Can Provider Incentives Reduce Health Care Costs?  
New Evidence From Germany

Jakob Schlockermann

September 25, 2017

Abstract

I study the design of provider incentives in acute hospital care. Specifically, I estimate the causal effect of moving hospital reimbursement from being increasing towards being flat in the patient’s length of stay. For German hospitals, pay is increasing in length of stay up to a certain number of days at which point it kinks and becomes flat. I investigate the degree to which hospitals bunch hospital discharges on the kink day. I show theoretically that my bunching design identifies the causal effect on length of stay of moving from a hypothetical linearly increasing to a flat payment schedule. The estimates are precise and one can reject a more than 0.04 days reduction in average length of stay from making the payment schedule flat. Reducing length of stay was a central political goal behind many countries’ — including Germany’s in 2004 — move towards hospital payment schedules that give less financial incentives to increase length of stay. My results imply that the German politicians failed to reach their goal. This paper’s findings contrast with the literature’s conventional wisdom that deems hospital payment incentives in the US effective in affecting length of stay. This contrast is particularly noteworthy, because the relevant German institutional features give no reason to expect a muted effect relative to the US.
1 Introduction

In the face of rising health care costs, countries around the world have reformed the way they reimburse health care providers and, in particular, hospitals. Prominent examples of traditional payment schedules for hospitals were fee for service systems, that pay health care providers for each individual service provided, and per diem systems, that pay a fixed amount per hospitalization day. Both systems give hospitals a financial incentive to increase their patients’ length of stay. The fact that traditional reimbursement systems might lead to an inefficient overprovision of care was not only recognized by health economists (see e.g. Ellis and McGuire 1986), but also by politicians. In 1983, Medicare famously reformed the way it reimbursed hospitals for inpatient care from a fee for service system to a so-called prospective payment system.

Prospective payment systems make the hospital’s pay less dependent on the actually incurred costs and more strongly tied to the expected costs based on case-characteristics. Specifically, Medicare patients since 1983 are grouped into Diagnostic Related Groups (DRGs) based on their diagnoses, major procedures (e.g. bypass surgery) and patient demographics such as age. Within one of those roughly thousand DRGs, the hospital’s pay is independent of the number of services provided and, in particular, the patient’s length of stay.¹ DRG systems for hospital reimbursement have since been introduced by many other industrialized countries, such as France, England, Spain, the Netherlands and, importantly for this paper, Germany (Geissler et al. 2011).

This paper provides rare quasi-experimental evidence on the causal impact of hospital financial incentives on length of stay by making use of a unique feature in the German reimbursement schedule for hospitals. As in Medicare, German patients are grouped into DRGs based on diagnoses, major procedures and patient demographics. Within a DRG, however, the hospital’s pay is not entirely independent of length of stay. Instead, the pay increases linearly in length of stay up to a certain number of days at which it kinks and becomes flat. Hence, the hospital’s marginal pay for keeping a patient in the hospital for another day drops discontinuously at the kink. Thus, discharging a patient becomes discontinuously

¹Medicare pays the hospitals some extra compensation in case the patient is an extreme outlier with respect to her length of stay.
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 that in a general class of models the amount of bunching informs about a parameter of interest to policy. Specifically, the bunching design identifies the change in average length of stay for the policy experiment of changing hospital reimbursement from a hypothetical payment schedule that pays linearly per day with the same slope that the actually observed kinked schedule features to the left of the kink to a schedule that is entirely flat in 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 prospective payment system on length of stay. First, there is no notable break in average length of stay around its introduction suggesting little causal impact. Second, eyeballing 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, due to the discrete nature of the assignment variable ‘days in the hospital’. 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 the policy experiment of changing hospital reimbursement towards a schedule that is entirely flat in length of stay when coming from a hypothetical payment schedule that pays
linearly per day with the same slope that the actually observed kinked schedule features to the left of the kink. Specifically, I can reject that length of stay would fall by more than 0.04 days.

Because of the way they are determined, the kink locations can only change over time for DRGs with a relatively high average length of stay. To show that my results carry over to less serious diseases, I analyze discharge behavior for transfers. For patients who are transferred from or to the hospital, there is a separate payoff schedule that allows me to apply the dynamic bunching design also to DRGs with smaller average durations. The estimates for the transferred cases are very similar to those from the main analysis suggesting that the results are not specific to severe diseases.

What do this paper’s estimates imply for the historically observed payment scheme reforms? Germany introduced its prospective payment system based on DRGs in 2004 coming from a per diem system, i.e. a system that pays a fixed amount per hospitalization day.\textsuperscript{2} German politicians 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:\textsuperscript{3}

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.

The estimated upper bound for the hypothetical policy experiment does not directly apply to the 2004 German reform, but is still informative. As I argue in detail in section 6, the change in financial incentives in the 2004 reform was smaller than the change in incentives for which the bunching design identifies the causal effect. The 0.04 day reduction in average length of stay therefore is also an upper bound for the impact of the 2004 German reform. This

\textsuperscript{2}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.

\textsuperscript{3}Dtsch Arztebl 2009; 106(26) - translation by this paper’s author
suggests that the politicians claiming success confused a secular declining trend in length of stay with the causal effect of their reform. For policy, my results imply that tying hospital pay more closely to length of stay again could improve welfare, because it would make the pay more closely connected to the hospitals’ actually incurred costs, thereby reducing the hospitals’ financial risk of drawing particularly sick patients.

As in Germany, a primary stated motivation for introducing prospective payment systems in Medicare was to give providers an incentive to contain costs and, in particular, length of stay. In the “Report to Congress: Hospital Prospective Payment for Medicare” from 1982 it says on page 102

> For the first time since the inception of the Medicare program there will be true incentives to match explicitly patient benefits with the costs of services provided to Medicare beneficiaries. Hospital managers will attempt to make their institutions produce patient care in all case types as efficiently as possible. Hospitals will seek to reduce both unnecessary lengths of stay and the quantity of unneeded routine and ancillary services, consistent with other institutional and social goals.

Indeed, the conventional wisdom in the academic literature deems the 1983 introduction of Medicare’s prospective payment system as having been successful in reducing length of stay. For instance, Coulam and Gaumer (1992) conclude on the effects on length of stay

> Even with this complication over the timing and extent of the change, there is little dispute that, for short-stay hospitals subject to the prospective payment system, length of stay declined, then stabilized.

My estimate of a 0.04 days reduction in length of stay when introducing a flat payment schedule is an order of magnitude smaller than what researchers have attributed to the 1983 Medicare reform. While comparing the effect sizes between the German and the Medicare reform must to a degree be speculative, I will show in section 6 that it remains interesting, because there is no institutional reason why one should have expected the effects in Germany to be so much smaller.
My results also paint a different picture than the literature studying the effects of provider incentives more generally. The conventional wisdom – mostly evaluating the effects of financial incentives for doctors in the US – deems provider incentives to be quite influential on treatment decisions (e.g. Rice 1983, Nguyen and Derrick 1997 Yip 1998, Gruber, Kim, and Mayzlin 1999, Jacobsen et al 2010, Clemens and Gottlieb 2014, Coey 2015, Alexander 2015).

The studies closest to my design are the recent papers by Kim et al. (2015), Einav et al. (2016) and Eliason et al. (2016) who exploit a discontinuity in the Medicare payment schedule for post-acute care. In contrast to my paper, they find strong bunching right at the discontinuity and, hence, sizable effects of marginal pay on length of stay.

Outline

Section 2 gives institutional background. In Section 3, I demonstrate theoretically which exact policy parameter the bunching design identifies. Section 4 discusses the data and the sample selection. Section 5 presents the results. Section 6 provides a discussion and concludes.

2 Institutional Background

Germany’s health care system is one of the most expensive among OECD countries. In 2015, Germany spent 11.1% of its GDP on health care (OECD average 9.0%) putting it on fourth position in the OECD. With 8.3 hospital beds per 1,000 people in 2013 (OECD average 4.8 beds) the hospital sector is very expensive in international comparison as well, reflecting a high average length of stay (7.7 days in 2013, putting it third among OECD countries only behind Japan and Korea) as well as a large total number of hospitalizations (25,602 per 100,000 people in 2014, putting it second only behind Austria).

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.\textsuperscript{5}

Until 2004\textsuperscript{6}, Germany reimbursed hospital using a cost-based per diem system in the

\textsuperscript{4}Grant (2009) replicates the study and gets quantitatively smaller but qualitatively similar results
\textsuperscript{5}This is slightly simplified - see details in the “supply side institutions” subsection.
\textsuperscript{6}Technically, the system already switched in 2003, but it only became compulsory in 2004
majority of cases. That is, the fee payable to the hospital increased linearly (with a hospital-specific slope depending on its 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 Diagnostic Related Groups (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 1-day-DRGs which determine reimbursement in the special case of a patient having a certain diagnosis and staying just one day. These 1-day-DRGs do not pose a 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. That is, DRGs for which the patient composition changes mechanically because e.g. a new 1-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 is 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) is a function of the hospital stay length as depicted in Figure 1. The parameters of the payment scheme are also based on the hospital cost data from two years

---

7The 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.
before. The payment increases linearly until a third of the average length of stay (rounded and measured two years prior) of 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) by the number of days at which the kink occurs. After the kink, the payment schedule remains flat until the average plus two times the standard deviation of the length of stay within the DRG in question is reached. 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 of days in the hospital). Interestingly, while the overall goal of the DRG reform was to reduce length of stay, the lower kink point was introduced in order to discourage hospitals from discharging patients extremely early. Given that this paper finds an absence of bunching, this policy appears to have failed.

Figure 2 demonstrates how strongly marginal pay changes at the kink. 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. That is, I calculate the percentage loss in hospital revenue 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 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.

There is another way of thinking about how big the slope changes at the kink are: In the theory section, I will demonstrate that the bunching design identifies the causal effect (on length of stay) of moving from a system that pays linearly per day with the same slope that the observed kinked payment schedule features to the left of the kink towards prospective pay (that is, towards a system that pays an amount independent of length of stay). If, however, we actually paid hospitals linearly per day with the same slope that the observed kinked payment schedule features to the left of the kink, then - holding length of stay for all patients constant - total hospital revenue would be (even if we did not pay any fixed

---

8To 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)

9This 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)
amount per hospitalization) more than twice of what it is given the current kinked payment schedule. Hence, any system which pays per day and which is realistic (i.e. that pays per day with a slope such that total hospital revenue remains about the same as with the current system) would feature a much less steep slope than the hypothetical fee for service system we estimate the causal effect for. In this sense, the observed kinks cause such a strong change in slope that the design overestimates the causal effect of moving towards prospective pay when coming from a realistic system that pays hospitals linearly per day. Therefore, this paper provides a very tight upper bound on the causal effect that is of interest for policy.

Figure 8 in the appendix shows a specific example of a DRG pay schedule. It is coming directly from the data (using the variable for revenue), showing the average reimbursement for the DRG K01C\textsuperscript{10} in 2010 in Bavaria\textsuperscript{11} plotted against the number of days in the hospital. As it is apparent from the graph, this DRG has its kink point at seven days.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{within-drg-payment-schedule}
\caption{Within-DRG payment schedule as a function of the length of the hospital stay}
\end{figure}

\textsuperscript{10}In case the reader is curious: the definition of this DRG is "Various procedures with diabetes with complications, without early rehabilitation, without complex geriatric early-rehabilitative treatment, without vascular intervention, with severe CC (complication or comorbidity) or complex arthrodesis of the feet"

\textsuperscript{11}I limit it to Bavaria, because - as discussed later - the payment schedule in 2010 shifts proportionally in the vertical direction depending on the state
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 lower kink day.

The graph shows a histogram of this measure across all DRGs and all years 2005-2013.

Transferred patients are subject to a different payment schedule. They are analyzed later in a separate section.

**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 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.

The health insurers can 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 workload 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.\textsuperscript{12} As discussed previously and demonstrated in the generalized model in the appendix, the presence of audits is not confounding the research design even with anticipatory behavior, since the causal effect of interest is the effect given the presence of demand side constraints.

\textbf{Supply Side Institutions}

In 2013, there were 1,995 hospitals in Germany, with 596 being public, 706 non-profit and 693 for-profit. 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 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.

At the beginning of each year, hospitals and insurers agree on a 'hospital budget' according to the expected amount of hospital revenue. Deviations of actual revenue from this hospital budget are only partially compensated. Hence, any change in hospital DRG revenue does translate proportionally into actual revenue but not one-to-one. This, however, does not affect the kinked nature of the schedule.

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

\footnote{12 see e.g. Faktenblatt Thema: Abrechnungsprüfung in Krankenhäusern from 06/06/2014}
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 pay 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, this is coded in the data and a very rare event. Discharges typically take place midday after the doctor’s ward round.

Medical liability risk is generally perceived to be small relative to that in the US due to comparatively small awards against physicians. 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

One concern for identification is that a patient’s DRG is not fix throughout her hospital stay, but can change if, for instance, the patient gets an infection and therefore a new major diagnosis. In qualitative interviews I conducted, doctors working in German hospitals confirmed that in typical cases the patient’s DRG is very predictable from day one. Moreover, 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 whom they work together with closely.

---

13 see e.g. Law Library Of Congress - Medical Malpractice Liability Systems In Selected Countries
14 American Medical News, May 3, 2010
Note that the presence of demand side constraints (such as audits) as well as supply side constraints (such as non-aligned incentives between hospitals and doctors) do not violate the identification assumptions. The causal effect of interest is the change in average length of stay when moving from fee for service to prospective pay given the structure of the demand and supply side institutions. That is, we want to know how hospitals’ treatment decisions change in response to marginal pay under the actual institutional constraints they face. I discuss this point in more detail in the extended theory section in the appendix.

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.

3 Theory

I present a simple model of hospital behavior to demonstrate that the causal response in average length of stay when switching from a fee for service system that pays the hospital per day to a prospective payment system is identified (for a well-defined subset of patients) by the ratio of bunching mass to the mass discharged on the day above the kink point. To be more precise, the design identifies the effect of moving towards prospective pay coming from a system that pays linearly per day with the same slope that the observed kinked payment schedule features to the left of the kink. As discussed above in the subsection about the payment schedule, a system that pays linearly per day in such a way would feature a much steeper slope than any realistic fee for service system, implying that the design identifies an upper bound for the causal effect of switching from a realistic fee for service system to prospective pay.

I present a simplified model here - in the appendix I show that the result holds in a more general model with risk-averse hospitals (the identified effect then applies to a payment

\footnote{This result holds as long as the bunching mass is smaller than the mass at the day above the kink - which is the empirically relevant case.}
scheme reform that keeps hospital profits constant in equilibrium), a generic formulation of
the cost function as well as the possibility of audits by the health insurers.

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

The hospital receives payment \( P(d_i) \) and incurs costs \( C(d_i) = \bar{c} + c \cdot d_i \). Because of asymmetric
information the hospital can choose any \( d_i \), but faces inducement costs proportional
to the forgone health benefit. The inducement costs are a reduced form for all costs propor-
tional to the health effect of changing \( d_i \) - for instance, it captures any intrinsic concern
the hospital might have about patient health, principal-agent costs that might arise from
having to bring the doctors to treat the patients in a non-health-optimizing way as well as
any liability or reputation costs.

The hospital trades of profits with inducement costs according to preference parameter
\( \lambda \) and solves
\[
\max_{\{d_i(\theta_i) \in \mathbb{R}\}} \int \left[ P(d_i) - (\bar{c} + c \cdot d_i) \right] + \lambda \int h (d_i, \theta_i)
\]

**Kinked Payment Schedule**

I start by characterizing hospital behavior under the actually observed kinked payment
schedule. It features slope \( p \) up until kink point \( d^* \) and is flat afterwards. Moreover, it pays
a lump-sum payment \( \bar{p}^{\text{kink}} \). That is, \( P^{\text{kink}}(d_i) = \bar{p}^{\text{kink}} + p \cdot d_i \) if \( d_i < d^* \) and \( P^{\text{kink}}(d_i) = \bar{p}^{\text{kink}} + p \cdot d^* \) if \( d_i \geq d^* \).\(^{16}\)

Optimal hospital behavior amounts to choosing cutoff values for \( \theta_i \) determining which

\(^{16}\)The actually observed payment schedule also features the upward sloping part for very long staying
patients. I assume that \( h (d_i, \theta_i) \) is sufficiently concave in \( d_i \) to nevertheless ensure a globally concave
objective function which makes the local optimality conditions for the hospital behavior sufficient. This
assumption is unlikely to be problematic even if not strictly true. What is key for my analysis is that the
local optimality conditions are sufficient around the lower kink point. For this to fail because of non-convexity
it would have to be that the hospital is indifferent between two very different length of stay options (one
around the lower kink point and one to the right of the upper kink point) for the same patient just because
of the convexity in the payoff schedule - this appears very implausible.
patient types are kept for how many days. $\theta_{d}^{kink}$ denotes the highest $\theta_i$ for which the patient stays $d$ days. A patient with a $\theta_i$ just above $\theta_{d}^{kink}$ would stay $d+1$ while a patient with a $\theta_i$ just beneath $\theta_{d}^{kink}$ would stay $d$ days. The cutoff values of interest are going to be those defining the range of patients who are discharged on kink day $d^*$. These are implicitly defined by equations

$$c = \lambda \left[ h \left( d^* + 1, \theta_{d^*}^{kink} \right) - h \left( d^*, \theta_{d^*}^{kink} \right) \right]$$

$$c - p = \lambda \left[ h \left( d^*, \theta_{d^*-1}^{kink} \right) - h \left( d^* - 1, \theta_{d^*-1}^{kink} \right) \right]$$

That is, for patient type $\theta_{d^*}^{kink}$ the hospital is just indifferent between the net profit of keeping her $d^* + 1$ instead of $d^*$ days (which is $-c$) and the net health benefit it would bring to the patient (which is $h \left( d^* + 1, \theta_{d^*}^{kink} \right) - h \left( d^*, \theta_{d^*}^{kink} \right)$) multiplied with how much the hospital values the health benefit, that is parameter $\lambda$. A patient with $\theta_i$ a little bigger than $\theta_{d^*}^{kink}$ would be kept $d^* + 1$ days, since her health benefit of staying another day is higher than for the $\theta_{d^*}^{kink}$ patient.

Similarly, for patient type $\theta_{d^*-1}^{kink}$ the hospital is indifferent between the marginal health benefit of keeping her $d^*$ instead of $d^*-1$ days and the profit impact (which now becomes $p - c$ because the kink is not reached yet). The resulting patient mass at $d^*$ under the kinked payment schedule is given by $M^{kink}(d^*) \equiv \theta_{d^*}^{kink} - \theta_{d^*-1}^{kink}$.

Next, I analyze hospital behavior under a counterfactual fee for service as well as a counterfactual prospective payment scheme to show how the bunching mass in the observed schedule identifies the causal effect of moving from fee for service to prospective pay.

**Fee for Service**

Let $P^{ffs}(d_i) = \bar{p}^{ffs} + p \cdot d_i$. The difference to the kinked payment schedule is the fact that the variable fee is paid for all $d_i$ and not just up to $d^*$. Without income effects, the level of the lump-sum payment $\bar{p}^{ffs}$ does not matter for hospital behavior - in the general model in the appendix I analyze the general case with risk-averse hospitals and a payment scheme reform that explicitly adjusts $\bar{p}^{ffs}$ to keep equilibrium hospital profits constant. As discussed before, $\bar{p}^{ffs}$ would have to be negative in order to keep hospitals profits constant (in the empirically relevant case in which length of stay does not respond dramatically to
financial incentives), because $p$ is so high that simply paying out $P_{ffs}(d_i) = p \cdot d_i$ would lead to a much higher total hospital revenue than under $P_{kink}(d_i)$.

Optimal hospital behavior is again characterized by cutoff values for $\theta_i$. The cutoff values for $d^*$ given the fee for service payment are now defined by equations

$$c - p = \lambda \left[ h \left( d^* + 1, \overline{\theta}_{d^*}^{ffs} \right) - h \left( d^*, \overline{\theta}_{d^*}^{ffs} \right) \right]$$

$$c - p = \lambda \left[ h \left( d^*, \overline{\theta}_{d^*}^{ffs} \right) - h \left( d^* - 1, \overline{\theta}_{d^*}^{ffs} \right) \right]$$

The only change relative to the kinked schedule is the fact that $p$ shows up in the equation for $\theta_{d^*}^{ffs}$. So we have $\overline{\theta}_{d^*}^{ffs} = \overline{\theta}_{d^*}^{kink}$ and $\overline{\theta}_{d^*}^{ffs} < \overline{\theta}_{d^*}^{kink}$. Hence, under the kinked payment schedule the mass of patients at $d^*$ is $M_{kink}(d^*) = \overline{\theta}_{d^*}^{kink} - \overline{\theta}_{d^*}^{kink} > \overline{\theta}_{d^*}^{ffs} - \overline{\theta}_{d^*}^{ffs} = M_{ffs}(d^*)$ which is the mass of patients at $d^*$ under fee for service. $B \equiv M_{kink}(d^*) - M_{ffs}(d^*) = \overline{\theta}_{d^*}^{kink} - \overline{\theta}_{d^*}^{ffs}$ is the bunching or excess mass at $d^*$ that is going to identify the causal effect of interest.

**Prospective Payment**

Under prospective pay the payment schedule has the form $P_{prospective}(d_i) = \bar{p}$. Again, I leave $\bar{p}$ undetermined here because it is irrelevant with risk-neutral hospitals, but I explicitly discuss it in the general model in the appendix. The relevant cutoff values at $d^*$ become

$$c = \lambda \left[ h \left( d^* + 1, \overline{\theta}_{d^*}^{prospective} \right) - h \left( d^*, \overline{\theta}_{d^*}^{prospective} \right) \right]$$

$$c = \lambda \left[ h \left( d^*, \overline{\theta}_{d^* - 1}^{prospective} \right) - h \left( d^* - 1, \overline{\theta}_{d^* - 1}^{prospective} \right) \right]$$

Relative to the fee for service scheme, both cutoff values go up under the prospective payment schedule. That is, $\overline{\theta}_{d^*}^{prospective} > \overline{\theta}_{d^*}^{ffs}$ and $\overline{\theta}_{d^* - 1}^{prospective} > \overline{\theta}_{d^* - 1}^{ffs}$. Hence, patients stay shorter under prospective pay than under fee for service which is exactly the key motivation for prospective pay the empirical validity of which this paper intends to test.

**What does the bunching design identify?**

The causal effect of interest is the effect of switching from a fee for service system (that pays linearly per day with the same slope that the observed kinked schedule features to the left of the kink) towards a prospective system that pays a flat amount. If this causal effect
is at most a reduction of one day (which given the observed amount of bunching is the only empirically relevant case\(^\text{17}\)), then the causal effect is equivalent to the share of patients that gets released a day earlier after introducing prospective pay. Hence, for the well-defined group of patients staying \(d^* + 1\) days under fee for service the average causal effect \(\Delta d_i\) of switching to prospective pay is

\[
\delta_{\text{switch}} = E \left[ \Delta d_i \bigg| \theta_i \in \left[ \bar{\theta}^{ffs}_{d^*}, \bar{\theta}^{ffs}_{d^* + 1} \right] \right] = \frac{\bar{\theta}^{\text{prospective}}_{d^*} - \bar{\theta}^{ffs}_{d^*}}{\bar{\theta}^{ffs}_{d^* + 1} - \bar{\theta}^{ffs}_{d^*}}
\]

That is, the average effect on length of stay for patients in \(M^{ffs}(d^* + 1)\) is the share of those patients that after switching to a prospective payment system falls beneath the new upper cutoff value for \(d^*\).

Since we do not have a sharp research design to directly compare the hazard rates under fee for service and prospective pay for similar patient populations, we need to use the bunching under the kinked schedule to identify \(\delta_{\text{switch}}\). Note that \(B = \bar{\theta}^{kink}_{d^*} - \bar{\theta}^{ffs}_{d^*} = \bar{\theta}^{\text{prospective}}_{d^*} - \bar{\theta}^{ffs}_{d^*}\) and \(M^{ffs}(d^* + 1) = \bar{\theta}^{ffs}_{d^* + 1} - \bar{\theta}^{ffs}_{d^*}\) which is \(\approx \bar{\theta}^{kink}_{d^* + 1} - \bar{\theta}^{kink}_{d^*} = M^{kink}(d^* + 1)\) if the bunching mass is small. Hence,

\[
\delta_{\text{switch}} = \frac{\bar{\theta}^{\text{prospective}}_{d^*} - \bar{\theta}^{ffs}_{d^*}}{\bar{\theta}^{ffs}_{d^* + 1} - \bar{\theta}^{ffs}_{d^*}} \approx \frac{B}{M^{kink}(d^* + 1)}
\]

So the local causal effect of moving from fee for service to prospective pay is identified by the ratio of bunching mass to the mass of patients at the day above the kink day. \(M^{kink}(d^* + 1)\) is directly observable. The remaining challenge lies in identifying \(B\) from the data which is what I am going to use the dynamically changing kink locations for.

### 4 Data and Sample Selection

\(\text{Data}\)

\(^{17}\)If the causal effect were one day or more, we would see bunching mass at the kink point in excess of 100\% of the expected mass. In the data, the bunching mass is at most a few percentage points of the expected mass.
I use administrative data from the Federal Statistical Agency in Germany. It covers the universe of in-patient hospitalizations covered by the DRG system\textsuperscript{18} 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 patients are readmitted within a certain time period (which depends on the DRG) the two cases 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.

Fortunately, I have two different measures for length of stay which are differentially affected by the two types of measurement error: First, the billed number of days. This measure yields the correct length of stay in the case of a readmission. In the case of an audit, however, it deviates from the true length of stay. Second, the difference between discharge and admission date. This measure 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 it is not affected by audits, since admission and discharge date appear to remain unadjusted after a successful audit - for a detailed discussion see appendix.

To summarize: the billed number of days is robust to readmissions, while the difference between discharge and admission date is mostly unaffected by audits. For my main analysis I use the latter measure, but I redid all of my analysis with the billed number of days instead of the difference between discharge and admission date and all results are very robust to which

\textsuperscript{18}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.
measure I choose. Therefore, the readmissions and audits do not pose a major problem for my data quality.

**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

**Suggestive Evidence**

I start by presenting suggestive evidence on the effect of the DRG system on length of stay. Figure 3 shows the development of average length of stay over time around the introduction of the DRG system. Average length of stay clearly is on a secular declining trend (the trend is also present before and after the time window shown in Figure 3). Had the introduction of prospective had any meaningful effect on length of stay, 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.
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 9 in the appendix shows the distribution of kink locations across DRGs, Figure 10 provides a histogram for length of stay and Figure 11 shows how often patients stay shorter respectively longer than their DRGs’ kink locations.

Figure 4 demonstrates the absence of bunching. To construct the graph I restrict the sample to DRGs with a kink location at 6 days or higher, because only for sufficiently high kink locations one can expect the hazards to evolve smoothly around the kink point. I center all observations around their respective kink location and pool them. I then show
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 pay 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 pay on length of stay decisions. The graph looks similar for other restrictions on the data like selecting only DRGs with kink location of at least 5 or 7 days.

Figure 4: Bunching behavior at the kink point

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 cannot be pinned down precisely. Therefore, I make use of DRGs with changing kink locations over time to get precise counterfactual hazards without strong functional form assumptions.

Since the DRG definitions are updated every year, I restrict my attention to DRGs the patient composition of which does not change 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 that feature 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. Clearly, the patients in the analysis sample have a longer average length of stay. This is reflective of the fact that in order to be in the analysis sample a DRG must have had a kink location of at least 3 days at some point which in turn means it must have had an average duration of at least 7.5 days. In this sense, the analysis sample is not representative, because only for fairly severe diagnosis there is variation in kink location over time. The other shown variables also show statistically significant differences (thanks to the large sample sizes), but the size of the differences are small. Hence, gender composition, age structure as well as year and month of admission seem to be fairly balanced between the samples.
Table 1: Summary Statistics Analysis Sample

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

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.

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 which is due to the secular trend towards shorter stays.
Table 2: Distribution of Changing Kink Locations

<table>
<thead>
<tr>
<th>Kink in t</th>
<th>Kink in t+1</th>
<th>DRGs</th>
<th>Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3</td>
<td>15</td>
<td>23 014</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>33</td>
<td>153 470</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>8</td>
<td>13 539</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>34</td>
<td>169 243</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>7</td>
<td>3 957</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>23</td>
<td>68 280</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>5</td>
<td>11 204</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>2</td>
<td>785</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>11</td>
<td>25 407</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>11</td>
<td>4 728</td>
</tr>
</tbody>
</table>

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.

Graphical Analysis

For the graphical analysis 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 5 shows the hazard rates in t and in t+1 for this pooled sample plotted against days relative to the kink location in t.

The shape of the hazard rates looks very similar in t and t+1 except for a tendency
towards higher hazards in $t+1$. This is reflective of the trend towards shorter stays. If there were meaningful bunching behavior, we would see relatively higher hazard rates at 0 for year $t$ and relatively higher hazard rates at $-1$ for year $t+1$. The absence of such a pattern is strong evidence that the hospitals do not adjust treatment length in response to marginal pay.

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. Figure 12 in the appendix shows the same graph, but restricted to October until December for year $t$ and January to March for year $t+1$. Figure 12 supports the conclusions from Figure 5, albeit being a little bit less precise due to the smaller underlying mass of data.\textsuperscript{19}

\textsuperscript{19}In contrast to Figure 5, Figure 12 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.
For comparison, Figure 13 in the appendix shows the evolution of hazard rates for DRGs for which the kink point did not change from one year to the next. As expected, there is no clear change in the shape of the hazards apparent here as well.

**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

$$\ln hazard_{drg,t,d} = \alpha_{drg,d} + \gamma_{drg,t} + \delta_{kink} + \epsilon_{drg,d,t}$$

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, the coefficient of interest is identified purely via changes in kink location and hazard rates from one year to the next. The indicators for each DRG-year-pair $\gamma_{drg,t}$ allow for a DRG-specific proportional shift in hazard rates each year. These capture general time effects for each DRG, e.g. a trend towards shorter stays/higher hazards for some DRGs. The $\delta_{kink}$ is an indicator that turns one if $d$ is the kink day for DRG $drg$ in year $t$.

An indicator for the kink day is sufficient and we do not need an additional variable to indicate the days to the right of the kink. This is because the research design only uses changes in kink location for identification. Consider the case when the kink location changes from 3 to 2 days. In terms of the model, the only change in hospital behavior is that $\theta_{kink}^{2}$ goes up from $\theta_{2}^{fts}$ to $\theta_{2}^{pps}$. Hence, just the hazards at 2 and 3 days change (which is then picked up by the $\delta_{kink}$), but the hazard rates elsewhere in the distribution are unchanged. If the kink location changes by more than one day, the hazards for the days in between the two kink locations could also change. This effect is negligible in my setting, however, due to the very small bunching amount. Moreover, my results are robust to conducting the analysis only with DRGs for which the kink location changes by one day at most (which is by far the more common case). Note also 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 (see Kleven 2016 - it is also implied by my model).

All standard errors are bootstrapped with $N = 400$ and clustered at the DRG-level. Since the parameter of interest is the bunching mass divided by the mass on the day above the kink point, I rescale the estimates of $\delta_{kink}$ accordingly. I want to limit the influence of days $d$ that are far away from the kink $\gamma_{drg,t}$. Therefore, I restrict the range of $d$ over which I estimate the equation to the lowest observed kink location of DRG $drg$ minus 10 days and the largest one plus 10 days. The results are barely changed by varying this range or not restricting the range at all.

Table 3 presents the regression results. Columns 1 and 2 present the rescaled estimates of $\delta_{kink}$ for the unweighted regression equation with and without the $\gamma_{drg,t}$ fixed effects. Columns 3 and 4 repeat these specifications, but weight the regression by the number of patients who are still at risk of being discharged on day $d$ with DRG $drg$ in year $t$. 

26
How are these coefficients to be interpreted? For instance, the point estimate in column 4 implies a reduction of .017 days in average length of stay when moving from fee for service to prospective pay. The estimates are very precise and throughout the 4 specifications one can reject an implied causal effect of more than a .04 days reduction. This is an order of magnitude less than what the literature has attributed towards the 1983 introduction of prospective pay in Medicare.

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 5 and 6 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 to bound the causal effect beneath .2 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 4 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. The estimates are a little less precise due to the smaller set of underlying data, but one can still reject an effect of more than .09 days.

All in all, the results robustly demonstrate that eliminating marginal pay by moving from fee for service to prospective pay does not have economically relevant effects on length of stay in the setting studied. Table 4 in the appendix presents the regression results using the billed number of days instead of the difference between discharge and admission date (see the discussion on audits earlier). None of my conclusions are affected.

Note that neither upcoming nor selective admission behavior pose a threat to identification. This is because both practices would, if anything, lead to larger bunching behavior in the data. Upcoding to a different DRG is least attractive for those patients who would otherwise be discharged at the kink, since discharge on the kink day maximizes profits. The
same argument applies with regards to selective admission, since the incentive to treat patients as an out-patient is weakest for those who would stay until their kink day otherwise.

**Heterogeneity**

The overall zero result might be masking stronger effects for subgroups of hospitals or patients. I start by investigating heterogeneity by hospital. Specifically, I estimate the specification \( \ln(hazard_{drg,d,t} = \alpha_{drg,d} + \delta_{kink} + \epsilon_{drg,d,t}) \) separately for each hospital (since they did not change the results in the main analysis much, I drop the \( \gamma_{drg,t} \) here to require a little less from the data). For many smaller hospitals, the degrees of freedom when estimating the equation are very small, because I cluster at the DRG level. If I consider hospitals for which the residual degrees of freedoms is at least 5, nearly 40% of coefficients are significant. If instead I require at least 60 residual degrees of freedom, less than 20% are. This share is larger than the 5% that would be expected by random chance. There is, however, an about equal share of negative and positive significant coefficients. This suggests that the more than 5% of rejections are not due to true underlying heterogeneity, but that the standard errors at the individual hospital level do not have sufficient data for the standard error estimates to be accurate. In Figure 6, I show a histogram of coefficients (weighted by the number of underlying patients) in .05 bins. The distribution appears fairly symmetric around zero and gives no reason to suspect that there is a substantial fraction of hospitals with meaningful \( \delta_{kink} \) values.

Moreover, I investigated the possibility of heterogeneous treatment effects by hospital size, by the degree to which hospitals manipulate birth weight in order to increase revenue\(^2\), 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), by hospital ownership (public and non-profit vs private) and by how strongly the slope changes at the kink (i.e. I took the ratio of the slope to the left of the kink and the pay in the flat portion of the schedule - the measure used for Figure 2 - and then interacted this measure with the kink indicator). None of these regressions showed

---

\(^2\)A newborn’s DRG depends on her birth weight discontinuously. Hospitals manipulate the birth weight to be able to code the higher paying DRG - that is, there is an excess mass of births with birth weight just below the respective cutoffs. I measure the degree to which each hospital manipulates and then interact this with the kink indicator.
Table 3: Regression Results

<table>
<thead>
<tr>
<th></th>
<th>all</th>
<th>all</th>
<th>all</th>
<th>all</th>
<th>decreases</th>
<th>increases</th>
<th>≥2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Causal Effect</td>
<td>-0.008 (0.018)</td>
<td>0.009 (0.014)</td>
<td>0.015 (0.011)</td>
<td>0.017 (0.011)</td>
<td>0.01 (0.009)</td>
<td>0.082 (0.047)</td>
<td>0.029 (0.029)</td>
</tr>
<tr>
<td>Year-DRG-FE</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>weighted</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Patients</td>
<td>761505</td>
<td>761505</td>
<td>761505</td>
<td>761505</td>
<td>713926</td>
<td>106993</td>
<td>362094</td>
</tr>
<tr>
<td>Clusters</td>
<td>147</td>
<td>147</td>
<td>147</td>
<td>147</td>
<td>116</td>
<td>56</td>
<td>98</td>
</tr>
</tbody>
</table>

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.
Figure 6: Histogram of Coefficients by Hospital

Shows patient-weighted share of estimates of $\delta_k$ for each hospital in .05 bins - restricted to estimates with degrees of freedom above 5. Moreover, shows patient-weighted share of significant coefficients in each bin.
any significant results, except for medical DRGs for which there is a significant $\delta_{kink}$ (yet still very small in magnitude). Hence, the result is appears to be very robust even within cuts of the data.

**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 be above 7.5 days at some point in order for the DRG to have a changing kink location). Transferred patients - which 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 pay for the hospital is linearly increasing until the rounded average duration from two years prior is reached. Afterwards, the payment schedule is 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 to implement the dynamic bunching design for all DRGs that are comparable from one year to the next and for which the average duration changes. Importantly, this includes DRGs with fairly small average durations such as two or three days. Table 6 in the appendix presents the results when estimating the main analysis equation

$$\ln \text{hazard}_{drg,t,d} = \alpha_{drg,d} + \gamma_{drg,t} + \delta_{kink} + \epsilon_{drg,d,t}$$

using transferred patients and the changing kink locations of their respective payoff schedules. The results are very close to those from the main analysis, suggesting that this paper’s conclusions are not limited to severely sick patients.

6 Discussion and Conclusion

**Implications for the German reform**

The last section has bounded the causal effect of moving from a hypothetical payment schedule that pays linearly with the same slope that the kinked payment schedule features to the left of the kink towards a flat payment schedule. This upper bound does not directly apply to the actual German move from per-diem to the current system in 2004. The pre-2004 per-diem system paid hospitals using a linear schedule in length of stay with hospital- and department-specific slopes, while the hypothetical linear schedule for which the effect is identified pays hospitals with hospital- and DRG-specific slopes. The slope in the hypothet-
ical schedule is, however, on average steeper than in the pre-2004 per diem system. In fact, the hypothetical schedule’s slope is so steep that if we started paying hospitals according to the hypothetical linear schedule, hospital revenue would be more than twice of what it is given the current kinked schedule. Because the hypothetical schedule’s slope is hospital- and DRG-specific, however, the slope is not necessarily steeper than under the pre-2004 per-diem system for every individual patient. Since I do not find any evidence of meaningful heterogeneity in effect size, however, the 0.04 days reduction in length of stay provides a reasonable upper bound for the effect of the 2004 German reform.

**Implications for the Medicare reform**

As also discussed earlier, 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 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. But 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), the incentive to increase length of stay in the Medicare pre-reform fee for service system was, if anything, smaller than in the German pre-reform per-diem system. Hence, the change in financial incentives induced by the Medicare reform was less drastic than in the German case.

Other institutional features give no reason to expect a muted effect of financial incentives for hospitals in Germany as well. In the German case as well as in Medicare, the doctors working in the hospitals are not directly affected by the financial incentives that their hospi-
tals face, since in Germany doctors are unionized and salaried and in the US they are mostly working based on a fee for service system or sometimes salaried. Moreover, the medical liability risk in the US is more severe in Germany, suggesting that, if anything, the doctors treating Medicare patients should be more strongly tied in their treatment decision to what is best for the patient from a purely medical perspective.\footnote{see e.g. Law Library Of Congress - Medical Malpractice Liability Systems In Selected Countries} Lastly, Medicare patients have a stronger incentive than the German patient population to counteract any profit-maximizing behavior that the hospital engages in, since Medicare patients have to carry a larger share of the hospitalization costs themselves. All in all, it therefore is quite striking that the estimated effect size for the German reform can be bounded an order of magnitude below what the literature has attributed to the Medicare reform.

**Implications for policy**

To get a better sense of why the financial incentives for hospitals show so little effect, I conducted interviews with doctors working in German hospitals. They are aware of the kinked payment schedule and its relevance for hospital profits, yet they (as one doctor put it) 'do not identify with their hospital and its motives' and therefore go by 'what makes sense from a medical perspective'. Given the apparent authority of doctors over the treatment decisions, politicians who want to incentivize shorter stays should possibly consider changing the financial incentives for the doctors directly. In the US, 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 importance of the doctors working in the hospital and their decision process for treatment decision is also an implication of Schlockermann (2017) who provides evidence of limited attention in German doctor decision making and its impacts on newborns’ length of stay.

My results also imply that Germany could improve welfare by paying hospitals per day again instead of prospectively. Paying hospitals in a way that is more closely tied to the actual incurred costs would compensate hospitals that draw a lot of patients who are comparatively sick conditional on their DRG. Compensating hospitals for such patients would reduce the incentive for the hospitals to discriminate against 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 in a profit-maximizing way than actual treatment
once admitted. Moreover, paying hospitals in a way that corresponds more closely to the actual incurred costs could improve welfare, because such a policy change would not affect total treatment volume, but it would reduce the hospitals’ financial risk. Hence, if hospital owners are risk-averse one could reduce equilibrium hospital profits without inducing hospital exit.

References


Appendix 1 - Measurement in Length of Stay

In the main text, I argue that audits do likely not introduce measurement error in my length of stay variable if defined via the difference between discharge and admission date. It is not known for sure how often the admission and discharge dates are adjusted after an audit, but I have evidence suggesting that in most cases they are not. Figure 7 shows the share of cases for which the billed number of days is smaller than the difference between discharge and admission date. This share is shown separately by month of admission. There are two ways in which Figure 7 suggests that after a successful audit only the billed number of days is adjusted but not the discharge and admission dates.

First, note that for patients admitted early in the year the share of cases for which the numbers diverge is a little above 4% - that is, very close to the 4.4%, the share of cases known to be successfully audited regarding length of stay in 2013.

Second, the share of cases with diverging number is clearly higher for cases from early in the year than for cases from later in the year. While a divergence between the billed number of days and the difference between discharge and admission date could in principle also be due to readmissions, only audits can explain the lower share of such cases later in the year. This 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.

I conclude that admission and discharge are typically not changed after a successful audit. Hence, the difference between discharge and admission date provides a measure of length of stay that is not affected by audits.
Appendix 2 - General Model

There is a continuum of identical hospitals (so that each hospital takes the reimbursement schedule as exogenous) each admitting a continuum of patients of type \(\theta_i\) who enjoy health benefit \(h(d_i, \theta_i)\) from staying in the hospital for \(d_i\) days. \(\frac{\partial h(d_i, \theta_i)}{\partial d_i} > 0\), \(\frac{\partial^2 h(d_i, \theta_i)}{\partial d_i^2} < 0\) and \(\frac{\partial^2 h(d_i, \theta_i)}{\partial d_i \partial \theta_i} > 0\), that is the health benefit is increasing and concave in the number of days in the hospital and people with higher \(\theta_i\) are those benefiting more from staying in the hospital.

W.l.o.g. assume \(\theta_i \sim U[0, 1]\).

The hospital receives payment \(P(d_i)\) and incurs costs \(C(d_i)\) with \(P'(\cdot) > 0\), \(P''(\cdot) \leq 0\),
The hospital values profits according to $U(\cdot)$, with $U'(\cdot) > 0$ and $U''(\cdot) \leq 0$. Because of asymmetric information the hospital can choose any $d_i$, but faces inducement costs proportional to the forgone health benefit. The inducement costs are a reduced form for all costs proportional to the health effect of changing $d_i$ - for instance, it captures any intrinsic concern the hospital might have about patient health as well as any liability or reputation costs.

Moreover, the health insurers can audit bills and reduce the billed number of days to a patient-specific $\tilde{d}(\theta_i)$ with $\tilde{d}'(\theta_i) > 0$. The health insurers audit whenever the amount of money they can save exceeds a threshold, that is whenever $P(d_i) - P\left(\tilde{d}(\theta_i)\right) > c + u_i$ with $c > 0$ and $u_i \sim F(u_i)$, some mean zero distribution. Thus, the probability of an audit is $F\left[P(d_i) - P\left(\tilde{d}(\theta_i)\right) - c\right]$. Therefore, expected audit costs are

$$A \left[P(d_i) - P\left(\tilde{d}(\theta_i)\right)\right] \equiv F\left[P(d_i) - P\left(\tilde{d}(\theta_i)\right) - c\right] \cdot \left[P(d_i) - P\left(\tilde{d}(\theta_i)\right)\right]$$

with $A'(\cdot) > 0$ and $A''(\cdot) > 0$.

Hence, the hospital solves

$$\max_{\{d_i(\theta_i) \in \mathbb{N}\}} U\left(\int_i \left[P(d_i) - C(d_i) - A\left[P(d_i) - P\left(\tilde{d}(\theta_i)\right)\right]\right]\right) + \lambda \int_i h(d_i, \theta_i)$$

**Kinked Payment Schedule**

I start by characterizing hospital behavior under the actually observed kinked payment schedule. It features slope $p$ (as discussed in the payment schedule section, the slope is set according to average variable costs) up until kink point $d^*$ and is flat afterwards. Moreover, it pays a lump-sum payment $\bar{p}^{\text{kink}}$. That is, $P^{\text{kink}}(d_i) = \bar{p}^{\text{kink}} + p \cdot d_i$ if $d_i < d^*$ and $P^{\text{kink}}(d_i) = \bar{p}^{\text{kink}} + p \cdot d^*$ if $d_i \geq d^*$. Define the marginal costs of increasing length of stay by one day as $\Delta C(d) = C(d) - C(d-1)$. Denote equilibrium profits (i.e. given $E[C(d_i)|\text{kink}]$ and $E[d_i|\text{kink}]$ which are endogenous to the payment system) hospitals make by $\pi$. Since

\[^{22}\text{The restrictions on the second derivatives are sufficient, but stronger than needed. All I need is that the overall objective function is globally concave so that local optimality conditions are sufficient. Therefore, the mild convexity my actually observed } P(d_i)\text{ features in the right tail is unlikely to matter. Also note that all this analysis is conditional on admittance already, i.e. any fixed costs of treating a patient do not make the cost function concave, but are just an additive fixed costs term independent of } d_i.\]
there is a continuum of hospitals, any individual one’s effect on the payment schedule or equilibrium profits via their effect on equilibrium costs is zero and can be ignored.

Due to the concavity of the objective function, optimal hospital behavior amounts to choosing cutoff values for \( \theta_i \) above which a patient is kept a day longer than people below the cutoff value. \( \theta^k_{d} \) denotes the highest \( \theta_i \) for which the patient stays \( d \) days. A patient with a \( \theta_i \) just above \( \theta^k_{d} \) would stay \( d + 1 \) while a patient with a \( \theta_i \) just beneath \( \theta^k_{d} \) would stay \( d \) days.

The cutoff values of interest are going to be those defining the range of patients who are discharged on kink day \( d^* \). These are implicitly defined by equations

\[
\triangle C (d^* + 1) \cdot U' (\pi) = \lambda \left[ h \left( d^* + 1, \theta^k_{d^*} \right) - h \left( d^*, \theta^k_{d^*} \right) \right]
\]

\[
\left( \triangle C (d^*) - p + A \left[ p \left( d^* - \tilde{d} \right) \right] - A \left[ p \left( d^* - 1 - \tilde{d} \right) \right] \right) U' (\pi) = \lambda \left[ h \left( d^*, \theta^k_{d^* - 1} \right) - h \left( d^* - 1, \theta^k_{d^* - 1} \right) \right]
\]

Next, I analyze hospital behavior under a counterfactual fee-for-service as well as a counterfactual prospective payment scheme to show how the bunching mass in the observed schedule identifies the causal effect of moving from fee-for-service to prospective pay.

**Fee-for-Service**

Let \( P_{fee\text{for}service} (d_i) = \bar{p}_{ffs} + p \cdot d_i \) with \( p \) as in the kinked section but paid for all \( d_i \) and not just up to \( d^* \). \( \bar{p}_{ffs} \) is set such that in equilibrium (i.e. given \( E [C (d_i) \mid fee\text{for}service] \) and \( E [d_i \mid fee\text{for}service] \) which are endogenous to the payment system) hospitals again make profits \( \pi \).

Define the marginal costs of increasing length of stay by one day as \( \triangle C (d) \equiv C (d) - C (d - 1) \). The mass of people that get discharged on day \( d^* \) is equal to \( M_{ffs} (d^*) = \bar{\theta}_{d^*}^{\text{ffs}} - \lambda \left[ h \left( d^*, \theta^k_{d^* - 1} \right) - h \left( d^* - 1, \theta^k_{d^* - 1} \right) \right] \)
\[
\left( \Delta C (d^* + 1) - p + A \left[ p \left( d^* + 1 - \bar{d} \right) \right] - A \left[ p \left( d^* - \bar{d} \right) \right] \right) \cdot U' (\pi) \\
= \lambda \left[ h \left( d^* + 1, \bar{\theta}_{d^*}^{ffs} \right) - h \left( d^*, \bar{\theta}_{d^*}^{ffs} \right) \right]
\]

\[
\left( \Delta C (d^*) - p + A \left[ p \left( d^* - \bar{d} \right) \right] - A \left[ p \left( d^* - 1 - \bar{d} \right) \right] \right) \cdot U' (\pi) \\
= \lambda \left[ h \left( d^*, \bar{\theta}_{d^*-1}^{ffs} \right) - h \left( d^* - 1, \bar{\theta}_{d^*-1}^{ffs} \right) \right]
\]

**Prospective Payment**

Now let \( P_{\text{prospective}} (d_i) = \bar{p} \) with \( \bar{p} \) set such that in equilibrium (that is, taking into account that \( E[C(d_i)] \) is endogenous to the payment schedule) hospitals profits remain \( \pi \). The mass of people that get discharged on day \( d^* \) is now equal to \( M_{\text{prospective}} (d^*) = \bar{\theta}_{d^*}^{\text{prospective}} - \bar{\theta}_{d^*-1}^{\text{prospective}} \) with \( \bar{\theta}_{d^*}^{\text{prospective}} \) and \( \bar{\theta}_{d^*-1}^{\text{prospective}} \) denoting the lower and upper marginal patients at \( d^* \) implicitly defined by equations

\[
\Delta C (d^* + 1) \cdot U' (\pi) = \lambda \left[ h \left( d^* + 1, \bar{\theta}_{d^*}^{\text{prospective}} \right) - h \left( d^*, \bar{\theta}_{d^*}^{\text{prospective}} \right) \right]
\]

\[
\Delta C (d^*) \cdot U' (\pi) = \lambda \left[ h \left( d^*, \bar{\theta}_{d^*-1}^{\text{prospective}} \right) - h \left( d^* - 1, \bar{\theta}_{d^*-1}^{\text{prospective}} \right) \right]
\]

If the causal effect of switching from fee for service to the prospective system is at most a reduction of one day (which given the observed amount of bunching is the only empirically relevant case\(^23\)), the average causal effect on patients staying \( d^* + 1 \) days under fee for service is

\[
\delta_{\text{switch}} \equiv E \left[ \Delta d_i \mid \theta_i \epsilon \left[ \bar{\theta}_{d^*}^{ffs}, \bar{\theta}_{d^*+1}^{ffs} \right] \right] = \frac{\bar{\theta}_{d^*}^{\text{prospective}} - \bar{\theta}_{d^*}^{\text{fee for service}}}{\bar{\theta}_{d^*+1}^{ffs} - \bar{\theta}_{d^*}^{ffs}}
\]

That is, is the average effect on length of stay for patients in \( M_{\text{fee for service}} (d^* + 1) \) is the

\(^23\)If the causal effect were one day or more, we would see bunching mass at the kink point in excess of 100% of the expected mass. In the data, the bunching mass is at most a few percentage points of the expected mass.
share of those patients that after switching to a prospective payment system falls beneath the upper cutoff value for \( d^* \). Next, I show that this causal effect is identified using a bunching design in a kinked payoff schedule.

**What does the bunching design identify?**

Despite the extensions the result from the simple model in the main part of the paper still goes through. If the causal effect of switching from fee for service to the prospective system is at most a reduction of one day (which given the observed amount of bunching is the only empirically relevant case\(^{24}\)), the average causal effect \( \Delta d_i \) on the well-defined group of patients staying \( d^* + 1 \) days under fee for service is

\[
\delta_{\text{switch}} \equiv E \left[ \triangle d_i \mid \theta_i \in \left[ \theta_{d^*}^{ffs}, \theta_{d^*+1}^{ffs} \right] \right] = \frac{\theta_{d^*}^{prospective} - \theta_{d^*}^{ffs}}{\theta_{d^*+1}^{ffs} - \theta_{d^*}^{ffs}}
\]

That is, the average effect on length of stay for patients in \( M^{ffs} (d^* + 1) \) is the share of those patients that after switching to a prospective payment system falls beneath the new upper cutoff value for \( d^* \).

Since we do not have a sharp research design to directly compare the hazard rates under fee for service and prospective pay for similar patient populations, we need to use the bunching under the kinked schedule to identify \( \delta_{\text{switch}} \). Note that \( B = \theta_{d^*}^{kink} - \theta_{d^*}^{ffs} = \theta_{d^*}^{prospective} - \theta_{d^*}^{ffs} \)

and \( M^{ffs} (d^* + 1) = \theta_{d^*+1}^{ffs} - \theta_{d^*}^{ffs} \approx \theta_{d^*+1}^{kink} - \theta_{d^*}^{kink} = M^{kink} (d^* + 1) \) if the bunching mass is not too big. Hence,

\[
\delta_{\text{switch}} = \frac{\theta_{d^*}^{prospective} - \theta_{d^*}^{ffs}}{\theta_{d^*+1}^{ffs} - \theta_{d^*}^{ffs}} \approx \frac{B}{M^{kink} (d^* + 1)}
\]

So the local causal effect of moving from fee for service to prospective pay (in a way that keeps hospital profits at \( \pi \) in equilibrium) is identified by the ratio of bunching mass to the mass of patients at the day above the kink day. \( M^{kink} (d^* + 1) \) is directly observable. The remaining challenge is identifying \( B \) from the data which is what I use the dynamically

\(^{24}\)If the causal effect were one day or more, we would see bunching mass at the kink point in excess of 100% of the expected mass. In the data, the bunching mass is at most a few percentage points of the expected mass.
changing kink locations for.

Appendix 3 - Further Results - Tables

Table 5: Distribution of Kink Locations in the Cross section

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

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.
Table 4: Regression Results - Using Billed Number of Days

<table>
<thead>
<tr>
<th></th>
<th>all</th>
<th>all</th>
<th>all</th>
<th>all</th>
<th>decreases</th>
<th>increases</th>
<th>≥2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Causal Effect</td>
<td>-0.045 (0.016)</td>
<td>-0.024 (0.014)</td>
<td>-0.01 (0.01)</td>
<td>-0.007 (0.01)</td>
<td>-0.012 (0.009)</td>
<td>0.034 (0.037)</td>
<td>-0.017 (0.023)</td>
</tr>
<tr>
<td>Year-DRG-FE</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>weighted</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Patients</td>
<td>761788</td>
<td>761788</td>
<td>761788</td>
<td>761788</td>
<td>714250</td>
<td>107059</td>
<td>362156</td>
</tr>
<tr>
<td>Clusters</td>
<td>147</td>
<td>147</td>
<td>147</td>
<td>147</td>
<td>116</td>
<td>56</td>
<td>98</td>
</tr>
</tbody>
</table>

Length of Stay is measured as the billed number of days in contrast to the table in the main text. The total number of patients is deviating slightly between the tables, because for data privacy reasons I can only make use of observations for which there are at least 3 patients that are discharged with a certain DRG after a certain number of days. 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.
Table 6: Regression Results for Transfers

<table>
<thead>
<tr>
<th></th>
<th>Transfers</th>
<th>Transfers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Causal Effect</td>
<td>0.0027 (0.0248)</td>
<td>0.0004 (0.0254)</td>
</tr>
<tr>
<td>Year-DRG-FE</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>weighted</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Patients</td>
<td>1971325</td>
<td>1971191</td>
</tr>
<tr>
<td>Clusters</td>
<td>236</td>
<td>226</td>
</tr>
</tbody>
</table>

The sample is restricted to transfers to and from the hospital, to DRGs that are comparable from one year to the next and to DRGs for which the kink location (here: the average duration two years prior) changes. Standard errors are clustered at the DRG level. Weighting refers to weighting by the number of patients still present in the hospital that day.
Appendix 4 - Further Results - Figures

Figure 8: Effect of Length of Stay on Reimbursement

Average pay by days of stay for DRG K01C in 2010 in the state of Bavaria
Figure 9: Histogram Kink Locations

Shows the share of DRGs with the respective kink location (for the lower kink) for years 2005-2013.
Figure 10: Histogram Length of Stay

Shows the share of patients staying 1, 2, 3 etc days in the hospitals across all years 2005-2013.
Figure 11: Histogram relative to the kinks

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. For all patients in the data, i.e. 2005-2013.
Figure 12: Changing kink location - October until March

Hazard Rates in year t (October to December only) and t+1 (January to March only) for DRGs that are comparable from t to t+1 and for which the kink point went down one day from t to t+1.

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 13: Hazard Rates across year for DRGs for which the kink location does not change

Hazard Rates in year $t$ and $t+1$ for DRGs that are comparable from $t$ to $t+1$ and for which the kink point remained constant from $t$ to $t+1$. 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: Heterogeneity in bunching by quartile of hospital size

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

Hazard rates for all patients pooled with a DRG with kink locations of at least 6 days. Done separately by quartile of hospital size.
Figure 15: Heterogeneity in bunching 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 long term care unit or a hospice.