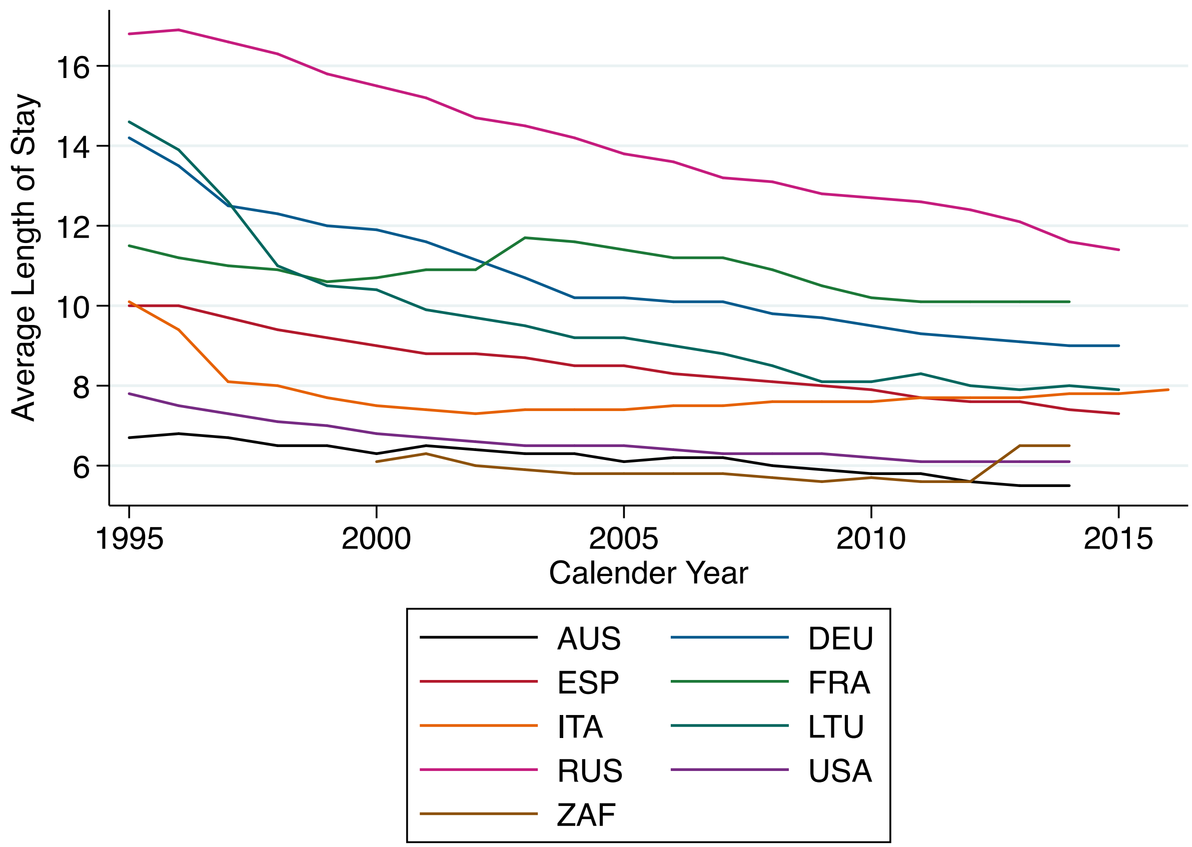
since the bunching mass is equal to the mass at the days from $d^{*}+1$ up to $\tilde{d}-1$ plus the fraction of the mass at $\tilde{d}$ that bunches. This fraction is the bunching mass that is not explained by the mass coming from $d^* + 1$ up to $\tilde{d} - 1$, i.e.

$$\frac{\bar{\theta}_{d^*}^{kink} - \bar{\theta}_{\tilde{d}-1}^{baseline}}{\bar{\theta}_{\tilde{d}}^{baseline} - \bar{\theta}_{\tilde{d}-1}^{baseline}} = B - f\left(d^* + 1\right) + ... + f\left(\tilde{d} - 1\right)$$

Hence, the remaining challenge is to estimate B and $f\left(d^* + 1\right)$, etc. in order to estimate the parameter of interest $E\left[\frac{d(d_i)}{d(p)} \triangle p \middle| \bar{\theta}_{\tilde{d}}^{baseline} > \theta_i > \bar{\theta}_{\tilde{d}-1}^{baseline}\right]$.

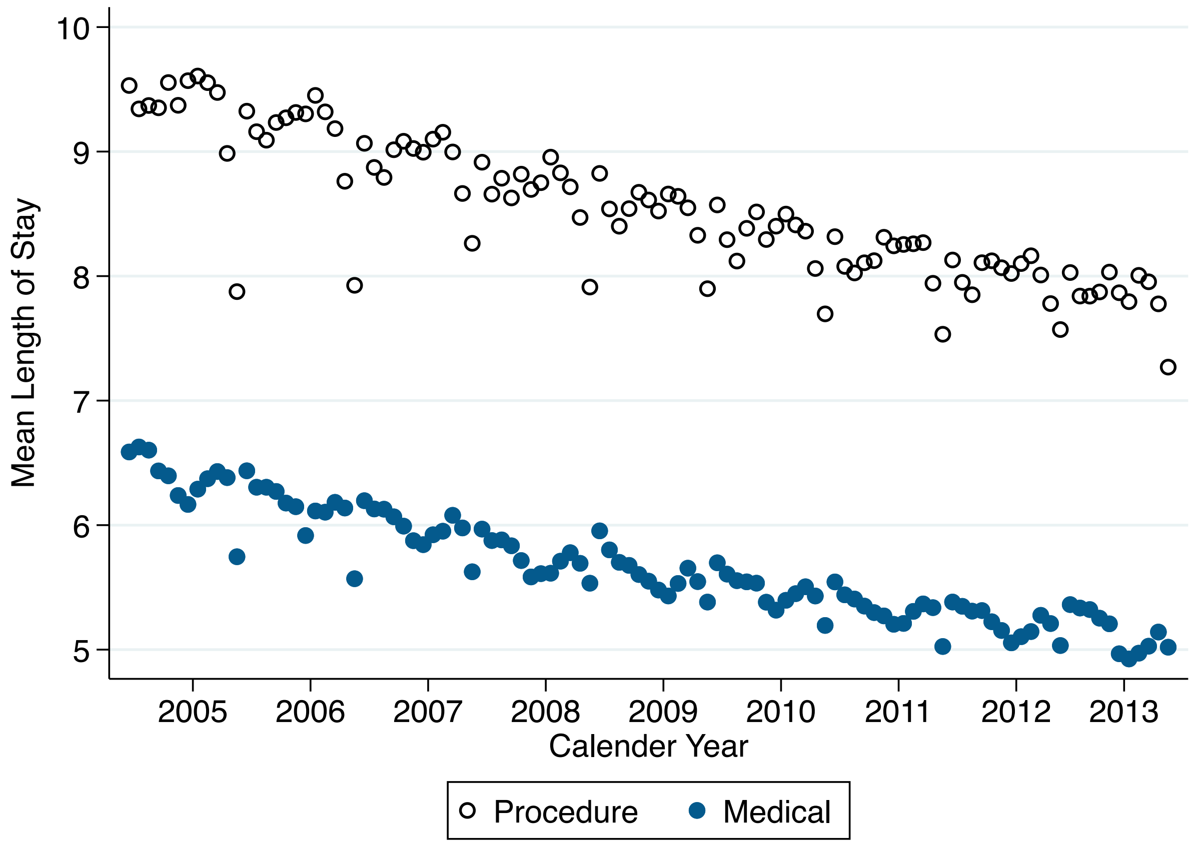
C Further Results

Figure 17: Time Series Length of Stay - Selection of OECD Countries



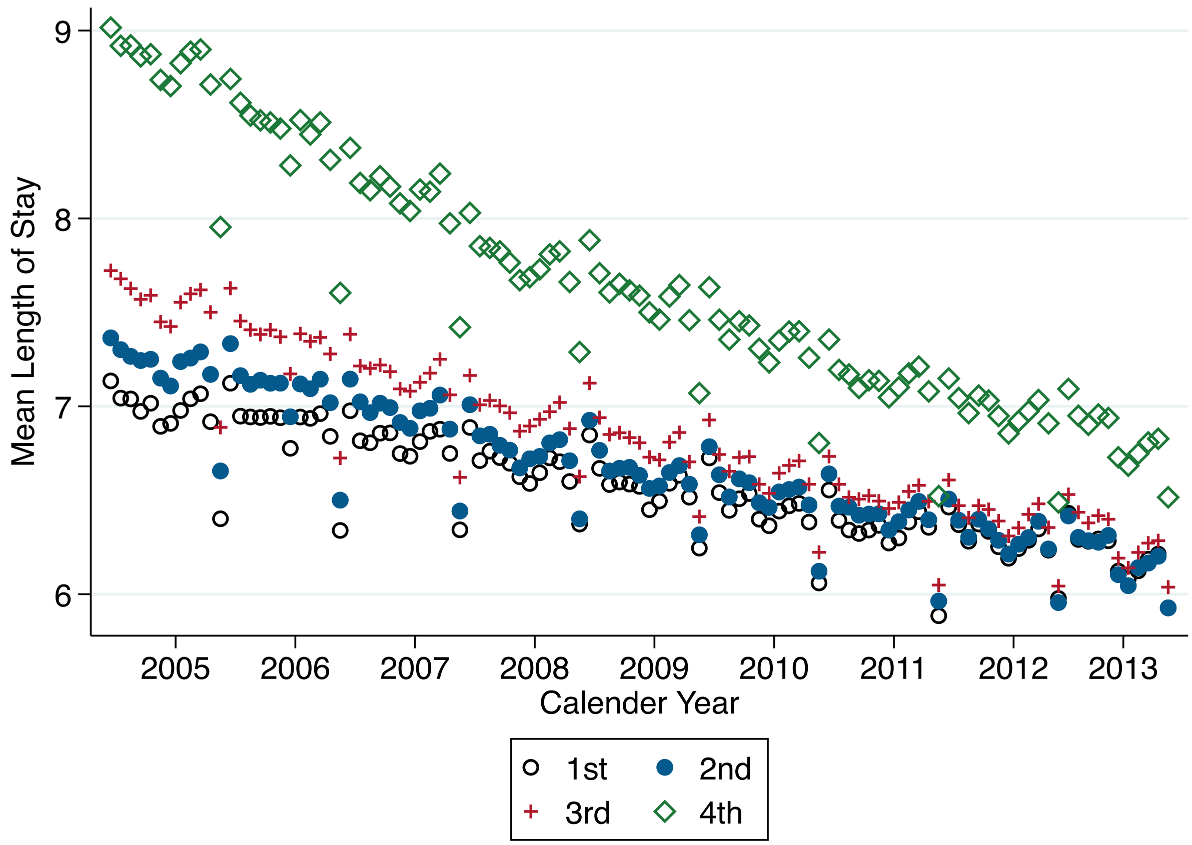
Average length of stay over time for selected OECD countries.
Note that the numbers for Germany do not necessarily agree with the numbers in Figure 5 due to the different source.
Source: OECD

Figure 18: Time Series Length of Stay - Procedural vs Medical DRGs



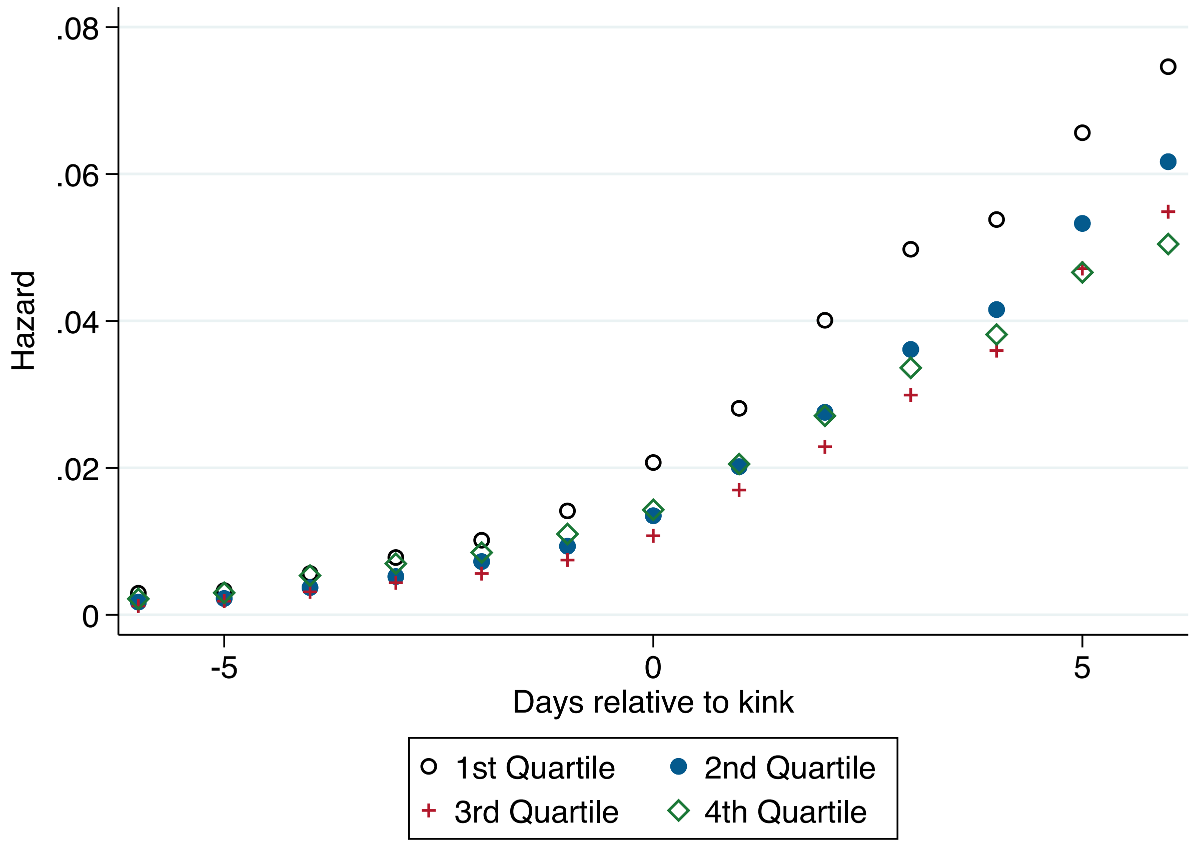
The graph shows the average length of stay for each month 2005-2013 separately for DRGs involving a medical procedures and so-called medical DRGs that do not.

Figure 19: Time Series Length of Stay - by Quartile of Hospital w.r.t. 2005 Average Length of Stay



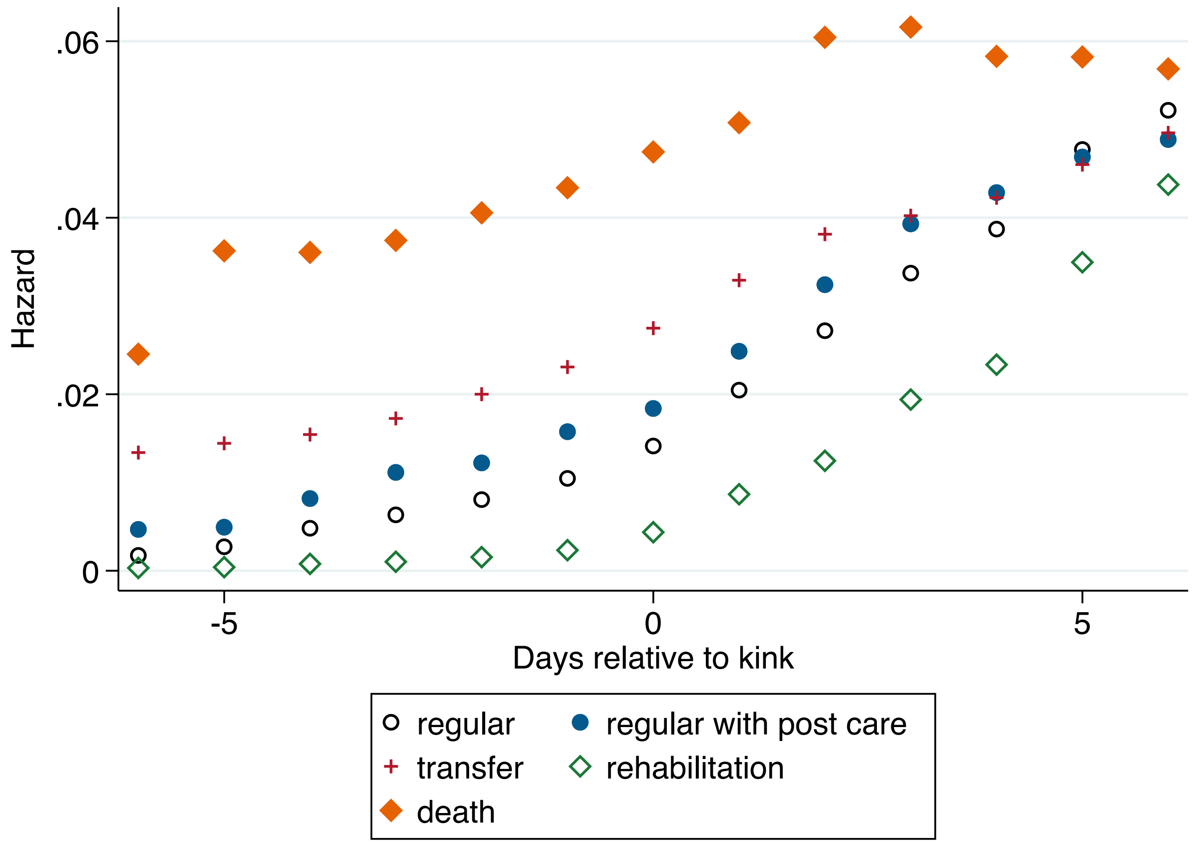
The graph shows the average length of stay (not residualized) for each month 2005-2013 separately by quartile of the hospital in terms of the hospital's residualized length of stay in 2005. I.e. the first quartile are those hospitals that - conditional on the patients' age, gender, birthweight and DRG - have the lowest length of stay in 2005.

Figure 20: Heterogeneity in Hazard Rates Around Kink - by Quartile of Hospital Size



Hazard rates for all patients pooled with a DRG with kink locations of at least 6 days. Done seperately by quartile of hospital size.

Figure 21: Heterogeneity in Hazard Rates Around Kink - by Discharge Reason



Hazard rates for all patients pooled with DRGs with kink locations of at least 6 days. Separately for different discharge reasons. Transfer refers to transfers to other hospitals (transfers are not subject to the kink in the payoff schedule).Rehabilitation refers to transfer to a rehabilitation or longterm care unit or a hospice.

Table 10: Distribution of Kink Locations in the Cross section

Kink Location	DRGs	Patients
n/a	899	7091698
2	3259	85555788
3	1442	18374780
4	1239	11334270
5	803	4806553
6	519	1880018
7	343	780616
8	276	438387
9	183	243005
10+	551	411956
Total	9514	130917071

For each kink location the table gives the number of DRGs that feature this kink location (each year counted separately, so the same DRG is counted several times if it exists in multiple years) as well as the total number of patients who are grouped in one of the respective DRGs. Each patient is only counted once.

D Bunching Estimate for Einav et al. Setting

This appendix describes how I construct the bunching estimate of a 0.34 day reduction in length of stay when cutting marginal pay for another day in the hospital by $1,000$€ for the setting studied in Einav et al. (2017), Eliason et al. (2016) and Kim et al. (2015). It builds notationwise on the derivation in the main text for my main analysis.

I use the distribution of downstream patient discharges across days reported in Figure 8 of Einav et al. (2017) for the time before the jump in payment at the threshold was introduced and for after it was introduced.[26] I assume that patients only shifted within the ± 15 day window around the threshold, that is, I restrict the data to this window.

At the threshold there is a notch (an upward jump in the payment) as well as a kink (the slope of the schedule turns from being increasing towards being flat). To estimate the amount of bunching I run the following type of specification

$$\begin{aligned} share_{d,t} &= \alpha_d + \beta_0 I\{post_t\} + \beta_1 d \cdot I\{post_t\} + \beta_2 I\{below\, threshold_d\} \cdot I\{post_t\} \\ &\quad + \beta_3 I\{above\, threshold_d\} \cdot I\{post_t\} + u_{d,t} \end{aligned}$$

with $share_{d,t}$ denoting the share of patients in time period t that are discharged after d days, α_d indicators for each day, $I\{post_t\}$ an indicator for whether t is for the time after the threshold was introduced, $I\{below\, threshold_d\}$ an indicator for whether d is for the days just below the threshold (how many days is determined within the process of looking for a solution, because it depends on the bunching mass) and $I\{above\, threshold_d\}$ an indicator for whether d is the day just above the threshold.

That is, the α_d imply that the coefficients of interest are only identified from the changes from the pre- to the post-period and not from smoothness assumptions. $I\{post_t\} + d \cdot I\{post_t\}$ allows the mass to shift in the post-period relative to the pre-period, i.e. allows for time trends. The $I\{above\, threshold_d\} \cdot I\{post_t\}$ captures the bunching mass and the $I\{below\, threshold_d\} \cdot I\{post_t\}$ allows for the hole in the mass of discharges that is expected for a notch with an upward jump.

Due to the presence of the notch as well as the kink, there is a marginal buncher caused by the notch and coming from below the threshold day d^* —denote the contrafactual day of that marginal buncher by d^{notch} and her type θ^{notch}— as well as a marginal buncher caused by the kink and coming from aboive the threshold day d^* —denote the contrafactual day of that marginal buncher by d^{kink} and her type θ^{notch}.

I approximate the notch induced change in marginal pay by $\frac{\triangle p}{d^* - d^{notch}}$, with $\triangle p$ (about \$13,000 on average) denoting the payment jump at the threshold. The kink induced change in marginal pay is simply the slope $\check{p}$ that the schedule has to the left of the threshold (about \$1,400 on average).

[26] The January 2017 version of their paper.

By definition, the marginal bunchers react to the notch and kink the same way as they would to a change in marginal pay throughout the schedule, i.e. θ^{notch} would also move from d^{notch} to d^* if marginal pay were increased by $\frac{\triangle p}{d^*-d^{notch}}$, throughout the schedule and θ^{kink} would also move from d^{kink} to d^* if marginal were decreased by $\check{p}$ througout the schedule. Hence, the causal effect of interest for θ^{notch} is $\frac{d(d_i)}{dp}\triangle p = d^* - d^{notch}$ and for θ^{kink} is $\frac{d(d_i)}{dp}\check{p} = d^{kink} - d^*$.

I make the same regularity assumption as in my main text: Patients who share the same length of stay under one payment schedule are at at most two different length of stay values under a different payment schedule - e.g. all patients who stay 6 days under one payment schedule stay 3 or 4 days or they stay 4 or 5 days, but never 3 to 5 days.

Under the regularity assumption, the causal effect for all patients with $\theta_i > \theta^{notch}$ and who in the absence of the notch would also stay d^{notch} is $\frac{d(d_i)}{dp}\triangle p = d^* - d^{notch}$, while those with $\theta_i < \theta^{notch}$ and who in the absence of the notch would also stay d^{notch}, have a causal effect $\frac{d(d_i)}{dp}\triangle p = (d^*-1) - d^{notch}$. Similarly, all patients with $\theta_i < \theta^{kink}$ and who in the absence of the kink would also stay d^{kink} have the causal effect $\frac{d(d_i)}{dp}\check{p} = d^{kink} - d^*$, while those with $\theta_i > \theta^{notch}$ and who in the absence of the notch would also stay d^{kink}, have the causal effect $\frac{d(d_i)}{dp}\check{p} = d^{kink} - (d^*+1)$.

Let $\theta_d^{contrafactual}$ denote the patient marginal between $d+1$ and d under the contrafactual distribution in the post period if no jump and kink had been introduced. The average causal effect for those at d^{notch} is $\frac{d(d_i)}{dp}\triangle p = \left(d^* - d^{notch}\right)\frac{\theta^{contrafactual}_{d^{notch}} - \theta^{notch}}{\theta^{contrafactual}_{d^{notch}} - \theta^{contrafactual}_{d^{notch}-1}} + \left((d^*-1) - d^{notch}\right)\frac{\theta^{notch} - \theta^{contrafactual}_{d^{notch}-1}}{\theta^{contrafactual}_{d^{notch}} - \theta^{contrafactual}_{d^{notch}-1}}$. The average causal effect for those at d^{kink} is $\frac{d(d_i)}{dp}\check{p} = \left(d^{kink} - d^*\right)\frac{\theta^{kink} - \theta^{contrafactual}_{d^{kink}-1}}{\theta^{contrafactual}_{d^{kink}} - \theta^{contrafactual}_{d^{kink}-1}} + \left(\left(d^{kink}-1\right) - d^*\right)\frac{\theta^{contrafactual}_{d^{kink}} - \theta^{kink}}{\theta^{contrafactual}_{d^{kink}} - \theta^{contrafactual}_{d^{kink}-1}}$.

I loop through all possible values for $\frac{d(d_i)}{dp}\triangle p$ and for $\frac{d(d_i)}{dp}\check{p}$ which imply the same $\frac{d(d_i)}{dp}$. For each of those values, I run the specification mentioned above

$$\begin{aligned} share_{d,t} &= \alpha_d + \beta_0 I\{post_t\} + \beta_1 d \cdot I\{post_t\} + \beta_2 I\{below\,threshold_d\} \cdot I\{post_t\} \\ &\quad + \beta_3 I\{above\,threshold_d\} \cdot I\{post_t\} + u_{d,t} \end{aligned}$$

. The $I\{below\,threshold_d\}$ is defined as d being in the $d^* - d^{notch}$ first days to the left of the threshold (note that d^{notch} is implied from the guessed value for $\frac{d(d_i)}{dp}\triangle p$).

I then use the regression result to get the amount of bunching from β_3 and I calculate the contrafactual mass distribution if the threshold had not been introduced from the predicted values not using the β_2 and β_3.

I then calculate the bunching mass implied from the guessed $\frac{d(d_i)}{dp}\triangle p$ and $\frac{d(d_i)}{dp}\check{p}$ together with the contrafactual distribution (if, e.g., $\frac{d(d_i)}{dp}\triangle p = 2.3$, I add the contrafactual masses at $d^* - 1$, $d^* - 2$ and 0.3 times the mass at $d^* - 3$. That way I calculate the total amount of bunching mass that is implied and check whether it is equal to the estimated amount of bunching.

My estimates for $\frac{d(d_i)}{dp}\triangle p$ and $\frac{d(d_i)}{dp}\check{p}$ are those for which the amount of bunching estimated corresponds to the one implied by $\frac{d(d_i)}{dp}\triangle p$ and $\frac{d(d_i)}{dp}\check{p}$.

$\frac{d(d_i)}{dp}\triangle p$ and $\frac{d(d_i)}{dp}\check{p}$ imply a value for $\frac{d(d_i)}{dp}$ which I then convert to an estimate with respect to a change in $1,000$ 2013-Euro using the American CPI and the 2013 Euro-Dollar exchange rate.

The result is a 0.34 day reduction in length of stay when cutting the marginal pay per day by $1,000$ 2013-Euro. The confidence interval (bootstrapped with 250 repetitions) is from 0.31 to 0.37 days.

E Back-of-the-Envelope Calculation for the Comparison to Medicare

Here I describe how I calculate the average reimbursement per hospitalization day in Medicare in 1984.

Levit et al. (1985) reports 44.24 billion dollars in 1984 in Medicare hospital spending. Kominski and Witsberger (1993) and Guterman and Dobson (1986) provide average length of stay and the number of hospitalizations for 1984 Medicare patients, resulting in a total of 103.46 million Medicare hospital days in 1984. Dividing total costs by the number of days and using the U.S. CPI and the 2013 average exchange rate from Dollar to Euro, I end up with approximately 750 2013-Euro per day.